SURROGATE MODELS FOR SEISMIC RESPONSE OF STRUCTURES

by

Sanjay Nayak

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Civil Engineering

Lyles School of Civil Engineering

West Lafayette, Indiana

August 2023

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Shirley J. Dyke, Chair

Lyles School of Civil Engineering and School of Mechanical Engineering

Dr. Arun Prakash

Lyles School of Civil Engineering

Dr. Ayhan Irfanoglu

Lyles School of Civil Engineering

Approved by:

Dr. Dulcy M. Abraham

To friends and family.

ACKNOWLEDGMENTS

Firstly, I would like to thank Dr. Shirley Dyke for providing her invaluable support and guidance throughout my graduate studies and for inspiring me to stay determined and become a better researcher.

I would like to thank Dr. Prakash and Dr. Irfanoglu for guiding me at various junctures in my graduate life and for serving on my committee and providing valuable insights to improve the quality of my work.

I would like to extend my thanks to Dr. Amir Behjat for guiding me at the start of my study and for helping me lay the foundation for my thesis.

I would also like to acknowledge my friends from the IISL team, Bowen Laboratory, and the Lyles School of Civil Engineering for constantly motivating me to stay focused throughout my studies. I am grateful to Purdue University for providing me with the resources, facilities, and environment to grow both professionally and personally.

Finally, I would like to thank my parents for constantly believing in me for being understanding throughout my graduate studies. I would not be here today without their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

ρ	Density
E	Young's Modulus
PGA	Peak Ground Acceleration
S_a	Spectral Acceleration
S_d	Spectral Displacement
S_v	Spectral Velocity

ABBREVIATIONS

CCD	Central Composite Design
DDM	Data-Driven Model
EDP	Engineering Demand Parameter
IM	Intensity Measure
LHS	Latin Hypercube Sampling
MRDR	Maximum Roof Drift Ratio
MSDR	Maximum Story Drift Ratio
PBM	Physics-Based Model

ABSTRACT

The seismic risks to a structure or a set of structures in a region are usually determined by generating fragility curves that provide the probability of a building responding in a certain manner for a given level of ground motion intensity. Developing fragility curves, however, is challenging as it involves the computationally expensive task of obtaining the maximum response of the selected structures to a suite of ground motions representing the seismic hazard of the region selected.

This study presents a methodology to develop surrogate models for the prediction of the maximum responses of buildings to ground motion excitation. Data-driven surrogate models using simple machine learning techniques and physics-based surrogate models using the space mapping technique to map the low-fidelity responses obtained using a multi-degree of freedom shear building model to the high-fidelity values are developed for the prediction of the maximum roof drift ratio and the maximum story drift ratio of a chosen 15-story steel moment-resisting frame building with varying structural properties in California. The predictions of each of these surrogate models are analyzed to assess and compare the performance, capabilities, and limitations of these models. Best practices for developing surrogate models for the prediction of maximum responses of structures to ground motion are recommended.

The results from the development of data-driven surrogate models show that the spectral displacement is the best intensity measure to condition the maximum roof drift ratio, and the spectral velocity is the best intensity measure to condition the maximum story drift ratio. Fragility analysis of the structure is thus conducted using maximum story drift as the engineering demand parameter and spectral velocity as the intensity measure. Monte Carlo simulation is conducted using the physics-based surrogate model to estimate the maximum story drifts for ground motions that are incrementally scaled to different intensity levels. Maximum likelihood estimates are used to obtain the parameters for a lognormal distribution and the 95% confidence intervals are obtained using the Wald confidence interval to plot the fragility curves.

Fragility curves are plotted both with and without variations in the structural properties of the building, and it is found that the effects of variability in ground motions on the fragility

are far higher than the effects of the randomness of structural properties. Finally, it is found that about 65 ground motion records are needed for convergence of the parameters of the lognormal distribution for plotting fragility curves by using Monte Carlo simulation.

1. INTRODUCTION

Seismic hazards pose a risk to buildings in several earthquake-prone regions. A way to quantify this risk on buildings is to develop fragility curves that provide the probability of failure (or the probability of a building responding in a certain manner) at a given ground motion intensity. An important step in the fragility analysis of structures is determining the maximum structural response (such as the maximum story drift ratio or the maximum roof drift ratio) for various ground motions. Maximum responses thus obtained are used to determine the state of the building. By repeating this for enough ground motion records, the probability that the building responds in a particular manner when excited at a particular ground motion intensity can be obtained. Different methods have been used to determine the maximum structural response of structures to ground motion excitations. Details regarding some of these methods along with more information regarding the fragility analysis procedure are provided in Chapter 2.

Outside of fragility analysis, estimates of maximum structural responses are also used in the design of seismic resistant structures[1], where one of the design objectives is to limit the maximum response (usually the story drifts) under a certain value for a design-level earthquake.

One straightforward way to obtain these peak structural responses is by using computational methods such as finite element modeling. However, obtaining the peak structural responses of a building under ground motion excitation computationally is a laborious task as it requires obtaining a complete time-history response first. This involves solving the equations of motion at every time step at which the ground motion data is recorded. Furthermore, ground motions are highly random and it is difficult to capture this randomness in just a few parameters. Adding to this complexity, ground motions with sufficiently high magnitudes drive the structures into nonlinear behavior. This nonlinearity could arise from various sources including material and geometric nonlinearities. All of this means that simulating the response of a structure to ground motion takes a long time, often magnitudes of order higher than the actual duration of the ground motion itself. This becomes a big-

ger issue in studies where multiple runs are required, such as fragility analysis, parametric analysis, optimization studies, or even preliminary design.

One possible solution to this challenge is reducing the complexity of the computational model used to obtain the time-history response. Computational models can be created with different levels of detail as required by the problem at hand. While complex high-fidelity models can provide more accurate results, these models require high-computational costs to execute. This could be problematic while using the models to run multiple times to explore the entire parameter space. Computational cost can be reduced by reducing the fidelity of the models at the expense of accuracy, by making approximations, by relaxing the convergence criteria, or by using a coarser mesh. The results generated using low-fidelity models, however, may not be accurate enough for certain studies. Determining the right balance between computational cost and accuracy, therefore, is always a challenging task.

Surrogate modeling is a technique to combine the advantages of both high-fidelity and low-fidelity modeling techniques to develop a model that gives acceptable results while requiring a low computational cost. In surrogate modeling, a limited number of high-fidelity data points are used to enhance the results obtained from low-fidelity models. These surrogate modeling techniques can be broadly classified into three types: data-driven, physics-based, and knowledge-based surrogate modeling techniques. Each of the three categories contains multiple different types of surrogate modeling choices based on the requirements of the problem. The goal of this thesis is to build surrogate models that can be used to estimate the maximum responses of buildings to ground motions without requiring nonlinear time-history analysis.

Data-driven surrogate modeling techniques use machine learning and statistical methods to fit the high-fidelity data obtained at specific points using high-fidelity models[2]. A model, trained on high-fidelity data in this way is then used to predict the desired outputs for regions of the input space where high-fidelity data is not generated. These are typically used when there is a lot of high-fidelity data available or when such data can be easily generated.

Physics-based surrogate modeling techniques use physical knowledge of the real-world system to obtain a simplified low-fidelity model[3]. These simplifications include reducing the complexity of the model by making certain assumptions to simplify the physics of the

problem or increasing the coarseness of the mesh and relaxing the convergence criteria to reduce the computational effort required. Usually, the results from the low-fidelity model are enhanced by mapping these low-fidelity results to high-fidelity data at certain points to reduce the error in prediction. Physics-based surrogate models are typically useful when certain assumptions can be made regarding the real-world system to build low-fidelity physics models without losing critical information about the system.

Knowledge-based surrogate modeling techniques use the existing knowledge about a real-world system to obtain the results[4]. Existing knowledge could include the observations made regarding the real-world system or empirical equations developed based on these observations. These kinds of models are typically used when little high-fidelity data is available and the exact physics of the system is not known, making it difficult to analytically model the system.

Once surrogate models with acceptable levels of errors have been developed, these can be used to obtain the peak structural response for many different ground motions with different levels of intensities, at a low computational cost. Thus, by evaluating the state of the buildings over many different ground motion excitation, a potentially better estimate can be obtained for the probability of the buildings exhibiting different types of responses. If the surrogate model is capable of making sufficiently accurate predictions for various combinations of structural properties, in addition to different ground motions, the fragility curves developed using these models can also incorporate uncertainties in the structural parameters.

In parametric analysis, these surrogate models can likely be used concurrently with complex nonlinear computations to explore the entire parametric space; the nonlinear computational models can be used for obtaining the responses at regular intervals, and the surrogate models for the responses at points within these intervals. This approach becomes more consequential when the results of high-fidelity computational models are already obtained at regular intervals to train the surrogate models.

In conceptual design and optimization studies, the surrogate models can be used to initially explore the parameter space entirely and narrow down the search to a small subset of the parameter space where complex nonlinear computational models can then be used to

capture more intricate details[5]. Furthermore, surrogate models can be potentially used to explore the effects of uncertainties in structural or material properties, and the assumed design loads of the building to supplement the data from complex nonlinear models constructed without incorporating these uncertainties.

While there are so many different types of surrogate modeling techniques available, there are not many studies that compare the performance, capabilities, and uses of different types. Most literature focuses on developing a single type of surrogate model and validating the use of that particular surrogate model for the application at hand. In this study, an effort is made to compare the data-driven and physics-based surrogate models for the nonlinear analysis of buildings to seismic ground motions.

A detailed overview of previous works on surrogate modeling and its types, fragility analysis, and the seismic response of structures by various authors is provided in Chapter 2. Chapter 3 provides more details regarding the building selected for this study and the scope of the study. Chapter 4 that follows provides the details regarding different models that are used to generate the data that is assumed to be the ground truth in this study. Following this, Chapters 5 and 6 provide the details of the surrogate modeling techniques used to predict the fundamental frequency and the peak structural responses (maximum roof drift ratio and the maximum story drift ratio) in this study. In Chapter 7, the surrogate models developed in this study are used to obtain the fragility curves for the building under consideration, while accounting for variations in structural as well as ground motion uncertainties. Finally, Chapter 8 draws the conclusions from this study and provides a list of improvements that can be made in the future.

2. LITERATURE REVIEW AND BACKGROUND

This chapter summarises the details of some of the previous research on surrogate modeling in seismic engineering and fragility analysis. The chapter is divided into several sections based on the topic being discussed.

2.1 Overview

The risks due to ground motion hazards on structures depends on both the seismic hazard of the region and the type of structure. The uncertainty in ground motion variability makes it difficult to quantify the seismic risk to structures in certain regions. Fragility curves have become a norm in the structural engineering community as a way to quantify the risk on a structure or a population of structures in a particular region due to ground motions. Fragility curves provide the probability that a structure or a population of structures exhibits certain levels of damage for a given intensity of ground motion[6]. An example of a fragility curve from Kircher et al.[6] is shown in Figure 2.1. In this way, if the ground motion hazard in a region is known (through ground motion hazard maps or using studies by researchers quantifying the ground motion hazard of particular regions), the corresponding risk on a structure or a set of structures in that region can be quantified using corresponding fragility curves. Bakalis and Vamvatsikos[7] provide a good guide to generating fragility curves.

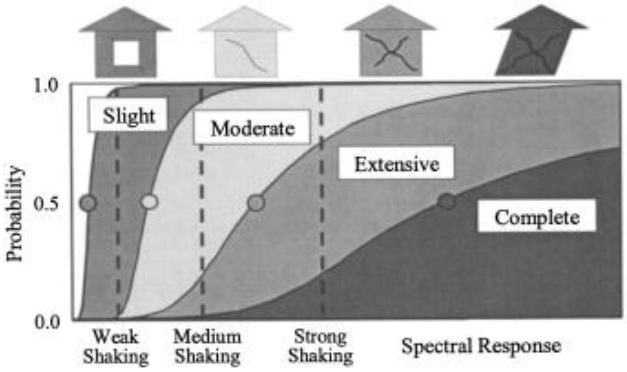


Figure 2.1. Example of a fragility curve[2]

[2]Source: Kircher et al.[6]

Fragility curves have been developed for a wide variety of structures. While most of the research done in fragility analysis is for bridges[8] and buildings[9], they have also been developed for dams[10]–[12], containment structures[13], tunnels [14], [15] and geostructures [16], and wind turbines[17]. Fragility analysis has also been conducted for nonstructural components. Soroushian et al.[18] for example, studied the fragility of fire sprinkler piping systems. Fragility curves can be obtained for a single structure[19], a set of similar structures in a region[20], or an entire population of structures encompassing different types of structures[21], [22].

Kircher et al.[6] provides a good account of the procedure to conduct fragility analysis. Usually, the first step in conducting a fragility analysis is selecting the structures and ground motions for the analysis. Following this, the response of the structures to ground motion is obtained in one of many ways. These can be analytical (physics-based), data-driven, or knowledge-based. Finally, the fragility of the structures is calculated and plotted to obtain the fragility curves. The following sections provide findings from past studies along each of these steps.

2.2 Uncertainties in Structural Properties

When developing fragility curves, a choice has to be made about whether to incorporate uncertainties in structural properties in the study or not. When not including any uncertainties, all structural properties can be assumed to be deterministic to obtain the fragility curves. Mandal et al.[13] in their study assumed that structural properties are deterministic and assumed that only the variability in ground motions affects the containment structure they chose.

However, when considering uncertainties, the structural properties are usually modeled as random variables with known probability distributions (either from previous experiments and measurements or by judgment). Modeling the uncertainties in all the structural and material properties is not possible as it adds needless complexity to the problem. Instead, it is important to determine a limited number of structural properties, the uncertainties which affect the fragility of the structure the most. Following this, only these variables can be

modeled as random variables, while the remaining properties are assumed to be deterministic. Seo et al.[23] developed Pareto plots and statistical significance measures to select structural parameters that affect the response of irregular steel moment frame structures the most. Glaister and Pinho[22] conducted a brief parametric analysis studying the effects of material and geometric uncertainties on the fragility curves of buildings. Pan et al.[24] followed a similar process using parametric analysis to study the effects of structural uncertainties on the fragility of bridges.

The most common sources of uncertainties modeled in material properties studied are the strengths and stiffness of structural materials[4], [9], [25]. In buildings, the sources of uncertainties in geometry typically arise in the form of dimensions of beams and columns, and the variations in configurations due to different numbers of stories[20]. In bridges, geometric uncertainties usually studied are the gaps in decks and different configurations of bridges[26], [27]. However, uncertainties in several other parameters have also been modeled to study the effects of these uncertainties on the fragility of structures.

Different researchers have previously shown that randomness in structural and material properties has a lower influence on the response of the structure than the randomness in ground motion properties. Kwon and Elnashai[9] modeled the strength of concrete and the yield strength of steel bars of an reinforced concrete building as random variables and found that the variability in material properties has a much lower effect on structural responses than variability in ground motions. Furthermore, they found that the mean value of the maximum story drift of individual analysis of 100 frames was different only at high ground motion intensities from the maximum story drift of a single structure with mean structural properties. They also found that the yield strength of steel bars had no significant effect on the structural response because the yield strength of steel did not affect the period of the building. They noticed that the effect of yield strength was even less influencial at low ground motion intensities.

Similarly, Rossetto and Elnashai[28] accounted for uncertainties in material properties by selecting unconfined compressive strength of concrete, compressive strength of infills, and yield strength of reinforcing bars of RC buildings as random variables and demonstrated that uncertainties in ground motions result in a larger variation in responses (maximum

story drift in their case) than uncertainties in the material properties. Lupoi et al.[29] found that the randomness in strength-related parameters do not have a significant influence on displacement-related quantities such as drifts. Reinhorn et al.[30] similarly found that varying the strength does not significantly change the fragility curves. On the other hand, Pang et al.[19] considered uncertainties in 15 material and geometric parameters in their study and found that modeling the variations in these parameters resulted in larger probability of damage as compared to when considering just the variations in ground motions.

Certain studies have developed a way to incorporate uncertainties in structural parameters directly in the method of developing fragility curves, without explicitly modeling them. Mosalam et al.[31] introduced uncertainties directly in the pushover curve to avoid the otherwise challenging task of considering many different sources of uncertainties in structural parameters.

2.3 Selection of Ground Motions

After determining the structure or a set of structures, it is important to select ground motion records that are characteristic of the region under study. These ground motions could be real ground motions or synthetic ground motions developed for the particular region. Obtaining different ground motion records from a single region for many intensities inside the range of expected ground motion intensities is not often feasible. This is more true when only real ground motions are used. One option is to use ground motion records from other regions with similar characteristics. The other option is to select just a few ground motion records and scale these records to different values of intensities to populate the whole range of ground motion intensities that are required in the study. Haselton et al.[32] and Cornell[33] provided recommendations on selecting and scaling ground motions without introducing bias.

2.4 Intensity Measures

The fragility of structures is conditioned on a given level of ground motion intensity to reduce the scatter in response which would otherwise be seen when the fragility of structures

would be obtained without mapping it to the particular ground motion intensity. Different types of intensity measures have been used by researchers in the past. However, there are very few papers that provide a comparison of more than 2 intensity measures for fragility analysis.

The most commonly used intensity measure is the peak ground acceleration of the ground motion[9], [28], [34]. However, several researchers found that peak ground properties such as the peak ground acceleration and the peak ground velocity are not the best measure to condition the fragility of a building since these values do not contain any information regarding the frequency content of the ground motion, and suggested using other measures such as the spectral response properties that capture the frequency content of ground motions instead, thus reducing the scatter in fragility.

Mandal et al.[13] noted that spectral acceleration correlated with maximum story drift ratio better than PGA. Similarly, Rossetto and Elnashai[21] demonstrated that vulnerability curves conditioned using PGA as the intensity measure showed higher scatter than vulnerability curves conditioned using spectral response properties. They found that using spectral displacement provided a better correlation to the empirical maximum story drift data they had used in their study than spectral acceleration. Despite this, peak ground acceleration continues to be the most commonly used ground motion intensity measure in fragility studies.

Certain researchers have found that using multi-dimensional intensity measures to develop fragility surfaces reduces the uncertainty in estimates of fragility curves. Pan et al. developed fragility surfaces by conditioning the probability of failure against 2 parameters, the moment magnitude and the distance from the epicenter at the site citing that these quantities are more readily available than other ground motion intensity measures[34]. In another study, Pan et al.[24] found that using multivariate regression against moment magnitude and epicentral distance provided a better fit for fragility values than using linear regression against PGA. Lagomarsino and Giovinazzi[35] used microseismic intensity to develop data-driven fragility curves of buildings using empirical studies.

Seyedi et al.[36] compared fragility surfaces obtained using spectral displacements at two different periods as intensity measures against fragility curves obtained using just one

of these measures and found that the uncertainty in fragility decreased when using two measures to condition the fragility instead of just one. Peña et al.[37] compared fragility estimates developed using single intensity measure (PGV) to data-driven fragility estimates developed using multi-dimensional intensity measures (PGV and spectral velocity) for a non-linear benchmark model of a twenty-story building. They noted that although the amount of uncertainty in fragility functions obtained using multi-dimensional intensity measures was higher than the uncertainty when using a single intensity measure, the uncertainty decreased the fastest when using multi-dimensional intensity measures as the number of observations used to build the fragility functions increased.

Baker and Cornell[38] demonstrated that using spectral acceleration as the sole intensity measure overestimated the response of the RC moment-frame building they studied, when compared to the responses obtained using a vector-valued intensity measure including both the spectral acceleration and a factor defined as epsilon (which measures the number of standard deviations by which log of spectral acceleration of a record varies from the log of the mean predicted spectral acceleration).

Mackie and Stojadinović[39] provides a list of commonly used intensity measures and engineering demand parameters. However, to my knowledge, no study is available comparing the results obtained using different combinations of intensity measures and response parameters to determine which combinations work the best.

Kiani and Pezeshk[40] analyzed how different ground motion intensity measures affect the structural responses for RC moment resisting frame buildings. They found that spectral intensity (SI) (or Housner intensity[41]), which is the area under the pseudo-spectral velocity curve between periods of 0.1 s and 2.5 s, is the best intensity measure to reduce the errors in the predictions of the maximum story drift ratio.

2.5 Engineering Demand Parameters

The damage states of a structure are usually determined by comparing the maximum response of the structure to a set of preset values. These response parameters are called engineering demand parameters (EDP). Different EDPs have been used by authors. For

buildings, the most commonly used EDPs are the maximum story[21], [28], [37] and roof drift ratios[20] or displacements. For nonstructural damage, the most commonly used EDPs are maximum drifts or maximum floor accelerations[42]. Other parameters such as plastic hinge rotations[9], [18], and effective stresses[17], [25] have also been used but less often.

For bridges, it is usually the practice to use different EDPs for different components. The commonly used combinations are curvature ductility for columns, and deformations for bearings and abutments[24], [26], [43].

2.6 Damage States

Damage states are defined by setting ranges for EDPs and comparing the observed values to figure out in which range the response falls. These limits and ranges are developed through engineering judgment, experience, or experimental observations.

HAZUS[6] defines four damage states: slight, moderate, extensive, and complete damage, for different structures based on the visible damages. FEMA-356[44] and FEMA-273[45] define 4 limit states: Operational, Immediate Occupancy, Life Safety, and Collapse Prevention, for different types of structures based on different types of engineering parameters such as maximum story drift and plastic hinge rotations, and also state the corresponding visible damages for each of those states. Several researchers use recommendations from these articles to classify damage and develop fragility curves[9], [20], [27] while several other authors define their limit states[21], [31].

When observational data from past earthquakes are used for fragility analysis, damage classification based on investigations can be used to classify responses[4]. Several times, however, schemes are required to convert qualitative data about damage in reports to quantitative measures. Lagomarsino and Giovinazzi[35] used fuzzy theory to build a damage probability matrix to accomplish this. Rossetto and Elnashai[21] developed a homogenous damage scale for RC buildings defining the damage index based on observed damage for various types of RC buildings.

2.7 Obtaining Responses

Once the suite of ground motions is obtained and the IMs and EDPs selected, the next step is to obtain the maximum responses of the structure. Different methods have been used to obtain the maximum response of structures to ground motions. These methods can be classified into analytical (physics-based), data-driven, and knowledge-based approaches.

2.7.1 Analytical

The most direct approach would be using analytical models to obtain the full time-history response of the structure to ground motions. Ideally, this analysis has to be repeated for a sufficient number of ground motions to incorporate the ground motion hazard of the selected region. Furthermore, to account for uncertainties in structural properties, a Monte Carlo simulation has to be conducted, obtaining responses for each combination of structural configuration and ground motion. Kwon and Elnashai[9] did exactly this to develop fragility curves for several classes of RC buildings.

In the absence of adequate ground motion data extending the range of intensities of expected ground motion hazard, a smaller set of ground motions can be selected and scaled to each intensity level. Then the scaled ground motions can be used to conduct full Monte Carlo simulations. This is called incremental dynamic analysis and is widely used in the scientific community[25].

However, conducting full time-history analysis of complex analytical models for these many configurations is computationally expensive. Several authors tackle this challenge by using time-history analysis of simplified analytical models for Monte Carlo simulation instead of complex models. These models include stick models[13], equivalent SDOF models[42], simplified and other physics-based surrogate models[46]. One commonly used method which uses equivalent SDOF models to obtain the maximum roof drift of a building is the displacement coefficient method, where the estimate for roof drift ratio obtained using an elastic SDOF system with the same fundamental frequency as that of the building is multiplied with the inelastic displacement ratio to obtain an estimate for the inelastic displacement ra-

tio. Lin and Miranda [47] instead used a first-mode elastoplastic SDOF system to estimate the inelastic roof drift directly.

Freeman[48] developed the capacity spectrum method to determine the performance of of structures to ground motions for the purpose of seismic design. In this procedure, pushover curves are first obtained by using static analysis by applying a lateral force distribution according to the fundamental mode shape on the complex analytical model to obtain the relation between the base shear force and roof displacement. This is then usually transformed to spectral acceleration-spectral displacement space by using an equivalent SDOF system with the same fundamental mode as the actual system, the dynamic properties for which are obtained from the characteristics of the pushover curve. The curve thus obtained is called the capacity spectrum. The intersection of the capacity spectrum with the response spectra is called the design point and is used as a measure of the performance of the structure. In this manner, the complex dynamic problem is converted to a simpler quasi-static problem. While this method was initially designed for design purposes, it has been extensively used for obtaining maximum responses for fragility analysis and is also a method that is suggested by HAZUS[6].Usually, authors validate the capacity spectrum method by comparing the results obtained to the responses obtained from full time-history analysis[14], [17]. Certain authors choose to use simplified analytical models instead of complex models for the pushover analysis further reducing the computational time[35].

Mosalam et al.[31] developed the adaptive pushover analysis method, where they update the shape of the static lateral load on the structure based on the changes in the fundamental mode shape as the structure starts to behave nonlinearly. In this way, the dynamic properties of the equivalent SDOF system obtained are modified to capture nonlinearities. Rossetto and Elnashai[28] demonstrated the use of a similar adaptive pushover analysis for the fragility analysis of RC structures.

Several authors have pointed out the limitation of the pushover analysis method. The pushover curve is obtained using the assumption that the fundamental mode dominates the response of the structure[31], which is not necessarily true for all types of structures. Chopra and Goel [49] tried to address this limitation by presenting an improved pushover analysis

method with evaluation of peak responses at multiple modes of vibration and then combining the results to get the estimate for the peak response.

In some studies, authors avoid having to conduct a full Monte Carlo simulation by sampling points selectively to cover the entire parametric space without having to run the analysis for every combination of structure and ground motion. Latin Hypercube Sampling (LHS) is a commonly used technique by many authors[11], [26], [34] to sample points uniformly throughout the entire space. Other sampling techniques such as the uniform design method have also been used by researchers[19].

Ultimately, the values of maximum response obtained, and thus the fragility calculations, vary based on the type of method used. Nielson and DesRoches[26] compared fragility results obtained for bridges using nonlinear time-history analysis to results obtained using capacity spectrum analysis and found significant differences between the two methods. They explained the differences by pointing out the differences in modeling techniques and inventory used. Park et al.[50] conducted a fragility analysis of low-rise unreinforced masonry buildings using three different models with different levels of out-of-plane wall stiffness considered and found that the seismic response was very sensitive to the level of out-of-plane wall stiffness considered. They also noticed variability in responses they obtained compared to the results obtained using recommendations from HAZUS[6] and used similar reasons to justify the differences.

2.8 Knowledge-Based

Knowledge-based methods for the generation of fragility curves rely on observed damage during past earthquakes and empirical equations to develop fragility curves.

Using observed damage for fragility analysis is expected to be the method since it is based on observed damage as opposed to a modeled system. Shinozuka et al.[4] used maximum likelihood analysis of damage data of 770 single support RC viaducts columns with similar geometry and reinforcements following the Kobe earthquake in 1995 to develop empirical fragility curves.

Lagomarsino and Giovinazzi[35] developed fragility curves using empirical equations as a function of the vulnerability index (V), and ductility index (Q), the values for which depend on the building type and construction. They accounted for randomness in structural properties by considering a range of vulnerability index values.

One difficulty of using knowledge-based methods is that post-earthquake damage observed in one region is not necessarily applicable in other regions due to the differences in ground motion hazards and structural types and practices. Rossetto and Elnashai[21] used damage data from 340,000 structures from 29 surveys covering 19 earthquakes for risk assessment of European RC structures. In their study, the non-European fragility curves can be used for the fragility analysis of similar European structures. However, they noted that this finding was based on a small amount of data and recommended not to apply such techniques. Straub and Der Kiureghian[51] proposed a method to include statistical dependencies among different observations to account for the uncertainty in the fragility of equipment in electric substations.

The second type of knowledge-based method is using empirical equations to obtain fragility functions. Mahsuli and Haukaas[52], [53] developed fragility curves to assess the seismic risk to the structures in Vancouver using different empirical equations.

2.8.1 Data-Driven Methods

Guan et al.[54] developed ML-based data-driven models using several ML algorithms and the spectral accelerations and displacements at different modal periods as inputs. Kiani et al.[55] used classification algorithms to classify the response of an 8-story SMRF building into two classes based on the maximum story drift ratios and developed fragility curves using the trained algorithms. Similarly, Wang et al.[56] used artificial neural networks for the fragility analysis of a single structure and obtained a R^2 value of 0.93 and 0.96 in training and testing using 80 training data points. Both of these studies did not consider uncertainties in structural properties. Mitropoulou and Papadrakakis[57] used neural networks to obtain the seismic demands due to ground motions on selected RC buildings.

Gidaris et al.[58] developed a surrogate framework using kriging for the prediction of maximum story drift and floor acceleration. They obtained models with R^2 as high as 0.96 and errors within 15% for the prediction of maximum story drifts. Micheli et al.[59] presented a multiple surrogate modeling framework using different kriging-based surrogate models for the prediction of maximum drifts and accelerations of each floor for wind loads. The most commonly used machine learning methods among these papers are artificial neural networks, support vector machines, Gaussian process regression (kriging), and linear regression[2].

Esteghamati and Flint [60] used several machine learning techniques for the prediction of seismic loss with R^2 values of up to 0.96 while varying the topology of the building considered. Ghosh et al.[61] used different classification algorithms to predict the seismic response of bridge components and obtained a R^2 value of 0.70 or higher for each type of model for different components of bridges to develop fragility curves.

Some authors also used neural networks to predict the entire time-history response of structures to ground motions. Huang et al.[62] used a deep neural network to predict the modal properties and the time-history response of a particular structure to various intensities of the ground motions from the Kobe earthquake. Oh et al.[63] used convolutional neural networks to predict the displacement response of a benchmark structure under different ground motions. Zhang et al.[64] used physics-guided convolutional neural networks using information from the equation of motion to build the loss function, to obtain the displacement time-history response of a 6-story hotel for different ground motions.

2.8.2 Comparisons of Different Types

To the best of my knowledge, few studies have been done to compare different types of modeling approaches used to obtain the values of engineering demand parameters. One study which does this comparison is conducted by Esteghamati and Flint[5] who make this comparison for the prediction of hazard performance measures for mid-rise concrete frame buildings. They used similar structures from INSSEPT database[65] for developing their knowledge-based surrogate models, support vector machines for data-driven surrogate models, and the capacity spectrum method for analytical models. They found that physics-based

models captured the trend in a better manner than data-driven and knowledge-based models. Furthermore, they found that physics-based models performed well when predicting the responses of structural configurations outside the training data, while the prediction of loss using the data-driven surrogate had higher errors. They recommended a technique to sequentially use the three different types of models for design, starting with the knowledge-based model to make initial decisions regarding the structural system to narrow down the possible options, followed by data-driven models to get bounds for obtaining constraints for the physics-based surrogate, which is finally used to obtain the optimum design.

2.9 Fragility Curve

Once the damage states are obtained for different structure-ground motion combinations, these can be converted to the probability of damage states in various ways. The most commonly used way is to use the ratio of structural configurations exhibiting a particular damage state at a given value of intensity measure to the total number of structural configurations analyzed as the probability of the structure exhibiting that particular damage state at the corresponding intensity level. Alternatively, instead of quantifying the EDPs, the fragility of structures can be directly quantified by using empirical equations.

A set of the probability of damage state-IM pairs is obtained for different damage states. These are plotted on the graph and a function is used to interpolate the values in between. Usually, it is assumed that a lognormal function is a good fit to determine the relationship between the probability of exhibiting a damaged state and the ground motion intensity measure. Buratti et al. performed three different statistical tests to demonstrate that the assumption of drift values being distributed log-normally is allowed[66]. The parameters of the best-fitting lognormal curve are obtained by regressing the probability values against the intensity measures. Many authors, however, use the parameters recommended by HAZUS[6] based on expert judgment to fit lognormal curves for fragility.

2.10 Component Level Fragility Curves

Fragility curves can also be obtained separately for each component or the system as a whole. This is more often done for bridges than other kinds of structures, owing to the differences in different bridge components. When component-level fragility curves are obtained, they need to be combined using appropriate techniques to obtain the system-level fragility curve.

Lang and Bachmann[67] used a simple rule assuming that the capacity of a a building with structural walls system is the sum of the capacities of individual walls.

Lupoi et al.[29] used a system function to determine the state of a highway viaduct system in terms of its components. Nielson and DesRoches[43] used component level fragility curves to determine the most critical components. They used the correlation between the probabilistic seismic demand models of different components to develop the system-level fragility curves for the bridges they considered. They found that considering just a single bridge component to depict the fragility of the entire bridge significantly underestimated the vulnerability. They also found that the bounds found using this method are narrower when a single component is significantly more fragile than the remaining components. In another study, Nielson and DesRoches[26] followed a similar approach to obtain system-level fragility curves and compared the results to results obtained using guidelines from HAZUS[6], which does not include the effects of abutments on the overall fragility of a bridge, and found that there was a considerable difference between the results.

Choi et al.[27] took this one step further by modeling the fragility of subcomponents and using the curve from the most vulnerable subcomponent to get the fragility of each component of the bridges they selected. They then used first-order reliability bounds to obtain the upper and lower bounds of the system-level fragility curves from component-level fragility curves assuming that bridge components are aligned in series, with the failure of one component causing the failure of the entire system. Pang et al.[19] also made use of first-order reliability to demonstrate that the overall fragility of a bridge is larger than the fragility of any single component. Pan et al.[24] demonstrated that using second-order

reliability to combine component fragility curves provided produced narrower bounds than first-order reliability.

2.11 Applications

While the main application of fragility curves is to quantify the risk to a structure or a set of structures to ground motion hazards, these have been used for other applications as well. Fragility curves can be used to determine the component of a structure that is most fragile[26].

Developed fragility curves can be converted to estimates of different types of losses (economic and casualties) to structures following an earthquake. FEMA[68] developed damage functions to convert the probabilities of failure to losses for various types of buildings.

3. PROBLEM DESCRIPTION

It is important to properly define the scope of the study at the start. This chapter sets up the study that we undertake in this thesis document and lists the set of questions that we answer in the following chapters.

3.1 Overview

The main objective of the study is to develop fragility curves by using different surrogate models to estimate the maximum responses to ground motions. A particular building is selected for this study from past work by Tabatabaiefar et al.[69]. The authors of the paper developed a computational model using OpenSees and validated the model using scaled-down shake table tests for a 15-story steel moment resisting frame building. Section 3.2 provides a detailed description of the selected building.

In this thesis, both the properties of the building as well as the ground motion data are varied. Surrogate models that are to be developed need to be able to make predictions across the range of these parameters. The surrogate models thus developed can be used for studies such as parametric analysis, optimization studies and conceptual design, and fragility analysis while considering uncertainties in the structure. More information about varying the structural properties and ground motion data are provided in Section 3.3 of the chapter. Finally, a uniform criterion needs to be fixed to evaluate the performance of each of the developed surrogate models. Section 3.6 describes the performance metrics used to evaluate and compare different surrogate models that are built in this study. The study intends to answer a set of questions. Some of these questions are listed in Section 3.7.

3.2 Building Model Selected

A 15-story regular moment resisting frame building is selected from literature[69]. The authors have constructed the scaled-down structure made of steel plates and have conducted shake table tests on the structure with amplitudes from four well-known ground motions (Kobe earthquake 1995, Northridge earthquake 1994, El Centro earthquake 1940, and Hachi-

nohe earthquake 1968). They scaled the structure to be able to conduct shake table tests using available facilities. They have provided the material and geometric properties of all the components used in their study, as well as the time history and peak responses obtained from the shake table tests.

These results are used in this thesis to build and validate an Abaqus model to obtain the high-fidelity data required to train and test the surrogate models. The paper by Tabatabaiefar et al.[69] shall be referred to as the "reference paper" from here onward in this thesis. The study in this thesis has been conducted on the scaled-down building provided in the reference paper since the scale of the building in question does not affect the performance of surrogate models generated in this study. The ground motion data has also been scaled accordingly to get results that would be consistent with the responses obtained when applying the unscaled ground motion to the unscaled building. Since the objective of this thesis is just to develop and compare various surrogate models, scaling down the building under consideration does not affect the results obtained. The following parts of this section provide the details of the scaled building as described in the reference paper.

3.2.1 Geometry

The building is a 15-story regular moment resisting frame building with a single bay and equal story heights throughout. The geometric scaling factor used by the authors of the paper is 1:30. Table 3.1 is a copy of Table 1 from the reference paper, providing the details regarding how different geometric properties scale with the scaling ratio λ.

Table 3.1. Scaling of different properties in terms of the geometric scaling factor (λ)

Property	**Scaling**	**Property**	**Scaling**	**Property**	**Scaling**
Mass Density	1	Acceleration	1	Length	λ
Force	λ^3	Shear Wave Velocity	$\lambda^{1/2}$	Stress	λ
Stiffness	λ^2	Time	$\lambda^{1/2}$	Strain	1
Modulus	λ	Frequency	$\lambda^{-1/2}$	EI	λ^5

Source: Tabatabaiefar et al.[69]

The height of each story is scaled down to 0.1 m, resulting in the total height of the structure of 1.5 m. The base plate is designed as a square plate with a side length of 500 mm and thickness of 10 mm. Each floor slab is designed in the form of identical square plates with a side length of 400 mm and thickness of 5 mm. 4 identical rectangular plates with a width of 40 mm and thickness of 2 mm are used as columns. All these dimensions are provided in Table 3.2. Figure 3.1 is the copy of the figure of the shake table specimen used in the reference paper.

Table 3.2. Dimensions of building elements

Element	**Length**(mm)	**Width**(mm)	**Thickness**(mm)
Base Plate	500	500	10
Floor Slabs	400	400	5
Column Plates	100[1]	40	2

3.2.2 Material Properties

The authors of the reference paper indicate that they designed the scaled-down prototype by using steel plate grade 250 of AS/NZS 3678-2011, with a minimum yield stress of 280 MPa and a minimum tensile strength of 410 MPa to build each of the parts of the building frame. The Young's modulus of the selected steel plate material is specified as 200 GPa. The authors of the reference paper specified a density of 7850 kg/m^3 in their numerical model. They also estimated a damping ratio of 1.1% for the building using free vibration tests.

3.3 Variation of Building Properties

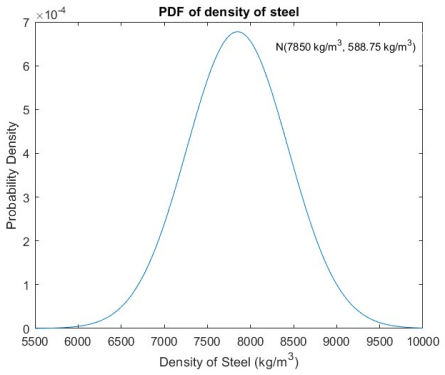
Section 3.2 described the base building selected for the study. In this section, we describe how the variations in building properties that are considered in this study when developing surrogate models. The building properties can be varied in a variety of ways; the geometry of the structure can be varied, or the properties of the material can be varied. Irregularities can also be introduced along the horizontal or vertical dimensions. It is outside the scope of

[1]The length here refers to the height of column per story

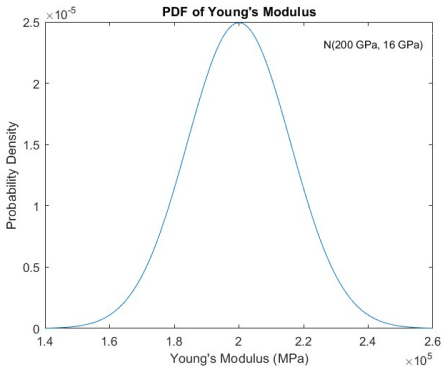
Figure 3.1. Shake Table Specimen Used by Tabatabaiefar et al.[69]

this study to incorporate each of these methods in the study. Therefore, for this study, we vary the building properties by just changing the material properties of the entire structure without changing the geometry or adding any irregularity along any dimension.

The two fundamental dynamic properties of a structure are its mass and its stiffness. Both these properties are varied by changing the material properties. The mass of the structure is changed by varying the density of the building material, while the stiffness of the structure is changed by varying Young's modulus of the building material. Both the density and Young's modulus of steel are modeled as normally distributed random variables,



(a) PDF for the Density of Steel



(b) PDF for Young's Modulus of Steel

Figure 3.2. Probability Distributions of Structural Parameters Modeled as Random Variables

staying consistent with past research, with the mean value being the deterministic value assumed in the reference paper and the variations selected from González and Zapico[70]. The values provided are assumptions made by the authors of the aforementioned paper and other works in the past have assumed different values. González and Zapico[70] assumed a coefficient of variation of 7.5% for the density of steel and coefficients of variation of 5% and 11.5% for stiffness of columns and beams respectively. For this study, we have assumed the same value for the coefficient of variation of density of steel and assumed that the coefficient of variation for Young's modulus of steel is 8% (about the average of the values provided by González and Zapico[70]). The means, coefficients of variation, and the standard deviation assumed are listed in Table 3.3, and the corresponding probability distributions are plotted in Figures 5.3a and 5.3b.

As described later in Chapter 2, different researchers have previously varied properties such as the yield strength of steel, and damping ratios. However, they found that changes in these parameters do not affect the response of a structure (or the fragility of the structure) to ground motions as do the stiffness and the mass of the structure[29]. Hence, in this study, just the density and Young's modulus of steel are considered to be random variables and all other structural properties are assumed to be deterministic. It is assumed that there is zero correlation between the density and Young's Modulus of steel. This follows what

is usually assumed in literature when the exact correlation between two random variables is not known[29] The values of all other structural properties are kept the same as in the reference paper and as listed in Section 3.2.2.

It is important to note that the geometry of the building is not varied in this study, although changing the geometry has significant effects on the response of the structure. Similarly, irregularity is not added to the building in any form.

Table 3.3. Parameters of Distribution of Density and Young's Modulus of Steel

Quantity	Mean	Coefficient of Variation	Standard Deviation
Density	$7850\,\mathrm{kg}/m^3$	7.5%[70]	$588.75\,\mathrm{kg}/m^3$
Young's Modulus	200 GPa	8%	16 GPa

Selecting a distribution as shown above helps sample the points accordingly when generating high-fidelity data and also helps to obtain regression parameters when conducting fragility analysis. The surrogate models we are developing should have the capability to accurately predict the responses for configurations of buildings across the range of structural properties defined above.

3.4 Ground Motion Data

Ground motion data for this thesis is obtained from the following ground motion databases: Pacific Earthquake Engineering Research Center (PEER) Ground Motion Database [71]–[73], and Center for Engineering Strong Motion Data (CESMD)[74]. Ground motions are selected randomly to reflect the distribution of ground motions in real life. Due to this random selection, similar to the real world, the number of ground motions selected at a given magnitude is inversely proportional to the magnitude. Although the distribution of the selected ground motion reflects the distribution observed in real, this could be a potential source of bias in the data as a majority of the ground motions are concentrated on the lower intensity side. Ground motions with magnitudes less than 3 are left out because there is almost no data recorded with such low magnitudes in databases and because these do not cause appreciable responses in buildings. Historically important ground motions and ground motions

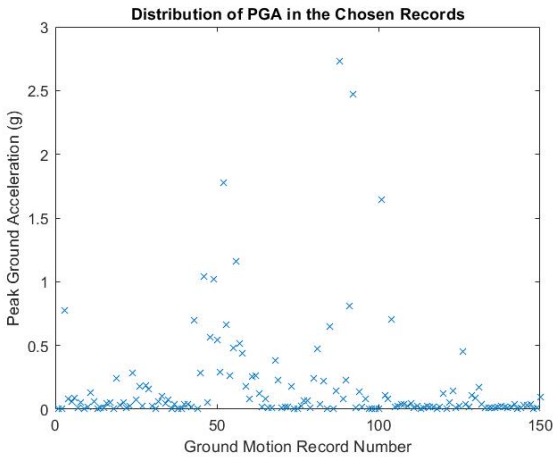


Figure 3.3. Distribution of Peak Ground Acceleration Among Selected Ground Motion Records

with unique characteristics such as high peak ground accelerations, are consciously included among the selected earthquakes. Initially, a set of 125 ground motions records is downloaded from the two databases. Later another set of 25 ground motions is added to the list. Figure 3.3 shows the distribution of peak ground accelerations for ground motions selected. The time steps of the selected ground motions are multiplied by the square root of the scaling factor ($\lambda^{1/2} = \frac{1}{5.48}$) to stay consistent with the geometric scaling used for the building dimensions.

3.5 Response Parameters

This section of the chapter provides details regarding the parameters of interest that need to be predicted by the surrogate models in this study. Three response parameters are chosen.

First among the three, is the fundamental frequency of the structure. This is not a response variable as it is an intrinsic property of the structure and does not vary with the ground motion input. The only factors that affect the value of the fundamental frequency of the structure are the structural properties: density and Young's modulus of steel. Determining the natural frequencies of a structure is important because the way a structure

behaves under ground motion excitations depends on these frequencies. In a lot of cases, the fundamental frequency of the structure dominates the response. This is why, in the structural engineering community, many different characteristics of a system such as its spectral acceleration, velocity, and displacement are defined using the fundamental frequency of the system. The fundamental frequency is also used to build equivalent single-degree-of-freedom or discrete multi-degree-of-freedom systems that are used to approximately determine the response of the system. In this study, the surrogates developed for the prediction of the fundamental frequency of the system are then used for both of the above purposes when building surrogates for the remaining two response parameters. Chapter 5 provides more details about the methods used to develop surrogate models for predicting the fundamental frequency of the structure with varying properties.

The next two response parameters are the maximum roof drift ratio and the maximum story drift ratio. These two parameters vary with both the ground motion characteristics as well as the structural properties. Structural damage a structure undergoes during a seismic event is dependent on the amount of drift of the structure during the said event. Even a part of the nonstructural elements of a structure is sensitive to the amount of drift caused during ground motions. Therefore the focus when designing structures for seismic excitation has moved towards limiting drifts in the recent past[1]. For the same reasons, in a lot of studies, the seismic fragilities, vulnerabilities, and risks of financial losses to a structure or a region are computed by comparing the amount of drift demand imposed by expected ground motions in the region to the drift capacities of the structures under consideration. Therefore, these drifts are important measures to determine during the design as well as risk assessment of buildings. More information regarding the development of surrogate models for the prediction of the maximum roof drift ratio and the maximum story drift ratio is provided in Chapter 6.

3.6 Performance Metrics

Various statistical values have been used to determine the performance of surrogate models. Since predicting the maximum responses of a structure to seismic ground motions

is effectively a regression problem, any performance metrics that are traditionally used for regression problems can be used here. The two performance metrics selected in this study are the Normalized Root Mean Squared Error (NRMSE) and the R^2 value (coefficient of determination).

The RMSE value provides an average measure of the error in prediction by averaging the squares of differences between the predicted values to the actual values and taking the square root. In other words, the RMSE provides the standard deviations of the residuals of the prediction and therefore measures the spread of the error. Therefore, the lower the value of RMSE, the more accurate the model.

$$RMSE = \sqrt{(\sum_{\mathrm{i}=1}^{N} \frac{(y_{\mathrm{i}} - \hat{y}_{\mathrm{i}})^2}{N}} \tag{3.1}$$

The RMSE value can be further normalized in various ways. Two of the most common ways used to normalize the RMSE values are dividing by the range of the actual data and dividing by the mean of the actual data. Here, the RMSE is normalized by dividing by the mean value of responses in the training set to obtain the NRMSE. Normalizing the RMSE value in this manner helps us to compare the performance of surrogates trained for the prediction of different types of data and with different magnitudes of data. This type of normalization could also be thought of as the coefficient of variation of the residuals.

$$NRMSE = \frac{1}{\bar{y}} \sqrt{(\sum_{\mathrm{i}=1}^{N} \frac{(y_{\mathrm{i}} - \hat{y}_{\mathrm{i}})^2}{N}} \tag{3.2}$$

where $\bar{y}$ is the mean of actual response values, $\hat{y}$ are the predicted values.

It is difficult to determine what a good value of NRMSE is. The determination of what value is good varies based on the application and usually involves judgment. In this study, an NRMSE value of 0.5 is targeted for the developed surrogate models, both in validation and testing. This value is selected from previous experience and is usually considered a good value in most fields and matches what has been targeted in previous studies[58], [60].

The R^2 value or the coefficient of determination measures the goodness of fit of a model in terms of how well the predictor variables capture the variations in the outputs. The R^2

value can also be thought of as the factor by which the variance in the prediction errors is less than the variance in the response.

$$R^2 = 1 - \frac{Sum\ of\ Squares\ of\ Residuals}{Total\ Sum\ of\ Squares} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{3.3}$$

The R^2 value is always between 0 and 1. An R^2 value of 1 means that the model captures all the variations in the output data. Meanwhile, an R^2 value of 0 that no variation in the output is captured by the input parameters of the model, or in other words, the output predicted is the same irrespective of the actual value. Again, deciding a good value of R^2 is a difficult task and involves judgment based on the problem at hand and previous experience. In this study, an R^2 value of 0.95 (which means that 95% of the variation in the response parameters or equivalently 78% of the standard deviation of the response parameter is captured by the predictor parameters) is chosen as the target for the surrogate models in both validation and testing.

The RMSE and the R^2 values together are usually considered sufficient to determine the performance of different models although this varies based on the criticality of application of results. In this study, these two performance metrics are computed for both validation and testing. For validation, k-fold cross-validation[75] is used. In k-fold cross-validation, the training dataset is divided into k subsets. In each turn, 1 of the subsets is kept for validation while the model is trained on the remaining k-1 subsets. This process is repeated k times until each of the subsets is used for validation for the model trained on the remaining subsets. Although this increases the time required for model training and validation, it is a more effective way to use all the training data and also reduces overfitting. Also, as opposed to holdout validation, in k-fold cross-validation, the entire dataset is used in training the model and it also gives a better idea regarding how the model generalizes on unseen datasets.

3.7 Research Questions

This section lists a subset of the questions that this study intends to answer:

1. What are the different ground motion parameters that capture the ground motion characteristics well enough to predict a particular response parameter? In other words, which ground motion properties correlate best with response parameters, so that they can be used as input parameters into surrogate models to predict these responses? Also, what is the right choice of ground motion intensity measure to condition the engineering demand parameter when developing fragility curves?

2. How many training data points are required to train different types of surrogate models to predict different response parameters accurately when not varying the properties of the building?

3. How does the number of training points increase as the properties of the building under study are also changed along with the ground motion data?

4. How does the performance of each type of surrogate model compare against each other?

5. What is the reduction in computational cost when surrogate models are used for various applications as compared to the use of complex nonlinear computational models?

6. Is the reduction in computational time significant enough to justify the error associated with using surrogate models instead of high-fidelity nonlinear computational models?

7. Are the fragility curves computed using surrogate models accurate enough to warrant the reduction in computational time?

4. HIGH-FIDELITY DATA

Before we start developing surrogate models, we have to develop a means to collect or generate high-fidelity data that meets the needs of generating a surrogate model required for a given task. High-fidelity data is required to train surrogate models and test the performance of the developed surrogate models by comparing them using the high-fidelity data (considered the ground truth). Several approaches are available to obtain high-fidelity data depending on the problem under consideration. In certain cases, high-fidelity data might be easily available, while in certain other cases, high-fidelity data might be scarce and difficult to generate because of the lack of adequate knowledge regarding the problem.

Three of the widely used approaches to obtain high-fidelity data that can act as ground truth are experimental, analytical/numerical, and observational/empirical approaches. Experimental approaches are used when it is feasible to develop physical experimental models that can closely resemble the real system. While experimental results are usually considered to be more accurate, as they resemble the actual system and capture the physics behind the actual system more accurately, it is not always practical to build experimental models. Furthermore, experiments might have their own set of errors because of inaccuracies in the setup and simplifications from the real system.

Analytical and numerical methods are useful to generate data in such cases where the experimental study becomes infeasible. This would however require an adequate understanding of the system to capture the physics of the real system in a set of mathematical equations. Therefore, there is always a level of simplification associated when using analytical and numerical methods (in other words, there is always a certain level of epistemic uncertainty built into data generated by analytical and numerical methods). The use of analytical and numerical methods becomes most advantageous when a lot of data is to be generated and it is not practical to build an experimental model for each of these configurations.

In certain cases, observational and empirical data could be available for the problem at hand. The availability of such data helps reduce the amount of time to obtain high-fidelity data. In recent years, researchers have started pooling their resources together to develop databases that act as sources of data for various types of problems. Benchmark problems and

competitions have encouraged researchers to develop and document standardized data that can be used by other researchers. However, the biggest restriction to such data is that the observational and empirical data is usually limited or incomplete. Furthermore, researchers documenting observational data usually use different methods to gather and document the observational data, making it difficult to use multiple sources together. Even when databases with high-quality data are available, these data are relevant for certain very specific problems only and cannot be generalized easily to other cases.

To test the behavior of structures under seismic ground motions, the experimental method of generating data is to build physical models of the building and conduct shake table tests by applying the ground motion data to the physical model using actuators. Several researchers scale down the building using certain laws, like done in the reference paper, to make it more practical to build the physical specimen and test on available shake table facilities. This however becomes impractical when the responses from many different configurations of a building are required as is the case in this study. Observational methods of obtaining high-fidelity data include using results from surveys and investigations conducted after previous seismic events. However, as mentioned before, observational data are very specific and the methodologies used to obtain and record these data vary from person to person. Empirical equations available in design standards or from other past literary works can be used to generate high-fidelity data, but these too are usually simplifications and apply to very specific cases. Numerical models can be developed and solved using finite element methods to obtain the response of a building to ground motions computationally. Many different commercial and open-source finite element software are available to conduct this analysis.

In this thesis study, computational and numerical models are used to generate high-fidelity data used to train and test the surrogate models. This chapter describes the methods that are used to generate high-fidelity data in this thesis, along with the associated assumptions and levels of complexity.

4.1 Abaqus Models

Abaqus FEA software[76] is used in this thesis to develop high-fidelity data. Abaqus is a versatile commercially available finite element software with various structural analysis functionalities that help to create computational models with various levels of complexity. Furthermore, Abaqus generates input files before solving the system of equations to obtain the response of the system. These input files are easy to understand and manipulate, thus allowing the use of simple codes to run the Abaqus models using different values of parameters to generate a whole set of data. This section provides the details of the different Abaqus models created to generate high-fidelity data for the studies conducted as a part of this thesis.

Different types of Abaqus models are created based on the level of complexity required for particular studies mentioned in the following chapters. Initially, a nonlinear 3D model is built in Abaqus and validated using the results from the reference paper. However, running the 3D model proved computationally expensive, requiring 2-4 days to generate data for a single sample point. Using this model would require tens of weeks to generate enough data to develop acceptable surrogate models.

Therefore, using the assumption that ground acceleration is happening along a single horizontal dimension, a nonlinear 2D Abaqus model is developed and validated. This simplification reduces the number of degrees of freedom at which the equation of motion has to be solved, without introducing any errors in the response along the direction of interest. The time for generating data using the 2D model is about 2-4 hours per sample point, which results in a big decrease in the time required to generate high-fidelity data.

A linearized 2D model is also created using the modal dynamics option in Abaqus to generate pseudo-high-fidelity data for preliminary studies where generating data quickly is more important than capturing the complexities of the problem. These preliminary studies are used to inform the generation of actual high-fidelity data using the nonlinear 2D model. This also allows us to understand how the nonlinearities affect the training of surrogate models. The linearized 2D model takes about 1-2 minutes per data point.

The following subsections provide a detailed description of the steps used to create these models. Enough details are provided to allow any future researcher to reproduce the models if required. The subsections follow the different tabs that are provided in Abaqus for modeling.

4.1.1 Geometry

This subsection describes the geometry created for each of the Abaqus models.

3D Model

For the 3D model, each element of the building is modeled as a 3D deformable element[76] in Abaqus. Perimeter beams are included in the frame, deviating from the actual building prototype, to make it easier to define contact between different elements accurately. However, the lateral dimensions of the beams are kept negligibly small (as shown in Section 4.1.3 below) to ensure that adding these beams to the frame does not change the responses significantly. Beams and columns are modeled using deformable 3D wire elements[76], while the base plate and floor plates are modeled as deformable 3D shell elements[76]. These choices are justified because beams and columns in moment-resisting frames act primarily as bending/axial members with the dimension along the length considerably larger than the other two dimensions and floor slabs act as membranes with 2 dimensions considerably larger than the dimension along the thickness. Using this simplification allows a reduction in the number of degrees of freedom without introducing significant errors compared to modeling all the members as solid elements. Dimensions for these elements are selected to be consistent with Table 3.2 (the length of beams is assumed to be the same as the length of floor slabs).

2D Model - Linear and Nonlinear

For the 2D model, each element of the building is modeled as a 2D planar deformable element[76] in Abaqus. Here unlike the 3D model, beams are not added, as it is easy to provide accurate contact without requiring the addition of beams. Columns are modeled using deformable 2D planar wire elements[76]. However, unlike in the 3D model, the base and floor slabs are also modeled using deformable 2D planar wire elements here and the

lateral dimensions are provided using the sections module in Abaqus. These choices are justified because columns act as axial members with the dimension along the axial dimension significantly longer than the lateral dimensions and since we are assuming that the ground motion happens in one horizontal direction only and are interested in the response along only this direction, we can simplify the slabs along the second horizontal dimension without causing any error. These simplifications significantly reduce the computational cost required to obtain data.

4.1.2 Material Properties

This subsection describes the material properties assumed for different elements of the building. The same material properties are used in all of the models used in this study. Steel is modeled as elastic-perfectly plastic to be consistent with the approach used in the reference paper. This approach is also followed in the linearized model since including material nonlinearities does not increase the computational cost significantly when using modal analysis. Therefore, the linearized model is not entirely linearized as it still has material nonlinearities (but it is linearized in the aspect that geometric nonlinearities are ignored and the responses of the structure are assumed to be a linear superposition of the responses of the natural modes of the structure).

The Poisson's ratio of steel is assumed to be the average value of 0.28 as it is not provided in the reference paper. Material damping is included to obtain a damping ratio of 1.1% for the building as in the reference paper. Rayleigh damping is used to achieve this and the values of α (coefficient for mass proportional damping) and β (coefficient for stiffness proportional damping) are selected based on trial and error while ensuring the individual damping term adds up to a damping ratio of 1.1% for the fundamental mode of response of the structure. It is observed that results closest to those provided in the reference paper are obtained when the entire material damping is assumed to be mass proportional and the portion of damping that is stiffness proportional is assumed to be 0. This however means that the damping for higher modes of responses is lesser, because their modal masses are smaller. Therefore, even though using just mass-proportional damping results in lower amount of difference with

respect to the results from the reference paper, it also means that the contribution of higher modes to the overall response of the structure is increased.

The Young's Modulus and density of steel are varied in the study as needed based on the sample point being considered (as described in Section 3.3). For validation of the Abaqus model, Young's Modulus value of 200 000 MPa and density of 7.85×10^{-9} are selected to be consistent with the values used in the reference paper. The material properties described in this subsection are summed up in Table 4.1 below.

Table 4.1. Material Properties of Steel for the Structure

E (MPa)[1]	ν	$\sigma_{\mathbf{y}}$ (MPa)	ρ(t/mm^3)[1]	α (s^{-1})
200000	0.28	280	7.85 $\times 10^{-9}$	0.3075

4.1.3 Sections

This subsection describes the section properties provided for different elements in the Abaqus models.

3D Model

Two types of sections are used in the 3D model. Homogeneous shell elements with appropriate values of shell thickness (from Table 3.2) are assigned to the base and floor slabs. Beam sections with appropriate profiles (consistent with dimensions in Table 3.2) are assigned to the columns members. Beam sections with width and depth of 0.01 mm are assigned to the beam members added in the model. These values are smaller than all the dimensions of all other elements by quite a few orders of magnitude and therefore adding these beams to the building model does not impact the response of the building significantly.

[1]These values are varied in the studies conducted in the thesis document as mentioned in Section 3.3

2D Model - Linear and Nonlinearized

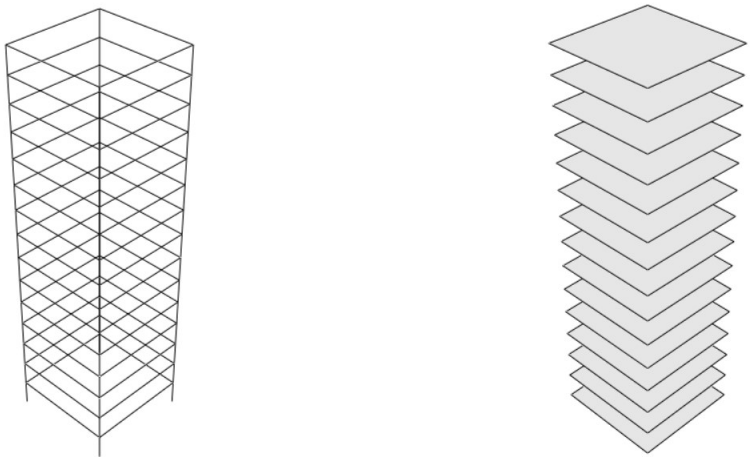
Beam sections are used for all the elements in the 2D model. Section dimensions are selected to be consistent with 3.2. Even the base and floor slabs are modeled using beam elements with appropriate width and thickness as section dimensions.

4.1.4 Assembly

This subsection describes the approach used to assemble the parts into the building model in Abaqus. The procedure used below is one of many equivalent ways in which this assembly can be obtained.

3D Model

First, the columns and beams are assembled to form the frame of the structure. This is merged to form a new part that resembles the entire frame. Similarly, all the floor slabs are assembled, spaced at a distance equal to the height of each story between each, and merged to form a part that resembles just the entire set of floors without the frame. This is shown in Figure 4.1. Finally, the frame, the set of slabs, and the base plate are all assembled by using the various tools available in Abaqus.



(a) Frame (b) Slabs

Figure 4.1. Merged Parts for 3D Model

2D Model - Linear and Nonlinear

First, the columns and the slabs (modeled as beams as described in 4.1.1) are assembled to form the frame of the structure. This is merged to form a new part that resembles the entire frame of the structure, with the slab included. Next, the frame and the base plate are assembled using the various tools available in Abaqus.

4.1.5 Step

This subsection describes the step option selected in the step module in Abaqus created for each of the Abaqus models.

3D Model and 2D Model - Nonlinear

Four steps are used to generate all of the data for the two nonlinear models. The first step is the default initial step in Abaqus. A frequency step using the Lanczos solver[76] is selected after the initial step to obtain the natural frequencies of the models. This step is chosen from the linear perturbation steps[76] and carries out eigenvalue analysis to calculate the natural frequencies of the structure. The third step is a static, general step[76] that is used to add self-weight to the structure. Adding the self-weight helps condition the problem better when applying the ground motion accelerations onto the structure. The self-weight added in this step is carried forward into the next step, which is a dynamic, implicit step[76]. The Nlgeom option[76] is kept on for both the third and fourth steps to allow the model to account for geometric nonlinearities. The fourth step is used to apply the ground motion loads onto the structure and obtain its dynamic time-history response. The increment size is modified to match the sampling rate of the ground motions being applied and the total time is selected to match the end time of the ground motion acceleration. Since all of the ground motion data selected have peaks well before the end of the ground motion data, the peak response of the building will be captured without running the analysis for times after the ground motion data ends. All other step options are kept at the default values.

2D Model - Linear

The first two steps for the linearized modal model are the same as the first two steps of the nonlinear models. The default initial step is followed by the frequency linear perturbation step[76] with the Lanczos solver. The third step in the linear model is the modal dynamics step[76] from the linear perturbation sets of steps. This step allows for the application of ground acceleration loads to the structure and like the third and fourth steps in the nonlinear models, the time step for this step is modified to match the sampling rate of the ground motion being applied and the end time is changed to match the end time of the ground motion data.

4.1.6 Interaction

This subsection describes the various contacts and connections provided between different parts of the structure.

3D Model

Since the frame elements are merged into a single part, the joints in the frame behave as a tie contact. Therefore, at joints, the columns and beams have the same displacement.

Tie constraints with node-to-surface discretization[76] are defined at the intersections of the columns with the base plate. The top surface of the base plate is selected as the master surface and the nodes at the bottom of the columns are selected as the slave surface.

Similarly, the contact between the beams (which were intentionally added to define the contact more easily) and the slabs are modeled using tie constraints with surface-to-surface discretization with the top surface of the column and the perimeter of the beam as the master surface and the floor slabs as the slave surface. All other properties of the tie constraints are kept default and the tie constraint is maintained through all of the analysis steps.

Using tie constraints for these contacts implies that the displacements of interacting pairs are the same at the point of interaction. This means that all the joints are perfect and there is no slip occurring at any of the joints. This assumption, while not always true in real-case

scenarios, helps improve the conditionality of the problem. This approach is consistent with the approach used in the reference paper and a more detailed study that allows for slip at joints is out of the scope of this study.

2D Model - Nonlinear and Linear

Since the frame elements (the column and the floor slabs) are merged into a single part, the joints between these elements behave as a tie contact.

Tie constraints with node-to-surface discretization are defined at the intersections of the columns with the base plate. The top surface of the base plate is selected as the master surface and the nodes at the bottom of the columns are selected as the slave surface. All other properties of the tie constraints are kept default and the tie constraint is maintained through all of the analysis steps.

Once again, using tie constraints for these contacts implies that the displacements of interacting pairs are the same at the point of interaction. This means that all the joints are perfect and there is no slip occurring at any of the joints. This assumption, while not always true in real-case scenarios, helps improve the conditionality of the problem. This approach is consistent with the approach used in the reference paper and a more detailed study that allows for slip at joints is out of the scope of this study.

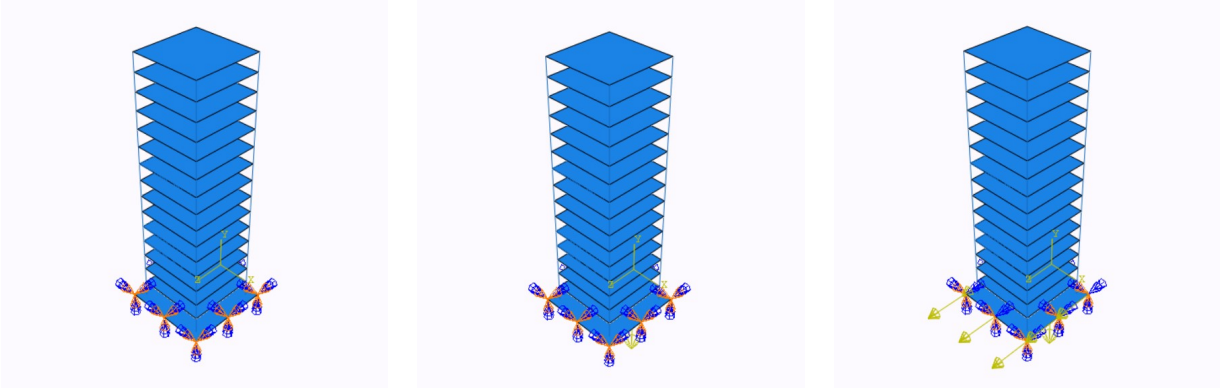
4.1.7 Loads and Boundary Conditions

This section corresponds to the load module of Abaqus and provides information regarding the loads and boundary conditions used in each of these models for each step.

3D Model

Two boundary conditions and a load is applied to the model. The first boundary condition is a displacement boundary condition on the bottom of the base plate, with all 6 degrees of freedom (3 translations and 3 rotations) fixed to zero at the initial step. This boundary condition is propagated as is to the next 2 steps and in the final step (dynamic, implicit), the constraint on the translation along the global z-axis (which is one of the horizontal axis for

the building model) is freed while keeping all other degrees of freedom fixed. An acceleration boundary condition is applied in the fourth step to apply ground acceleration along the z-axis (the same horizontal direction along which the building is free to move in this step). To do this, first, the ground motion data is added to the set of amplitudes in Abaqus, and the magnitude of acceleration along the z-axis is modified to apply the acceleration in mm^2/s (in this case, since all the ground motion data is in units of g's, the multiplier used is 9806.65). Gravity load is applied in the third step (static, general) and is propagated to the fourth step as is (gravity load helps to condition the problem better and reduce instabilities when solving). Figure 4.2 shows the boundary conditions and loads applied at each step in the 3D model.



(a) Steps 1 and 2 (b) Step 3 (c) Step 4

Figure 4.2. Boundary Conditions and Loads in the 3D Model

2D Model - Nonlinear

Similar to the 3D model, two boundary conditions and a load are applied to this model too. The first boundary condition is a displacement boundary condition applied on the base plate (which is modeled as a 2D planar wire), with all 3 degrees of freedom (2 translations and 1 rotation) fixed to zero at the initial step. This boundary condition is propagated as is to the second and third steps and in the final step (dynamic, implicit), the constraint on the translation along the global x-axis (the horizontal axis along which the degrees of freedom are being considered and along which the ground acceleration will be applied) is freed while

keeping the remaining two degrees of freedom fixed. An acceleration boundary condition is applied in the fourth step to apply the ground acceleration along the x-axis (the same horizontal direction along which the building is free to move in this step). To do this, first, the ground motion data is added to the set of amplitudes in Abaqus, and the magnitude of acceleration along the x-axis is modified to apply the acceleration in mm^2/s (in this case, since all the ground motion data is in units of g's, the multiplier used is 9806.65). Figure 4.3 shows the boundary conditions and loads applied at each step in the 2D model.

2D Model - Linear

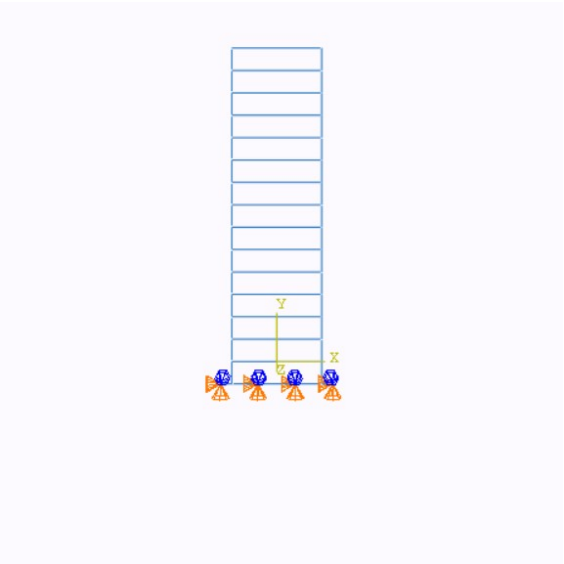
Just two boundary conditions are applied in the linear model. The first boundary condition is a displacement boundary condition applied on the base plate, with all 3 degrees of freedom (2 translations and 1 rotation) fixed to zero at the initial step. This is propagated as is to the second step. In the third step (Modal dynamics), the displacement boundary condition is built into the model automatically. The ground acceleration is applied to the model using the acceleration base motion boundary condition in the modal dynamics step. To do this, first, the ground motion data is added to the set of amplitudes in Abaqus, and the magnitude of acceleration along the x-axis is modified to apply the acceleration in mm^2/s (in this case, since all the ground motion data is in units of g's, the multiplier used is the value of acceleration due to gravity.

4.1.8 Mesh

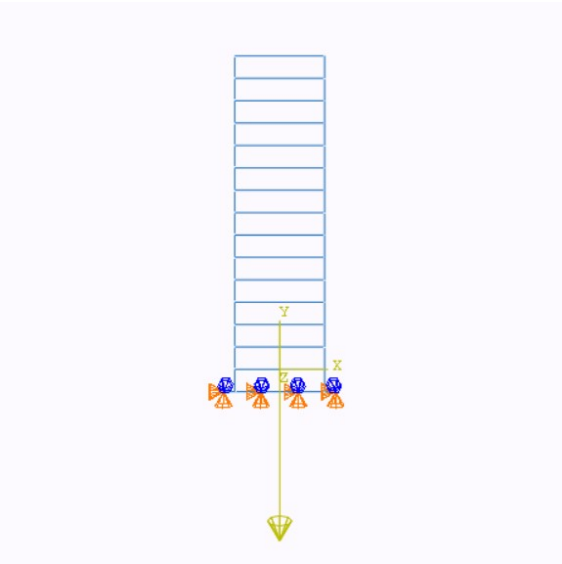
This section corresponds to the mesh module of Abaqus and provides the details of meshing along with the reason for selecting particular types and sizes of mesh. The meshes are kept as coarse as possible to get accurate results without any errors due to discretization.

3D Model

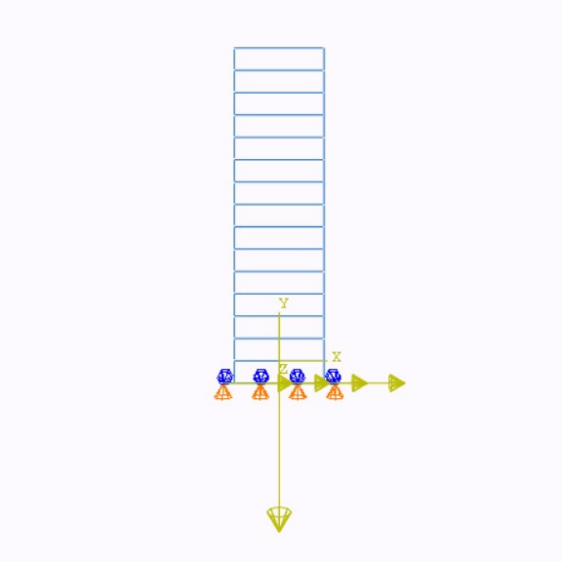
2-node linear beam elements in space (B31)[76] are used to discretize the frame elements. The use of linear elements with just 2 nodes instead of quadratic elements reduces the computational cost significantly but it also introduces small inaccuracies into the model.



(a) Steps 1 and 2



(b) Step 3



(c) Step 4

Figure 4.3. Boundary Conditions and Loads in the 2D Nonlinear Model

This decision can be justified in the following way: it is assumed in the study that the results obtained from the Abaqus model are the ground truth. Therefore, the overall complexity of the model is of the most importance and that does not reduce with the simplification in the number of nodes. Hence, this simplification does not affect complexities that need to be captured by the surrogate models when making predictions. In other words, the

surrogate modeling process developed in this study will work similarly regardless of whether the elements are modeled with 3 nodes or 2 nodes. Furthermore, the response parameters of interest are the peak drifts, which can be obtained by extracting the displacements at nodes, thus rendering the integration scheme inside the element less significant. An approximate global seed size of 1 mm is selected resulting in a total of 29944 nodes and 30000 elements for the whole frame. The remaining mesh properties are kept at the default values.

A linear 4-node doubly curved shell element, with reduced integration, hourglass control, and finite membrane strains (S4R)[76] is used for the discretization of the floor plates and the base plate. Using reduced integration is found to reduce the computational cost without significantly affecting the accuracy of results, and the finite membrane strains option is selected to accommodate nonlinearity in deformations. An approximate global seed size of 1.75 mm is selected, resulting in 82369 nodes and 81796 elements for the base plate and 793500 nodes and 786615 elements in total for the floor plates. The remaining mesh properties are kept at the default values.

Therefore, the model has 905813 nodes and 898411 elements consisting of 868411 S4R elements and 30000 B21 elements. Figure 4.4 shows the mesh obtained in this manner.

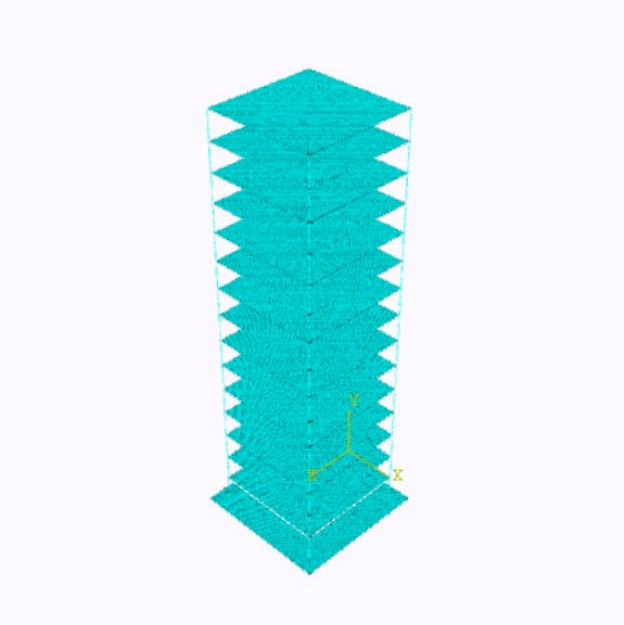


Figure 4.4. Mesh for 3D Model

2D Model - Linear and Nonlinear

Planar 2-node linear beam elements (B21)[76] are used to discretize all the elements in the model. Again, the use of linear elements with just 2 nodes instead of quadratic elements with 3 nodes can be justified because it does not reduce the overall complexity of the model, but does reduce the computational time. Hence, the approach of developing surrogate models is not affected when considering the results obtained from the Abaqus models as the ground truth. For the frame (which includes the floor slabs, modeled as beams as described in 4.1.1, and columns), the global seed size is set to approximately 10 mm, resulting in a total of 887 nodes and 900 elements. The global seed size for the base (which is also modeled as beams as described in 4.1.1) is also set to approximately 10 mm, resulting in a total of 51 nodes and 50 elements. The remaining mesh properties are kept at default values.

Therefore, there are a total of 938 nodes and 950 elements, all of type B21, in the 2D model. This is 3 orders of magnitude smaller than the number of nodes and elements in the 3D model, thus reducing the computational cost considerably, without affecting the accuracy of the responses obtained. Figure 4.5 shows the mesh obtained in this manner.

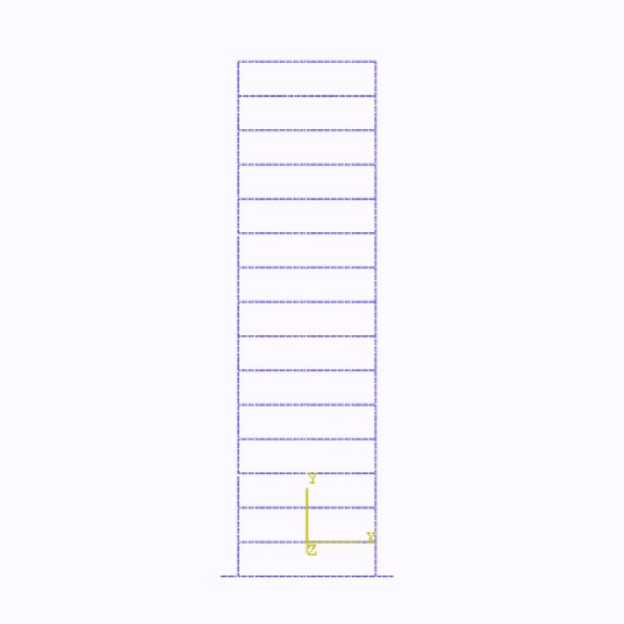


Figure 4.5. Mesh for 2D Model

4.1.9 Outputs

This subsection describes the set of outputs requested from Abaqus during the analysis. When computational time is of immense importance, only the outputs that are required can be requested and if higher computational power is available then other outputs may be requested. The outputs that are required for this study are listed in this section. The outputs that are requested are the same for all 3 types of Abaqus models used in this study. In the frequency step, the fundamental mode frequency and the corresponding mode shape are requested. In the remaining steps, displacement/velocity/acceleration and force/reactions are the outputs that are requested in Abaqus. All of these are field outputs, and no history outputs are required for this study.

4.1.10 MATLAB Code to Run Abaqus

As described in Section 3.3, the Abaqus model has to be run by changing the values of several parameters including Young's modulus and the density of steel, and the ground motion parameters (which includes the ground motion data, step increment time, and end time). Doing this manually to collect the amount of data required for this study would be impractical. Fortunately, Abaqus generates an input file that can easily be manipulated using any programming language to change the parameters as required. Although Python is usually the preferred programming language to alter input files generated from Abaqus, MATLAB has been used in this study. Using MATLAB to generate data allows us to post-process the obtained results and save the calculated responses of interest such as the frequencies and peak drift ratios as MATLAB data files (.mat files). This makes it easier to access the data later and use the functionalities in MATLAB to run various studies. Therefore, a MATLAB code is written to change the Abaqus input file for multiple sample points and run the models automatically. The code is used to update the input file obtained from Abaqus to include the correct values for Young's Modulus, the density of steel, and ground motion data. In this study, the data is generated using 2 different versions of MATLAB, version R2021b, and R2023a. Although the version of MATLAB does not affect the generation of data in any way,

version R2023a has better functionalities for later studies, which is the reason MATLAB was updated to this version as the project progressed.

4.1.11 Assumptions

The high-fidelity data obtained in this study from Abaqus models have simplifications built in, in terms of the assumed material behavior and the physical behavior that is modeled. In the real world, the behavior of the structure as well as material properties are much more complex and it is expected that the challenge to develop a surrogate model to make predictions with such levels of detail would be higher. This section lists a few of the assumptions that are made while developing these models.

1. Just the moment resisting frame is modeled in Abaqus. The contribution to stiffness and mass by other parts such as the infill walls and non-structural elements is assumed to be negligible.

2. Steel is modeled as elastic-perfectly plastic. In reality, however, the behavior of steel after yielding is more complex, with nonzero stiffness. However, the post-yield stiffness of steel is usually negligible and the results do not vary by a lot when not considering this portion. Using this assumption gives conservative results as the stiffness is modeled to be lower than its actual value.

3. Usually, during ground motion excitations with significant magnitudes, it is observed that buildings undergo damage, reducing their stiffness as the excitation progresses. While this model considers a reduction in the post-yield stiffness of steel, it does not account for certain damage such as slippage in joints or damage due to the presence of defects in joints. Joints are assumed to be perfect, which is rarely the case in the real world but is useful for this study on surrogate modeling.

4. It is assumed that the ground motion excitation acts just along one horizontal dimension of the building. However, this is not necessarily the case, and ground motion excitations usually have significant components along both horizontal directions.

5. The effect of vertical accelerations during ground motions is ignored. Although this is a simplification, this follows what has been observed in past research[43].

There are additional assumptions for the modal analysis[77]. These are the following:

1. Model is linear. Both material and geometric non-linearities are ignored.

2. It is assumed that the dynamic response of the building is a linear combination of the natural modes of the system.

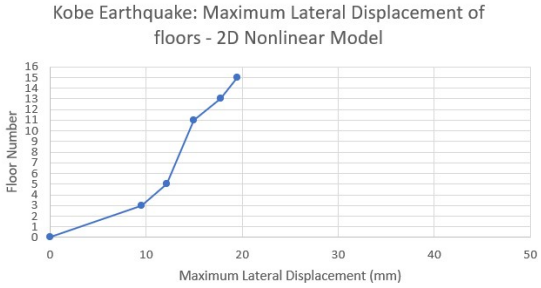
3. Because non-linearities are ignored, the resulting deformations are limited to smaller values than the corresponding values when considering non-linearities.

4.1.12 Validation of the Abaqus Models

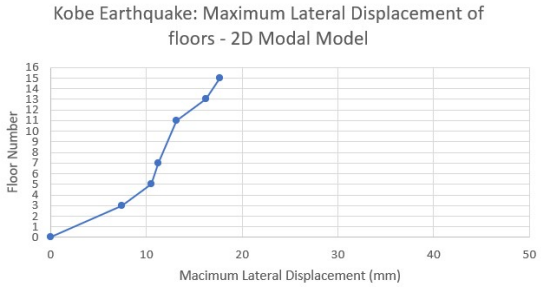
This section provides the details regarding the validation of the Abaqus models by comparing them with the results from the reference paper. Since the responses of interest are the maximum drifts, the maximum drifts obtained from Abaqus at various floors are compared with the corresponding values provided in the reference paper. This process is undertaken for 2 of the 4 ground motion excitations used in the reference paper, namely, the Kobe earthquake in 1995 (recording station: Kobe Japanese Meteorological Agency station) and the El Centro earthquake in 1940 (recording station: El Centro). It is not done for the Northridge earthquake in 1994 (recording station: Sylmar County Hospital ground) because of a discrepancy in the time step of the obtained ground motion data set with the one used by the authors in the reference paper, and for the Hachinohe earthquake in 1968 (recording station: unknown) because lack of credible source with the ground motion recording data for this earthquake. Care is taken to ensure that data from the same ground motion recording station is used as used in the reference paper.

Figures 4.6 and 4.7 show the validation plots for both the 2D Nonlinear and 2D Modal Abaqus models for Kobe and El Centro earthquakes respectively. While there are small errors in the magnitudes of lateral displacements between the results from the Abaqus model to the results from the reference paper (green or olive-colored curves), the trend is similar to

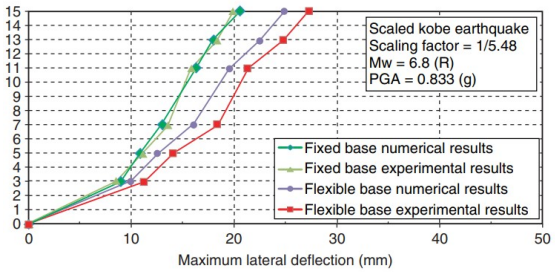
the results from the reference paper. For the development of surrogate models, it shall be assumed that results from the 2D Nonlinear Abaqus model are the exact high-fidelity data.



(a) Results from Abaqus 2D Nonlinear Model

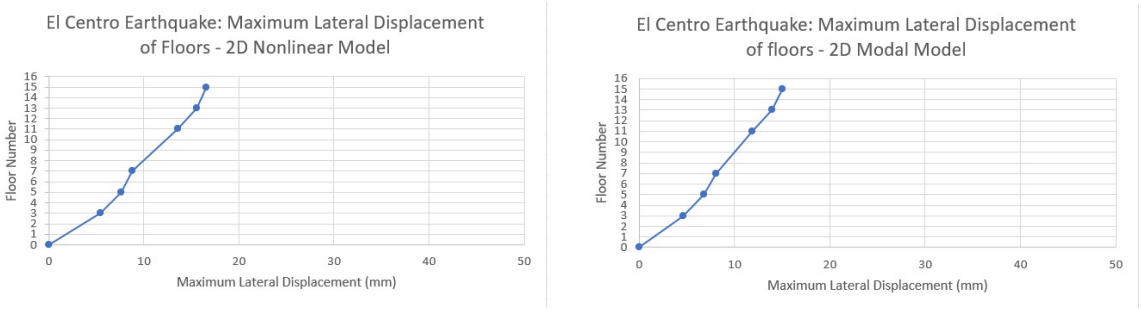


(b) Results from Abaqus 2D Modal Model



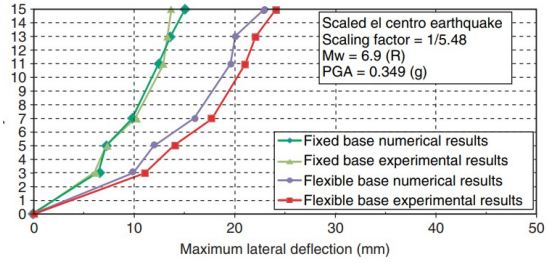
(c) Results from Reference Paper

Figure 4.6. Validation of Abaqus Models: Kobe Earthquake



(a) Results from Abaqus 2D Nonlinear Model

(b) Results from Abaqus 2D Modal Model



(c) Results from Reference Paper

Figure 4.7. Validation of Abaqus Models: El Centro Earthquake

5. FUNDAMENTAL FREQUENCY

The first parameter of interest that is to be predicted in the study is the fundamental frequency of the building. This parameter, unlike other parameters considered in this study (described in the subsequent chapter), is an inherent property of the structure. In other words, the fundamental frequency of a building is independent of the ground motion data and is only dependent on the mass and stiffness of the building. Therefore, among the parameters that are varied in the study, just the density and Young's modulus of steel affect the value of frequency.

Hence, the predictor variables used as inputs to the surrogate models are the two structural properties as shown in Figure 5.1. The variation of these parameters is not random like the variation of ground motion characteristics, consequently, it is easier to build accurate surrogate models to predict the fundamental frequency of the structure. This implies that the number of training points required to develop an accurate surrogate model to predict the fundamental frequency will be lower than the number of training points required to predict the structural responses considered in Chapter 6. This also suggests that more accurate surrogate models can be built to predict the fundamental frequency of the structure than to predict the structural responses considered in the following chapter.

Figure 5.1. Flowchart of the Structure of the Surrogate Models for the Predictions of Fundamental Frequency

This chapter describes the study on the development of surrogate models to predict the fundamental frequency of the structure. The chapter is divided into 3 sections. Section 5.1 describes the methods used to sample points in the parameter space at which the high-fidelity data needs to be collected. Following this, Sections 5.3 and 5.3.3 provide information about the data-driven and physics-based surrogate models developed in this study for the predic-

tion of the fundamental frequency of the structure. Both these sections have subsections describing the methodologies used, the results obtained, and the lessons from developing these types of surrogate models. Finally, Section 5.5 compares the different types of surrogate models developed for predicting the fundamental frequency of the structure using the performance metrics described in Section 3.6 and provides the conclusions.

5.1 Design of Experiments

The first step in developing surrogate models is to select points in the parameters where high-fidelity data will be generated to train the surrogate models on. This is referred to as the design of experiments. Similarly, one also has to select a set of points to test the performance of the surrogate models as well. This section covers the details of this process for developing surrogate models to predict frequencies.

As mentioned in the 3.3, the properties of the structure are varied by changing just the mass and the stiffness of the structure. Since the fundamental frequency of the structure depends on only these two properties, a small amount of data points are enough to train the surrogate models to obtain accurate predictions. In this study, a central composite design (CCD) with nine points across the Young's Modulus - Density space is selected for obtaining the high-fidelity data. The range for both these parameters in training is chosen as ± 3 standard deviations to cover 99.87% of the parametric space. This is done to prevent the surrogate models from requiring to extrapolate when making predictions as most statistical methods struggle to make accurate predictions when extrapolating. MATLAB function ccdesign is used to obtain this set of points for this study. Figure 5.2 shows the points selected for training and Figure 5.3 shows the corresponding points along the probability distributions of the corresponding random variables.

A set of 25 points ranging across Young's Modulus - Density parameter space is selected for testing the accuracy of the surrogate models. Latin-Hypercube Sampling is used for creating the testing data set to ensure that points are selected with equal probability across the parameter space to determine if the developed surrogate models make accurate predictions

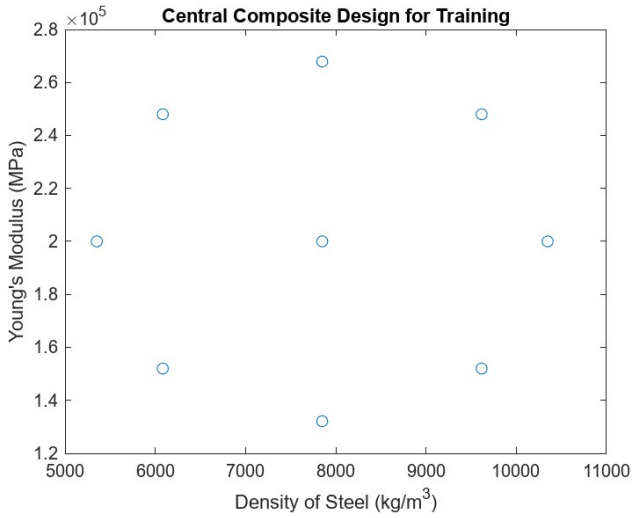
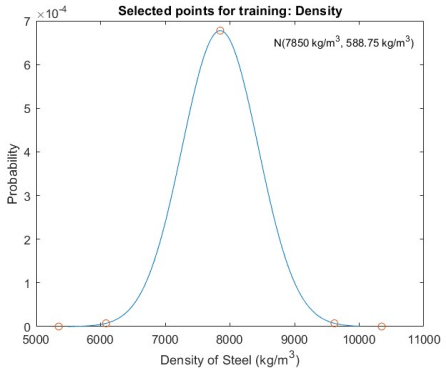
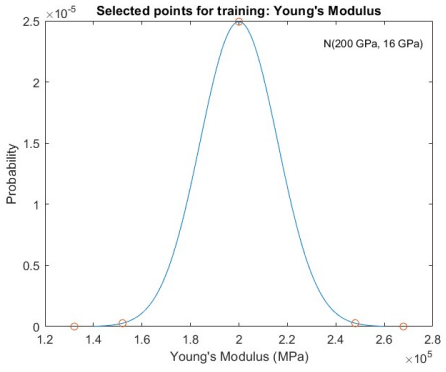


Figure 5.2. Central Composite Design for Training Surrogates to Predict Fundamental Frequency



(a) Selected Points for Density of Steel



(b) Selected Points for Young's Modulus of Steel

Figure 5.3. Selected Points for Training Surrogate Models to Predict Fundamental Frequency[2]

throughout the space. Figure 5.4 shows the points used for testing the accuracy of surrogate models.

[2]Source: González and Zapico[70]

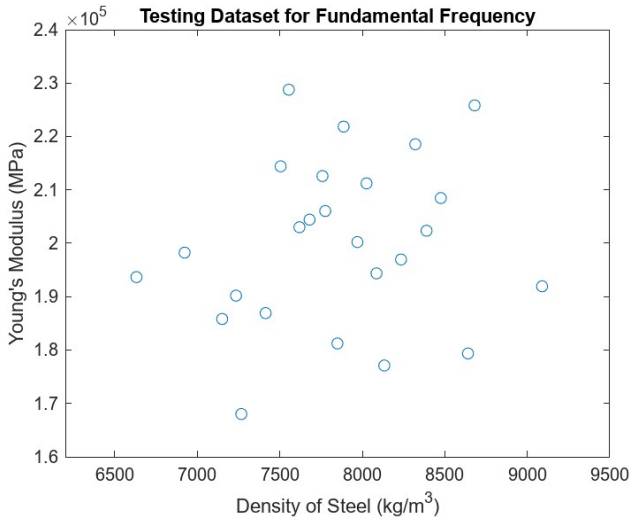


Figure 5.4. Testing Data Points for Fundamental Frequency

5.2 High-Fidelity Model

Once the points are selected, the next step is to generate high-fidelity data at these points and form input-output pairs on which the surrogate models can be trained and tested. The details regarding the generation of high-fidelity data are provided in this section. The validated Abaqus model is used to obtain the high-fidelity values of the fundamental frequency of the structure. Since all three of the Abaqus models are more or less the same until the second step and provide similar results, any of these models can be used. In this thesis, the 2D-Nonlinear Abaqus model described in Chapter 4 is used to generate high-fidelity fundamental frequency values.

In Abaqus, the natural frequencies of a structure can be obtained using the Frequency step which uses eigenvalue analysis to obtain the frequencies of interest. There are different solvers for eigenvalue analysis, and in this study, the default Lanczos solver is used. The computational time for the extraction of natural frequencies of a structure using Abaqus is not as high as when obtaining the responses of a structure to ground motions by conducting time-history analyses, and the running 2D-Nonlinear Abaqus model to generate fundamental frequency values takes about a minute on average per sample point.

5.3 Data-Driven Surrogate Techniques

The regression learner app in MATLAB is used to train different types of data-driven models for the prediction of the fundamental frequency of the structure. Since the value of the fundamental frequency depends on both Young's Modulus and the Density of Steel, these two parameters are taken as predictor variables that are input to the surrogate models, with the fundamental frequency being the sole output of interest. Bayesian optimization[78] option that is available in the regression learner app is used for the automatic tuning of hyper-parameters of the models to obtain the model with the best accuracy and fit. From trial and error, 200 iterations of Bayesian optimization are determined to be ideal to obtain the best hyper-parameters. It is found that using a lower number of iterations did not allow for sufficient convergence of the error during optimization and allowing a higher number of iterations often did not result in significant improvements. The option to standardize data is present in the regression learner app, and this option is kept on for all surrogate models. Four types of data-driven modeling techniques are used to obtain a surrogate model for the prediction of frequency: 1) Support Vector Machines (SVM), 2) Gaussian Process Regression (GPR), 3) Artificial Neural Networks (ANN), and 4) Linear Regression. 3-fold cross-validation is used to validate the model while using all of the 9 points in training and the 25 points testing set is used to test the accuracy and fit of prediction.

5.3.1 Results

The results using the set of 9 points in training and the set of 25 points in testing are shown in Table 5.1. This table lists the RMSE and R^2 values in validation and testing for the most accurate and best-fit data-driven models of each type. Table 5.2 lists the most accurate data-driven method that is developed along with the ideal hyperparameters tuned using Bayesian optimization.

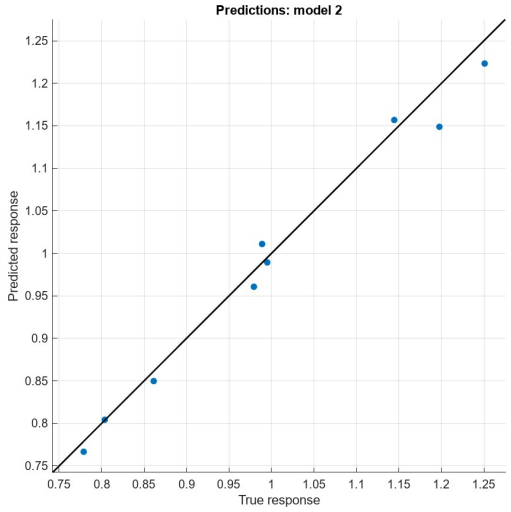
Figure 5.5 shows the performance of the SVM model in validation and testing by plotting the predicted values of fundamental frequency against the actual values.

Table 5.1. Performance of Data-Driven Models for Fundamental Frequency Prediction

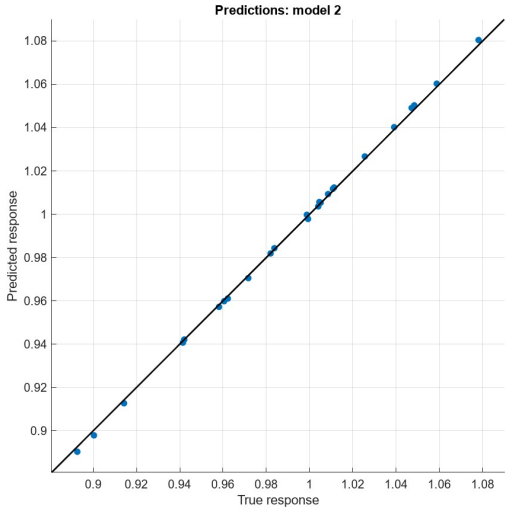
Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.0222	0.99	0.0013	0.999
GPR	0.0508	0.93	0.0107	0.94
ANN	0.0520	0.93	0.0045	0.991
Linear Regression	0.0469	0.94	0.0102	0.95

Table 5.2. Best Data-Driven Surrogate Trained for Fundamental Frequency

Model	Box Constraint	Kernel Scale	Epsilon	Kernel Function
SVM	654.8406	18.8175	2.8425^{-4}	Gaussian



(a) Predicted vs Actual Values of Fundamental Frequency in Validation for the Best Data-Driven Surrogate



(b) Predicted vs Actual Values of Fundamental Frequency in Testing for the Best Data-Driven Surrogate

Figure 5.5. Performance of the Best Data-Driven Surrogate Model for Fundamental Frequency

5.3.2 Discussion

Table 5.1 shows that a data-driven surrogate model with great accuracy and fit is obtained using just 9 points in training. In fact, with the SVM model, an NRMSE of 0.0013 and an R^2 value of nearly 1 is obtained. It can be seen in Figure 5.5b that the prediction error is very

small, if not negligible. This shows that, as expected it is easy to predict the fundamental frequency of a structure with variation in its structural properties.

Given that generating the high-fidelity data for fundamental frequency using the 2D-Nonlinear Abaqus model requires on average 1 minute per sample point, the initial cost to generate points to train the data-driven surrogate is about 9 minutes plus the time to obtain the best hyperparameters, which is about 3 minutes on average. Therefore, an initial cost of about 12 minutes is required to build this data-driven surrogate. However, once the surrogate model has been developed it can be used to obtain the fundamental frequencies at all other points in the parameter space at virtually no computational expense. Therefore, if the fundamental frequency has to be evaluated for a lot of different values of structural properties, then first training a surrogate model like this one and then using it will be more effective than using Abaqus for each configuration.

It is worth noting that the performance of the surrogate models is better in testing than in training as seen in Table 5.1. This is because extreme values are selected by using the central composite design for training, while the points selected for testing are distributed according to the assumed distribution of the structural properties. The reason for using this approach in training is to ensure that the model does not have to extrapolate when making predictions, as statistical models usually have high errors when extrapolating.

5.3.3 Physics-Based Surrogate Techniques

A low-fidelity physics-based analytical model can be obtained by assuming that the building behaves as a shear building with lumped masses. There are multiple ways to map the results obtained from the high-fidelity model to the high-fidelity fundamental frequency values; this can be done using various types of machine learning techniques or can be done manually using analytical equations. In this study, the second approach is used. Usually when building a lumped mass, shear building model, the mass of the columns is ignored and it is assumed that all of the structure's mass is concentrated in the floor slabs. Similarly, it is assumed that all of the stiffness of the structure is provided by the columns only and the contribution of the floor slabs to the overall stiffness is ignored. This helps transform

a continuous structure into a discrete model with a significantly lower number of degrees of freedom. The natural frequencies of this model are easily obtained by solving the characteristic eigenvalue problem of the model numerically, by inverting the mass matrix using Gaussian elimination and multiplying the inverse of this with the stiffness matrix. In this study, the low-fidelity results are matched to the high-fidelity data points by augmenting the mass matrix to include a part of the mass of the columns. This is done manually as shown in Equation 5.1

$$M_{\mathrm{i,i}} = M_{floor\ \mathrm{i}} + factor * (M_{half\ single\ column\ above\ floor\ \mathrm{i}} + M_{half\ single\ column\ below\ floor\ \mathrm{i}}) \quad \forall \mathrm{i} = 1 : number\ of\ floors \tag{5.1}$$

which reduces to

$$M_{\mathrm{i,i}} = M_{floor} + factor * (M_{single\ column})\ \forall \mathrm{i} = 1 : 15 \tag{5.2}$$

due to the selection of a regular building. High-fidelity data points are used to solve for the factor in 5.2. This factor can be chosen to be a function of Young's Modulus and Density of steel. However, in this study, it is found that a constant value is enough for an exact match in the predicted fundamental frequency up to 4 decimal values.

5.3.4 Results

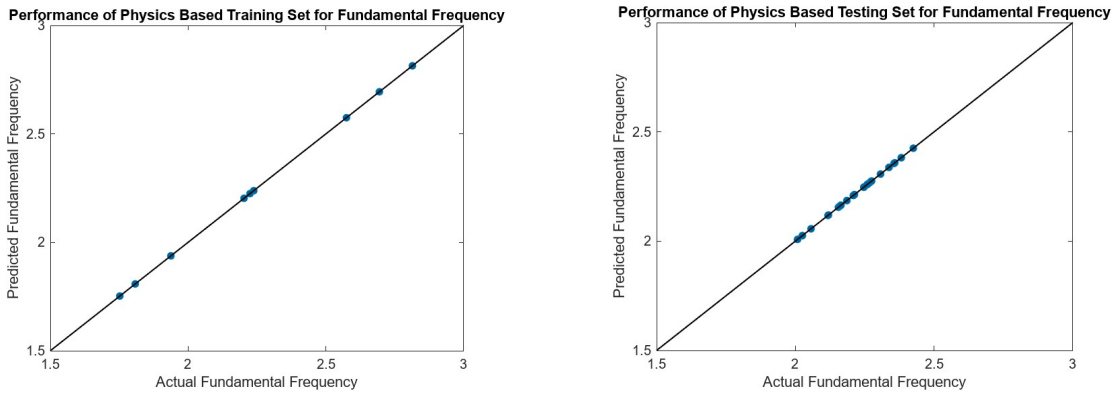
The value of the factor in Equation 5.2 can be found in several ways, including regression. In this study, however, a trial and error method is used. It is found that a constant value of 7.05 is enough to get an exact match with the high-fidelity values of the fundamental frequency at every point.

Since this is an analytical method, there is no training for the model here. Therefore, all of the 9 selected points using central composite design are used for validation. The same 25 points set is used for testing the model. The performance metrics for this system are provided in Table 5.3.

Table 5.3. Performance of Physics-Based Surrogate Model for Fundamental Frequency

Validation NRMSE	**Validation R^2**	**Testing NRMSE**	**Testing R^2**
$7.0857*10^{-10}$	1.00	$8.0321*10^{-10}$	1.00

Figure 5.6 shows the performance of the physics-based surrogate model in validation and testing by plotting the predicted values of fundamental frequency against the actual values.



(a) Predicted vs Actual Values of Fundamental Frequency in Validation for the Physics-Based Surrogate Model

(b) Predicted vs Actual Values of Fundamental Frequency in Testing for the Physics-Based Surrogate Model

Figure 5.6. Performance of the Physics-Based Surrogate Model for Fundamental Frequency

5.3.5 Discussion

The results show that by using the value of 7.05 in Equation 5.2, an exact match is obtained in the predicted and actual values of fundamental frequencies for this structure. The analytical model computes the fundamental frequencies almost instantly once the value of the factor is determined. In this study, since trial and error were used to determine the value of this factor, it is difficult to determine the time required to find the value of the factor. However, if some sort of regression was used to obtain the value of the factor, we can expect a similar time to develop the physics-based surrogate model as was required to develop the data-driven surrogate models in Section 5.3. Therefore, as with the data-driven

surrogate model, when the fundamental frequency of the structure has to be computed for many different configurations, the time required using the physics-based surrogate model is significantly lower than the time required to generate the same amount of data using the high-fidelity 2D-Nonlinear Abaqus model.

5.4 Findings

The findings from the study for the prediction of fundamental frequency for the selected building using surrogate models are listed below:

- A central composite design with 9 training points was enough to develop data-driven surrogate models with NRMSE of less than 0.5 and R^2 of over 0.95 in both training and testing.
- The predictions of fundamental frequency from the physics-based surrogate model, developed using the shear building assumption followed by modification of the mass matrix to account for the mass of the columns, match the high-fidelity fundamental frequency values up to four decimal places.

5.5 Conclusions

As expected both data-driven and physics-based surrogate models can be easily developed to predict the fundamental frequency of the structure for changing values of density and Young's modulus of steel with great accuracy.

For both the data-driven and the physics-based models, the prediction process involves just solving a system of equations. They both require almost no time to make predictions. However, an advantage of the physics-based model over the data-driven surrogate model it is easier to understand the prediction process for the physics-based model because it is based on an analytical simplification of the problem at hand, unlike the data-driven model which is a black box with the best hyperparameters which are fairly random, determined by Bayesian optimization. Furthermore, if the factor is known, the time to develop the physics-based surrogate model is less than the initial time required to train the data-driven model.

Another advantage of the physics-based surrogate model is that the lessons from this problem can be applied to other structures as well. However, the data-driven model with the same hyperparameters cannot directly be applied to other structures, making this type of model more specific to the problem at hand. The data-driven surrogate model can make accurate predictions only for this particular building with the selected structural properties varying inside the selected range. One potential disadvantage of the physics-based model is that it could be difficult to build simple analytical models for structures with higher levels of complexity. However, in such cases, one could expect to require more points to train the data-driven models to obtain satisfactory performance as well.

6. PEAK STRUCTURAL RESPONSES

It is usually the practice in seismic engineering to determine certain peak structural responses to given ground motions and determine the state of the building by comparing these values to predefined limit values of the corresponding parameters[6]. Some of the common response parameters of interest are the maximum roof drift ratio, the maximum story drift ratio, and the maximum acceleration experienced by the structure.

In this study, surrogate models are developed to predict the maximum roof drift ratio (MRDR) and the maximum story drift ratio (MSDR) of the structure under consideration for various ground motions. The maximum roof drift ratio is the ratio of the maximum displacement experienced by the roof of a structure with the overall height of the structure. The maximum story drift ratio measures the maximum value of the ratio of the displacement experienced by any floor with respect to its adjacent floor with the height of the corresponding story enclosed by these floors.

In fragility analysis, the amount of structural damage that a building undergoes is usually defined as a function of either the maximum roof drift ratio or the maximum story drift ratio, while the amount of damage to non-structural components can be divided into two parts, one that is proportional to the overall drift, and another that is proportional to the maximum acceleration experienced by the structure.

Since these parameters are responses of the structure to ground motions, the values of these parameters vary with both the structural properties as well as the ground motion characteristics. Since the ground motions are fairly random, developing surrogate models to obtain the values of these responses is a challenging task. Furthermore, nonlinearities in the problem add to the complexity of predicting these response parameters. Therefore, one can expect to require a higher number of high-fidelity data points to train an acceptable surrogate model to predict these values as compared to the prediction of the fundamental frequency of the structure that is presented in the previous chapter.

Section 6.1 provides the details about the data-driven surrogate models developed in this study, while Section 6.3 provides the information regarding the physics-based models

developed in this study for the prediction of these maximum structural responses. Following this, Section 6.4 compares the two types of models and draws conclusions based on the study.

6.1 Data-Driven Surrogate Techniques

Owing to the complexity of the problem, the study to develop data-driven surrogate models for determining these peak structural responses is divided into multiple sub-studies. The complexity of the problem is reduced in the initial studies, to begin with, and slowly added in later studies as more information is gathered about the system. Each of these sub-studies answers a particular set of questions, providing key learning that is used in the subsequent studies.

To start with, two preliminary studies are conducted using high-fidelity data from linearized models to get an approximate measure of the number of high-fidelity data points needed and figure out which predictors capture the response and its variations accurately. This is done in the interest of saving time by using linear models instead of complex nonlinear models. Findings from these preliminary studies are used to set up the method for the problem for the study with the actual nonlinear model.

Each subsection in this section is divided into 5 more subsections. First, information regarding the design of experiments is provided. This is followed by a section providing the details of the model selected for generating high-fidelity data. Following that, information regarding the surrogate models that are developed for the corresponding substudy is provided. Following that, key results are provided and finally, the findings from each of these studies are listed.

6.1.1 Preliminary Study 1: Same Building Scenario

For the first preliminary study, the properties of the building are kept the same and only the ground motions are varied. The objective of this preliminary study is to answer the following questions:

1. Which ground motion parameters capture the trend in peak structural responses the best?

2. Which of the 2 peak structural responses is easier to predict?

3. How many high-fidelity data points are required for training the data-driven models to obtain acceptable performance?

Design of Experiments

Since structural properties are kept constant at the mean value for this study, just the ground motion records need to be changed. 125 real ground motions are selected from different ground motion databases as described in 3.4. These ground motions are randomly shuffled to reduce bias in training sets.

25 ground records are kept aside for testing the surrogate models. Out of the remaining 100 records, 25 are chosen for the initial training process. Surrogate models are trained using this initial set and the performance in training and testing is checked. If no surrogate models satisfy the required accuracy or fit, 25 more ground records are added to the training set to increase the number of ground records to 50. The process is repeated until 100 ground motions are reached in training or a model with a satisfactory level of accuracy in both validation and testing is obtained.

This cycle is repeated for different combinations of input parameters for both the prediction of MRDR and MSDR.

High-Fidelity Model

As mentioned above, a linearized model is used to obtain high-fidelity data for this preliminary study. This is done to reduce the computational time required to obtain high-fidelity data for the study. Since we do not need exact answers yet and will be building upon the learning from this study in a more complete study with the fully nonlinear high-fidelity model, this approximation is acceptable. It is important to however remember the assumptions made when making these simplifications, the most important one being that non-linearities are ignored.

For this preliminary study, the responses generated using the fundamental frequency matched shear building model, developed in 5.3.3 is assumed to be the pseudo-high-fidelity

data. The peak structural responses from this model are obtained by first obtaining the time-history response of the shear building using numerical integration (using ODE45 in MATLAB [79]) and then determining the maximum value of the response.

Data-Driven Modeling Techniques

As was the case when predicting fundamental frequencies, the MATLAB regression learner app is used to build data-driven surrogate models. Again 4 types of modeling techniques are tried: 1) Support Vector Machines (SVM), 2) Gaussian Process Regression (GPR), 3) Artificial Neural Networks (ANN), and 4) Linear Regression (LR). Bayesian optimization[78] with 200 iterations is used to tune the hyperparameters of the models.

Multiple different parameters capture various characteristics of the ground motions as listed in past literature2. Since ground motions are highly complex and nonlinear, it is also true that just a single parameter cannot capture all the characteristics of the ground motions. This is why some authors have used multiple predictor variables when developing fragility curves[37], [38].

Since the main objective of this preliminary study is to understand which parameter/set of parameters provides the most accurate surrogate models, different parameters and their combinations are tried as predictor variables to the surrogate models in this study. Two types of parameters are used as predictor parameters:

1. Peak-based Values: Peak ground acceleration. This value just captures the maximum ground acceleration and does not capture any information regarding the frequency content or the duration of ground motion excitation.

2. Frequency-based values: spectral response properties (spectral acceleration, spectral velocity, and spectral displacement) capture the frequency content of a ground motion. For this study, the spectral response properties are obtained using the numerical response of an equivalent Single Degree of Freedom (SDOF) system. An equivalent single-degree-of-freedom system is a single-degree-of-freedom system with the same fundamental frequency and the fundamental mode damping ratio as the actual structure. The spectral response properties are the maximum responses (absolute acceleration,

relative velocity, and relative displacement) of this equivalent SDOF. The response can be obtained by numerically conducting a time-history analysis of the SDOF and noting the peak values. ODE45 in MATLAB is used to numerically solve the equations of motions for the SDOF in this study. The computational time to do this is 25 seconds on average.

Since the equivalent SDOF is usually created to have the same fundamental frequency as that of the actual structure, the spectral response properties depend on both the structural properties (because the fundamental frequency varies with the structural properties) as well as the ground motion characteristics. However, for this preliminary study, since we are keeping the structural properties constant, the fundamental frequency of the equivalent SDOF remains constant and the spectral response properties just vary with the ground motion properties.

5-fold cross-validation is used to validate the model while using all of the points in training and the 25 points testing set is used to test the accuracy and fit of prediction. The results from this study are used to understand which set of predictor parameters provides the most accurate results when predicting the respective peak responses.

Results

Results from different combinations of input parameters to the data-driven surrogate model for this preliminary study are provided in Appendix A. The results of the hyperparameters for the best model obtained are fairly random and hence are thus not provided.

Table 6.1 provides a summary of the performance of the best models for each combination of input parameters along with the corresponding number of training data points required for the prediction of maximum roof drift ratio.

Table 6.2 provides a summary of the performance of the best models for each combination of input parameters along with the corresponding number of training data points required for the prediction of maximum story drift ratio.

Table 6.1. Data-Driven Model Results: Same Building - Maximum Roof Drift Ratio

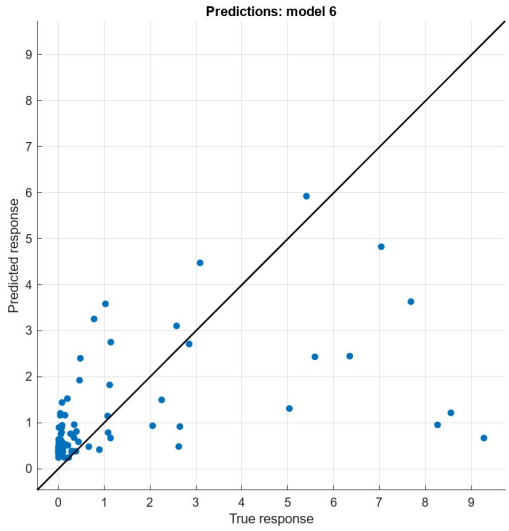
Inputs Parameters	**Points Required**	**Best Model**	**Validation NRMSE**	**Validation R^2**	**Testing NRMSE**	**Testing R^2**
PGA	100	ANN	1.72	0.32	1.85	0.43
S_a	25	ANN	0.23	0.995	0.34	0.98
S_v	100	GPR	0.49	0.94	0.47	0.96
S_d	25	SVM	0.18	0.997	0.34	0.98
PGA, S_a	25	GPR	0.16	0.997	0.14	0.997
PGA, S_v	75	SVM	0.48	0.96	0.38	0.98
PGA, S_d	25	SVM	0.13	0.998	0.22	0.99
S_a, S_v	25	ANN	0.12	0.999	0.25	0.99
S_a, S_d	25	LR	0.24	0.994	0.32	0.98
S_v, S_d	25	GPR	0.14	0.998	0.20	0.99
PGA, S_a, S_v	25	SVM	0.31	0.99	0.29	0.99
PGA, S_a, S_d	25	LR	0.22	0.996	0.19	0.99
PGA, S_v, S_d	25	ANN	0.14	0.998	0.28	0.99
S_a, S_v, S_d	25	ANN	0.11	0.999	0.25	0.99
PGA, S_a, S_v, S_d	25	SVM	0.13	0.998	0.29	0.99

Table 6.2. Data-Driven Model Results: Same Building - Maximum story Drift Ratio

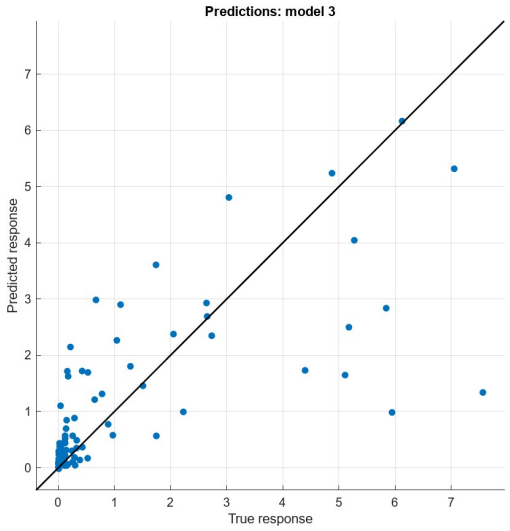
Inputs Parameters	**Points Required**	**Best Model**	**Validation NRMSE**	**Validation R^2**	**Testing NRMSE**	**Testing R^2**
PGA	100	GPR	1.18	0.55	1.09	0.77
S_a	100	ANN	0.55	0.90	1.05	0.79
S_v	100	ANN	0.39	0.95	0.62	0.93
S_d	100	GPR	0.56	0.90	0.99	0.81
PGA, S_a	100	LR	0.36	0.96	0.30	0.95
PGA, S_v	100	ANN	0.31	0.97	0.58	0.94
PGA, S_d	50	ANN	0.33	0.97	0.50	0.95
S_a, S_v	25	SVM	0.26	0.99	0.40	0.97
S_a, S_d	100	LR	0.46	0.94	0.81	0.87
S_v, S_d	25	SVM	0.29	0.99	0.45	0.96
PGA, S_a, S_v	100	ANN	0.33	0.97	0.53	0.95
PGA, S_a, S_d	100	ANN	0.34	0.96	0.55	0.94
PGA, S_v, S_d	50	GPR	0.32	0.97	0.50	0.95
S_a, S_v, S_d	25	SVM	0.42	0.97	0.35	0.98
PGA, S_a, S_v, S_d	50	ANN	0.30	0.98	0.50	0.95

Discussion

The first thing worth noting is that just PGA by itself is not a good predictor variable for the prediction of either maximum roof drift ratio or maximum story drift ratio. Although PGA does have some correlation with both these peak responses (the R_2 value is not 0), it does not correlate well enough with either response to satisfy the target performance metrics. Figure 6.1 shows the scatter on using 100 training points with PGA as the sole input parameter to predict the maximum roof drift and story drift ratios.



(a) Scatter in Validation of MRDR with Input: PGA

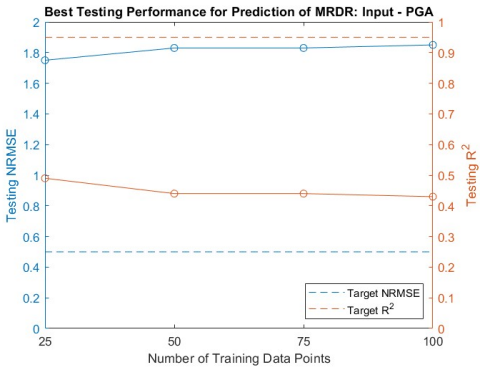


(b) Scatter in Validation of MSDR with Input: PGA

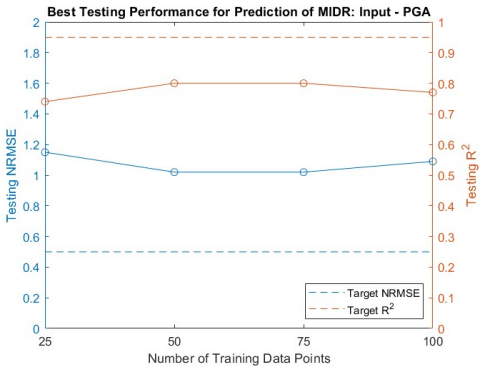
Figure 6.1. Scatter in Validation of Peak Structural Responses with Input: PGA for 100 Training Points

Furthermore, increasing the number of points does not significantly decrease the error in validation or testing when using PGA as the only input parameter for the surrogate model. This is shown in Figure 6.2. This suggests that just the PGA of ground motions does not capture the variation in peak responses well enough. This finding is expected because PGA does not capture the frequency content of the ground motions.

Among the remaining input parameters for the prediction of the maximum roof drift ratio, it is seen that both spectral acceleration and spectral displacement provide surrogate models that satisfy the performance criteria. Spectral displacement as the predictor provides



(a) Best Testing Performance for Prediction of MRDR for Input: PGA



(b) Best Testing Performance for Prediction of MSDR for Input: PGA

Figure 6.2. Performance Metrics for the Best Data-Driven Models for Predicting Peak Responses for Input: PGA

the best performance of the developed surrogate model. There could be several explanations for this behavior. To start with, spectral displacement measures the maximum displacement of a single story structure and therefore has a good correlation with the maximum roof drift ratio. Another reason could be that the structural response for this structure falls in the constant displacement region of the response spectrum.

As expected, the prediction becomes more accurate when more than one parameter is used as input parameters. It is seen that combining PGA with any of the spectral parameters results in a surrogate model with performance better than any single parameter input surrogate model. Therefore, while PGA by itself does not capture the characteristics of the ground motion, when it is used along with any spectral response property, the trend in the maximum roof drift ratio is captured very well. This confirms the findings from Pena et. al. [37]. Similarly, using any two spectral response properties together as the inputs to surrogate models also results in surrogate models with satisfactory performance. The same trend is seen as any three or all four of the input parameters are used together as inputs to the surrogate model.

The results for the prediction of the maximum story drift ratio are not as straightforward as that for the prediction of the maximum roof drift ratio. It is seen that among the remaining input parameters, the spectral velocity is the best parameter to use as input to

the surrogate models used for the prediction. However, using just the spectral velocity as the input parameter did not result in a surrogate model with satisfactory performance even with 100 training points.

As seen for the prediction of the maximum roof drift ratio, the prediction of the maximum story drift ratio also becomes more accurate when more than one parameter is used as input parameters. Using the spectral velocity with either or both of the remaining spectral response properties resulted in surrogate models with satisfactory performance with just 25 training data points.

Findings

This subsection lists the findings from the first preliminary study where we used pseudo-high-fidelity data and neglected the randomness in structural properties.

- Using PGA alone as a predictor causes large errors in the prediction of peak structural responses.
- Spectral displacement is the best predictor for the prediction of the maximum roof drift ratio.
- Spectral velocity is the best predictor for the prediction of the maximum story drift ratio.
- The prediction of maximum roof drift ratio is easier than the prediction of maximum story drift ratio.
- 25 points in training are sufficient to get good predictions of maximum roof drift ratio for most combinations of input parameters.
- For the prediction of maximum story drift ratio, anywhere from 25 to more than 100 points are required in training based on the combinations of input parameters.

6.1.2 Preliminary Study 2: Linearized Fixed Base Scenario

After determining the best predictor sets in the previous preliminary study, the next question to be answered is how the number of high-fidelity data points required increases when the structural properties of the building are also varied. Therefore, in this preliminary study, both the structural properties as well as the ground motions are varied and the learning from previous preliminary studies is applied to develop surrogate models which provide accurate predictions throughout the range of the parametric space.

The objective of this preliminary study is to answer the following questions:

1. What is the increase in the number of high-fidelity data points required for training the data-driven surrogate models as the structural properties are also varied?

2. Which parameters capture the properties of the structure as well as the ground motion characteristics well enough to be used as predictors in data-driven surrogate models?

3. How does the variation in structural properties affect the response when compared to the variation in ground motion characteristics? (Is it necessary to consider the randomness in structural properties?)

Design of Experiments

Latin hypercube sampling (LHS)[80] is used to sample the values of structural properties (Young's Modulus and Density of Steel) at which high-fidelity data is to be collected. Using LHS reduces the total number of high-fidelity data points required by selecting points randomly based on the cumulative density function of the parametric space. MATLAB function lhsnorm is used to obtain the points[81]. This function is selected because it is assumed that both Young's Modulus and the Density of Steel are normally distributed. When generating the points, it is assumed that the covariance between Young's Modulus and the Density of Steel is zero. This follows what has been traditionally used in literature when the exact covariance of two structural properties is unknown[19].

The ground motion records are randomly shuffled to reduce any biases. 25 of these ground motions are kept aside for the training set. With the remaining 100 ground records,

four training sets with an increasing number of training points are created. The first set has 25 data points, the second has 50, the third has 75 and the fourth one has 100 data points. These sets have been named Set 25 through 100 respectively. Latin hypercube sampling is run separately for each of these sets ensuring that different sets have different values of Young's Modulus and density of steel. Figures 6.3 and 6.4 show the points sampled for each of these sets on the respective distribution. Figure 6.5 shows the points sampled for the testing data set. Figures 6.6 and 6.7 show the values of peak ground accelerations of ground motions selected for training and testing respectively.

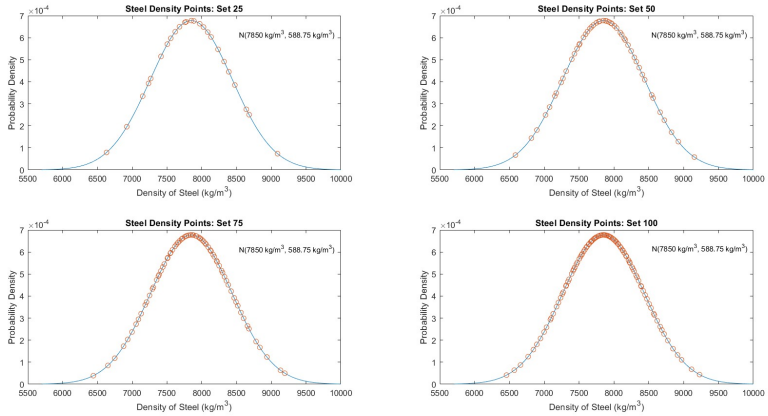


Figure 6.3. Steel Density Points for Different Training Sets

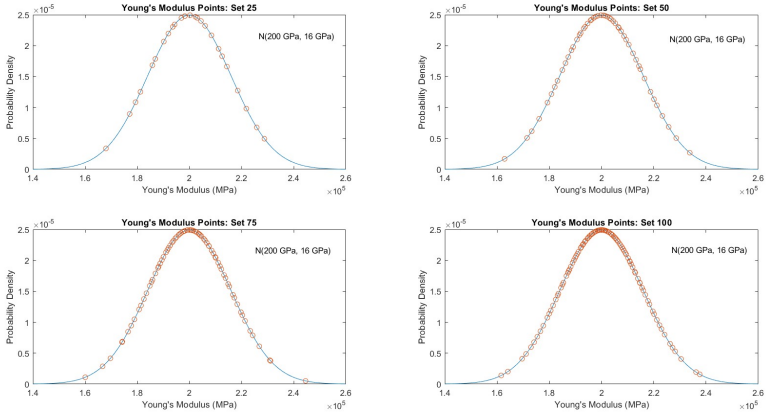
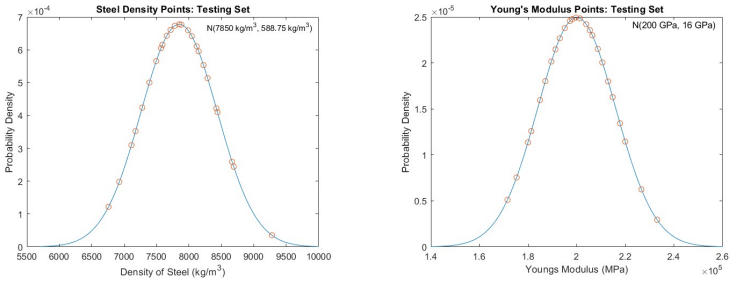


Figure 6.4. Young's Modulus Points for Different Training Sets



(a) Steel Density Points: Testing Set

(b) Young's Modulus Points: Testing Set

Figure 6.5. Sampled Points for Testing Set

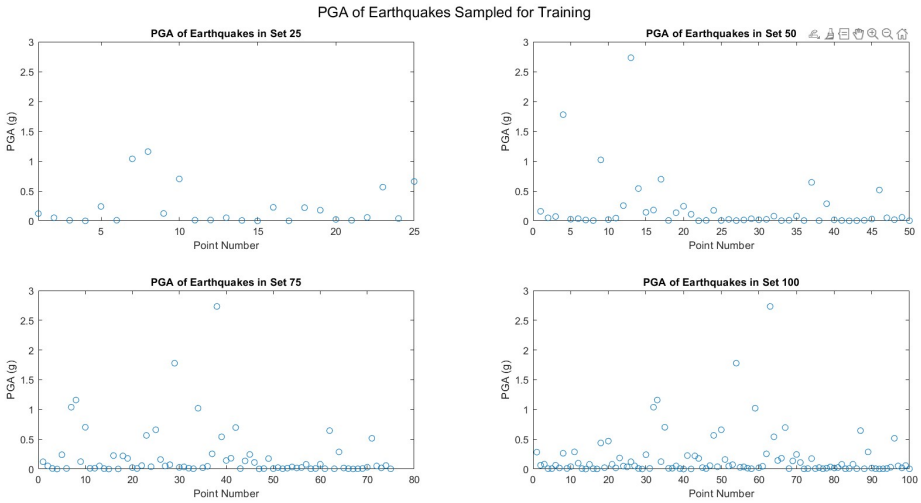


Figure 6.6. Peak Ground Accelerations of Earthquakes Sampled for Training

High-Fidelity Model

Once again, for this preliminary study, the responses generated using the fundamental frequency matched shear building model, developed in 5.3.3 is assumed to be the pseudo-high-fidelity data. The peak structural responses from this model are obtained by first obtaining the time-history response of the shear building using numerical integration (using ODE45 in MATLAB) and then determining the maximum value of the response.

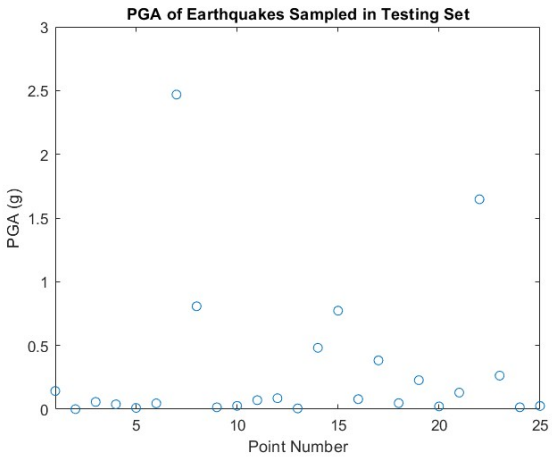


Figure 6.7. Peak Ground Accelerations of Earthquakes Sampled for Testing

Data-Driven Modeling Techniques

Similar to the previous preliminary study 4 data-driven options are tried out using MATLAB's regression learner app: 1) Support Vector Machines (SVM), 2) Gaussian Process Regression (GPR), 3) Artificial Neural Networks (ANN), and 4) Linear Regression (LR). Bayesian optimization with 200 iterations is used to tune the hyperparameters. 5-fold cross-validation is used to validate the model while using all of the points in training and the 25 points testing set is used to test the accuracy and fit of prediction.

In this preliminary study, however, only the set of predictors that performed well in the previous study are used. The two structural properties (Young's modulus and the density of steel) are also used as predictors in the surrogate model along with the selected ground motion parameters. One important consideration here is that the spectral response properties vary with both the structural properties as well as the ground motion characteristics. Therefore there are two options here:

1. Using the spectral response property at the mean values of the structural properties as used in the previous study. Here the spectral response properties do not capture the variation in structural properties. However, this variation is captured by the peak responses and can be provided to the surrogate models by providing Young's modulus

and the density of steel as input parameters if required. The flowchart for this surrogate model is shown in Figure 6.8.

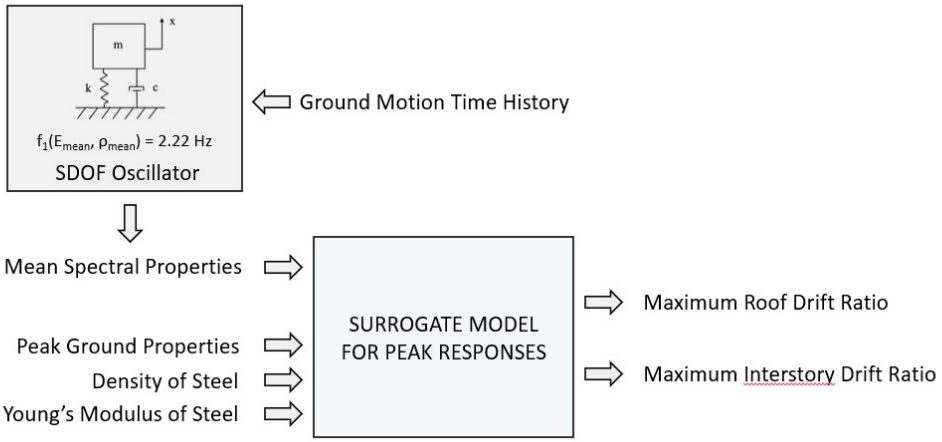


Figure 6.8. Flowchart of the Structure of the Surrogate Models for the Predictions of Peak Structural Responses Using Mean Spectral Response Properties

2. Using the actual spectral response properties by considering the variation of structural properties. This would require estimating the fundamental frequency of the structure first and then creating an SDOF system with this as its natural frequency. The surrogate model for the prediction of frequency that was developed in 5.3.3 is used for the first step. This option is viable since these surrogate models predict the frequencies exactly up to 4 decimal points and at a fast computational speed. The flowchart for this surrogate model is shown in Figure 6.9.

A sensitivity study is conducted to see if the second option described above would be required. For 30 of the 125 ground records, the frequency is varied within the range of interest and the spectral response properties are calculated at each of these frequency values. The coefficient of variation is measured for each of the spectral response properties, for each ground motion and is used to determine if the second option needs to be considered.

Figures 6.10, 6.11, and 6.12 show the variation of spectral acceleration, spectral velocity, and spectral displacement respectively of the structure at various frequencies for 30 ground

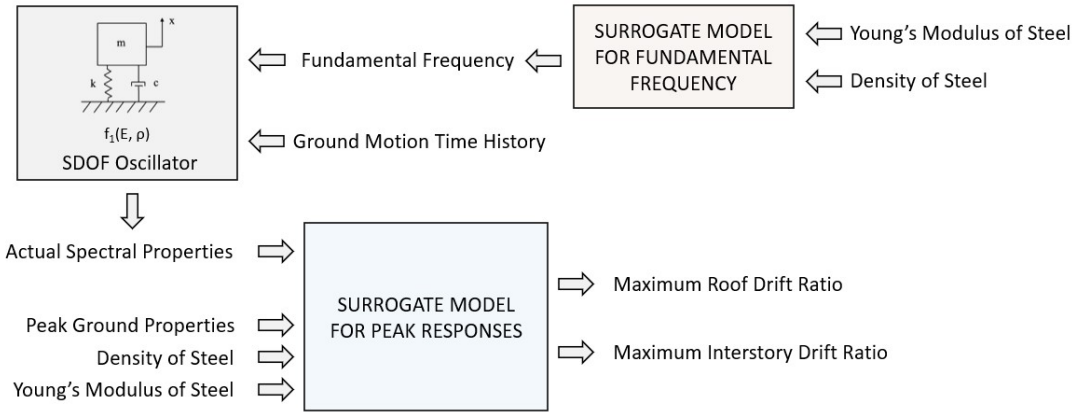


Figure 6.9. Flowchart of the Structure of the Surrogate Models for the Predictions of Peak Structural Responses Using Actual Spectral Response Properties

motions. The spectral velocity and spectral displacement plotted here are obtained using an equivalent SDOF system of the scaled model and hence these values are inherently scaled in a similar manner. Figure 6.13 shows the coefficients of variations against the fundamental frequency for different earthquakes. These plots show a moderate variation in spectral response properties as the fundamental frequency of the structure is changed within the range of sampled structural properties. This variation is different for different earthquake records as well. Therefore, it is worth studying the differences in determining the actual spectral response properties and using that as inputs to surrogate models against using the spectral response properties at mean structural properties as the inputs to surrogate models.

Results

The results for individual models for various combinations of input parameters for predicting maximum roof drift ratio and maximum story drift ratio for preliminary study 2 are provided in Appendix B and C. Appendix B lists the results when using actual spectral

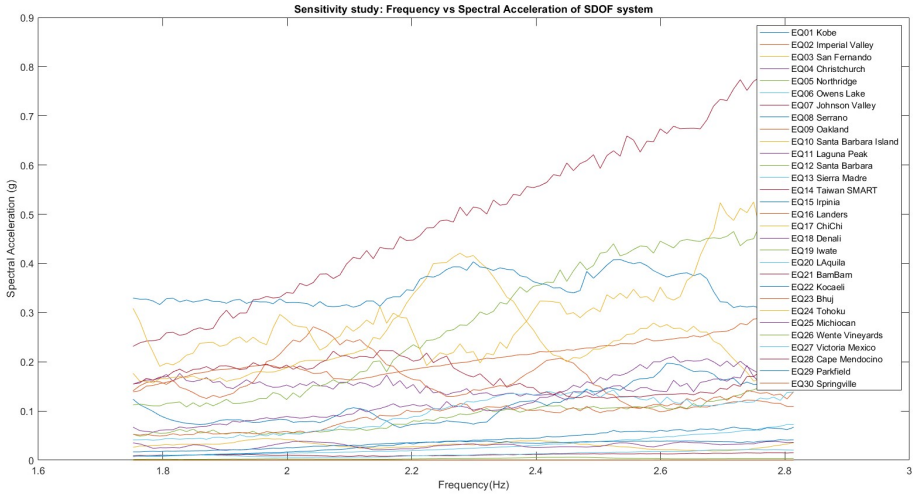


Figure 6.10. Spectral Acceleration vs Fundamental Frequency

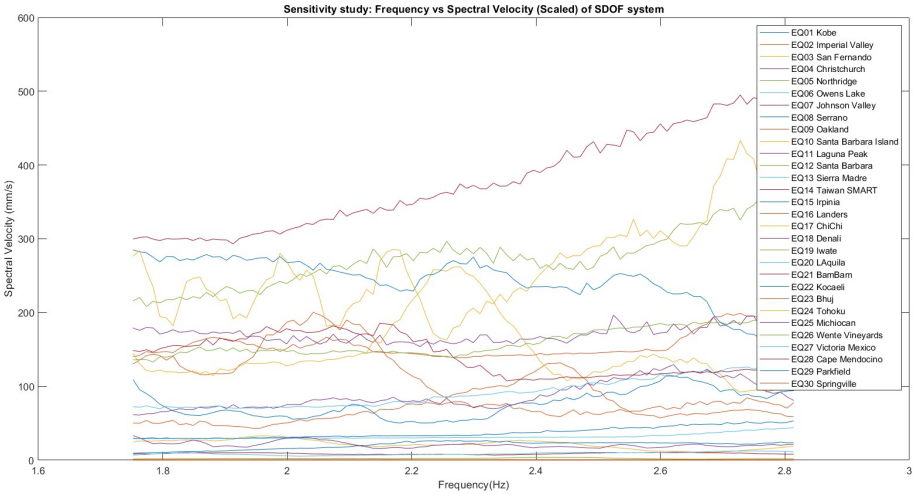


Figure 6.11. Spectral Velocity vs Fundamental Frequency

response properties as inputs, while Appendix C lists the results when using mean spectral response properties.

A summary of these results, listing the best models for each combination of input parameters, is provided in Tables 6.3 to 6.6.

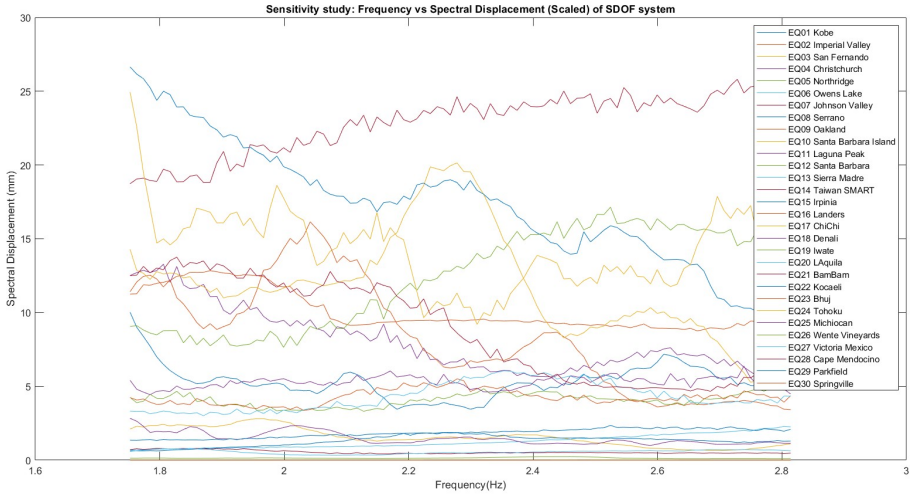
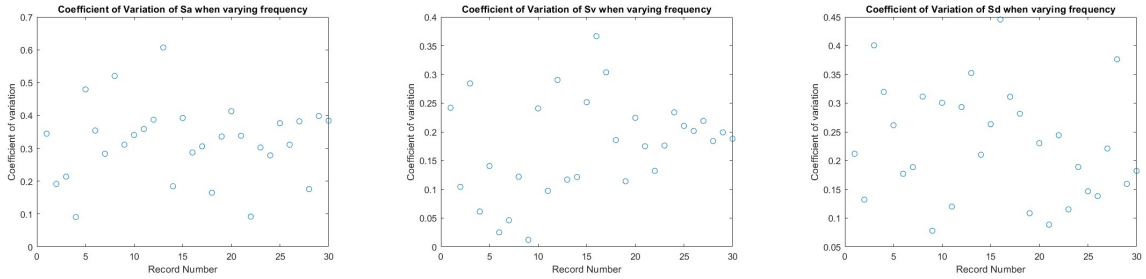


Figure 6.12. Spectral Displacement vs Fundamental Frequency



(a) Coefficients of Variation of Spectral Acceleration with Frequency for 30 Earthquakes

(b) Coefficients of Variation of Spectral Velocity with Frequency for 30 Earthquakes

(c) Coefficients of Variation of Spectral Displacement with Frequency for 30 Earthquakes

Figure 6.13. Coefficients of Variation of Spectral Response Properties with Fundamental Frequency for 30 Earthquakes

Discussion

The following is the discussion of the results obtained for the prediction of maximum roof drift ratio using data-driven surrogate in preliminary study 2 with linear high-fidelity data and randomness in structural properties.

- When using actual spectral parameters as inputs, there was no significant difference in performance by providing structural properties as inputs to the surrogate models

Table 6.3. Data-Driven Model Results: Varying Structural Properties - Maximum Roof Drift Ratio - Actual Spectral Response Properties

Inputs Parameters	Points Req	Best Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
$S_{a\ act}$	50	LR	0.40	0.97	0.43	0.95
E, ρ, $S_{a\ act}$	50	SVM	0.19	0.99	0.32	0.97
$S_{d\ act}$	25	LR	0.23	0.99	0.39	0.96
E, ρ, $S_{d\ act}$	25	GPR	0.24	0.99	0.48	0.95
PGA, $S_{a\ act}$	100	LR	0.09	0.98	0.57	0.95
E, ρ, PGA, $S_{a\ act}$	50	ANN	0.27	0.98	0.39	0.96
PGA, $S_{d\ act}$	25	LR	0.18	0.99	0.35	0.97
E, ρ, PGA, $S_{d\ act}$	25	LR	0.19	0.99	0.33	0.97
$S_{a\ act}$, $S_{v\ act}$	50	GPR	0.19	0.99	0.39	0.96
E, ρ, $S_{a\ act}$, $S_{v\ act}$	50	LR	0.20	0.99	0.23	0.99
$S_{a\ act}$, $S_{d\ act}$	25	LR	0.21	0.99	0.33	0.97
E, ρ, $S_{a\ act}$, $S_{d\ act}$	25	LR	0.27	0.98	0.37	0.97
$S_{v\ act}$, $S_{d\ act}$	25	LR	0.24	0.99	0.33	0.97
E, ρ, $S_{v\ act}$, $S_{d\ act}$	25	LR	0.19	0.99	0.25	0.99
PGA, $S_{a\ act}$, $S_{v\ act}$	50	LR	0.31	0.98	0.42	0.96
E, ρ, PGA, $S_{a\ act}$, $S_{v\ act}$	50	LR	0.29	0.98	0.33	0.98
PGA, $S_{a\ act}$, $S_{d\ act}$	25	LR	0.15	0.995	0.36	0.97
E, ρ, PGA, $S_{a\ act}$, $S_{d\ act}$	25	LR	0.18	0.995	0.36	0.97
PGA, $S_{v\ act}$, $S_{d\ act}$	25	LR	0.21	0.99	0.32	0.98
E, ρ, PGA, $S_{v\ act}$, $S_{d\ act}$	25	LR	0.27	0.98	0.26	0.98
$S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	25	LR	0.25	0.98	0.26	0.98
E, ρ, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	25	LR	0.21	0.99	0.32	0.98
PGA, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	25	LR	0.21	0.99	0.35	0.97
E, ρ, PGA, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	25	LR	0.20	0.99	0.36	0.97

in addition to the ground motion parameters as compared to when the structural properties were not provided as inputs. Figure 6.14 shows the testing NRMSE values for the prediction of maximum roof drift ratio using surrogate models with and without structural properties as inputs alongside actual spectral response properties.

This could mean two things: 1) the spectral parameters used as inputs capture the changes in structural properties, therefore providing structural properties as inputs adds no extra information, or 2) randomness in structural inputs does not have a

Table 6.4. Data-Driven Model Results: Varying Structural Properties - Maximum Roof Drift Ratio - Mean Spectral Response Properties

Inputs Parameters	**Points Req**	**Best Model**	**Val NRMSE**	**Val R^2**	**Test NRMSE**	**Test R^2**
S_a	25	LR	0.45	0.95	0.30	0.98
E, ρ, S_a	50	LR	0.43	0.95	0.31	0.98
S_d	25	GPR	0.39	0.97	0.25	0.99
E, ρ, S_d	25	LR	0.39	0.96	0.31	0.98
PGA, S_a	25	SVM	0.10	0.998	0.20	0.99
E, ρ, PGA, S_a	25	SVM	0.27	0.98	0.30	0.98
PGA, S_d	25	ANN	0.25	0.99	0.22	0.99
E, ρ, PGA, S_d	25	LR	0.28	0.98	0.30	0.98
S_a, S_v	25	LR	0.32	0.98	0.32	0.98
E, ρ, S_a, S_v	25	LR	0.32	0.97	0.32	0.98
S_a, S_d	25	LR	0.38	0.98	0.35	0.97
E, ρ, S_a, S_d	25	ANN	0.29	0.98	0.31	0.98
S_v, S_d	25	ANN	0.46	0.95	0.24	0.99
E, ρ, S_v, S_d	25	LR	0.32	0.98	0.32	0.98
PGA, S_a, S_v	25	LR	0.30	0.98	0.29	0.98
E, ρ, PGA, S_a, S_v	25	GPR	0.25	0.99	0.26	0.98
PGA, S_a, S_d	25	SVM	0.23	0.99	0.20	0.99
E, ρ, PGA, S_a, S_d	25	LR	0.31	0.98	0.29	0.98
PGA, S_v, S_d	25	LR	0.19	0.99	0.33	0.98
E, ρ, PGA, S_v, S_d	25	LR	0.31	0.98	0.29	0.98
S_a, S_v, S_d	25	ANN	0.40	0.96	0.25	0.99
E, ρ, S_a, S_v, S_d	25	LR	0.39	0.96	0.29	0.98
PGA, S_a, S_v, S_d	25	ANN	0.32	0.98	0.25	0.99
E, ρ, PGA, S_a, S_v, S_d	25	SVM	0.28	0.98	0.34	0.97

significant effect on the maximum roof drift ratio and therefore, the surrogate models developed without structural properties as inputs can make similar predictions as the ones with structural properties as inputs.

- Similar observation was seen when using mean spectral parameters as inputs. Here, since mean structural properties are being used, they for sure do not capture the randomness in structural properties. Therefore, it can be concluded that the randomness in structural properties does not affect the maximum roof drift ratio significantly

Table 6.5. Data-Driven Model Results: Varying Structural Properties - Maximum story Drift Ratio - Actual Spectral Response Properties

Inputs Parameters	Points Req	Best Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
$S_{a\ act}$, $S_{v\ act}$	100	ANN	0.47	0.94	0.80	0.91
E, ρ, $S_{a\ act}$, $S_{v\ act}$	100	GPR	0.40	0.96	0.82	0.88
$S_{v\ act}$, $S_{d\ act}$	100	ANN	0.33	0.97	0.46	0.96
E, ρ, $S_{v\ act}$, $S_{d\ act}$	100	ANN	0.41	0.96	0.56	0.90
PGA, $S_{v\ act}$, $S_{d\ act}$	100	ANN	0.28	0.98	0.73	0.90
E, ρ, PGA, $S_{v\ act}$, $S_{d\ act}$	100	SVM	0.29	0.98	0.73	0.91
$S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	100	ANN	0.41	0.96	0.69	0.92
E, ρ, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	100	GPR	0.38	0.71	0.84	0.88
PGA, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	100	SVM	0.24	0.98	0.72	0.91
E, ρ, PGA, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	100	ANN	0.29	0.98	0.72	0.91

Table 6.6. Data-Driven Model Results: Varying Structural Properties - Maximum story Drift Ratio - Mean Spectral Response Properties

Inputs Parameters	Points Req	Best Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
S_a, S_v	100	SVM	0.49	0.94	0.78	0.90
E, ρ, S_a, S_v	100	ANN	0.63	0.90	0.89	0.86
S_v, S_d	100	SVM	0.48	0.94	0.72	0.91
E, ρ, S_v, S_d	100	SVM	0.55	0.92	0.74	0.90
PGA, S_v, S_d	100	SVM	0.50	0.94	0.78	0.90
E, ρ, PGA, S_v, S_d	100	GPR	0.37	0.97	0.72	0.91
S_a, S_v, S_d	100	ANN	0.47	0.94	0.69	0.92
E, ρ, S_a, S_v, S_d	100	SVM	0.58	0.91	0.75	0.90
PGA, S_a, S_v, S_d	100	ANN	0.41	0.96	0.74	0.90
E, ρ, PGA, S_a, S_v, S_d	100	ANN	0.21	0.96	0.76	0.90

when compared to randomness in ground motion records. Figure 6.15 shows the testing NRMSE for the prediction of maximum roof drift ratio using surrogate models with and without structural properties as inputs alongside mean spectral response properties.

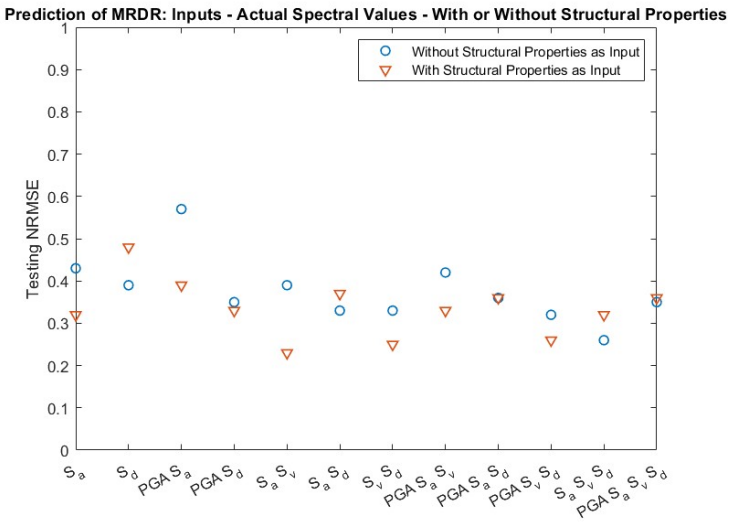


Figure 6.14. Testing Performance: MRDR: Actual Spectral Response Properties - With and Without Structural Properties

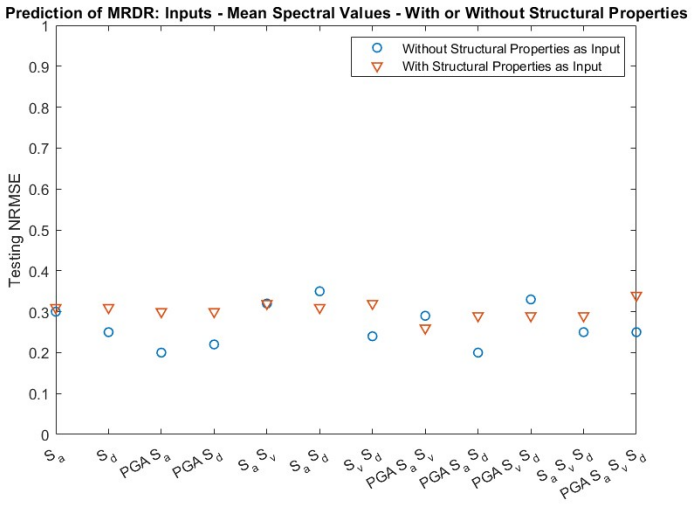
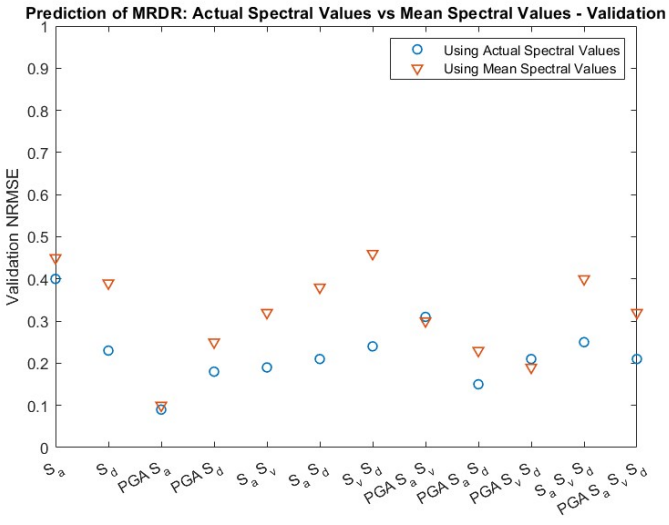


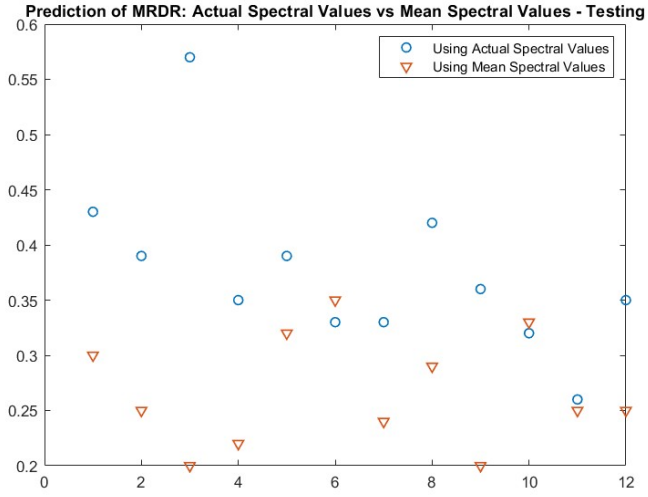
Figure 6.15. Testing Performance: MRDR: Mean Spectral Response Properties - With and Without Structural Properties

- Using mean spectral response properties provided similar, if not slightly better performance than using actual spectral response properties. This reinforces the conclusion in the previous point that the randomness in structural properties does not have a significant enough influence on the maximum roof drift ratio, therefore, not capturing

this randomness does not result in a reduction in the prediction error. Figure 6.16 compares the NRMSE in validation and testing for the prediction of maximum roof drift ratio when using actual spectral response properties to when using mean spectral response properties.



(a) Validation NRMSE



(b) Testing NRMSE

Figure 6.16. Performance of Data-Driven Surrogate Model: MRDR: Using Actual Spectral Response Properties vs Using Mean Spectral Response Properties

- It is seen that the best predictor for the prediction of the maximum roof drift ratio is spectral displacement.

- Increasing the number of inputs to the surrogate model while keeping the size of the training data set the same, reduces the errors in general as expected. This is because each input parameter captures different features of the ground motion excitation and the structural response. The only time prediction NRMSE increases on adding another input to the surrogate model is when PGA is the new input added. This is shown in Figure 6.17

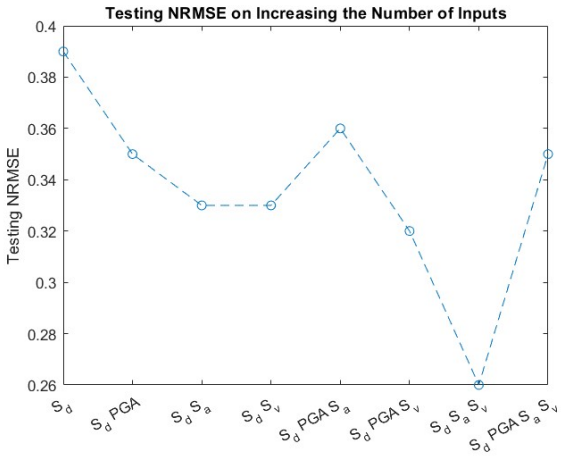


Figure 6.17. Testing NRMSE on Increasing the Number of Inputs to the Data-Driven Surrogate Model Using Actual Spectral Response Properties

- There is virtually no increase in the number of points required when considering the randomness in structural properties when opposed to not considering this randomness for the prediction of the maximum roof drift ratio.

- It is seen that linear regression provides the best fit for data in most cases for the prediction of the maximum roof drift ratio in this study. This is expected because the high-fidelity data chosen for this preliminary study is linear.

The following is the discussion of the results obtained for the prediction of maximum story drift ratio using data-driven surrogates in preliminary study 2 with linear high-fidelity data and randomness in structural properties.

- Like was the case for the prediction of maximum roof drift ratio, for the prediction of maximum story drift ratio too, when using actual spectral parameters as inputs, there was no significant difference in performance by providing structural properties as inputs to the surrogate models in addition to the ground motion parameters as compared to when the structural properties were not provided as inputs. Figure 6.18 shows the testing NRMSE values for the prediction of the maximum story drift ratio for surrogate models with and without structural properties as inputs alongside the actual spectral response properties.

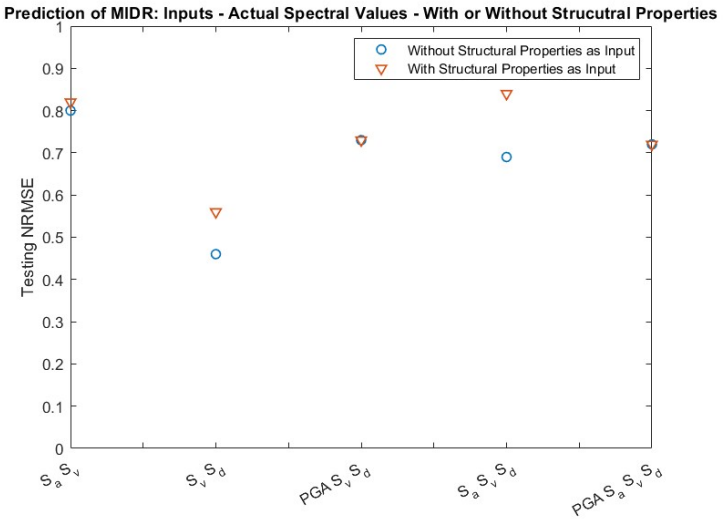


Figure 6.18. Testing Performance: MSDR: Actual Spectral Response Properties - With and Without Structural Properties

- Similar observation was seen when using mean spectral parameters as inputs. Here, since mean structural properties are being used, they for sure do not capture the randomness in structural properties. Therefore, it can be concluded that the randomness in structural properties does not affect the maximum roof drift ratio significantly when compared to randomness in ground motion records. Figure 6.19 shows the testing NRMSE for the prediction of maximum story drift ratio using surrogate models with and without structural properties as inputs alongside mean spectral response properties.

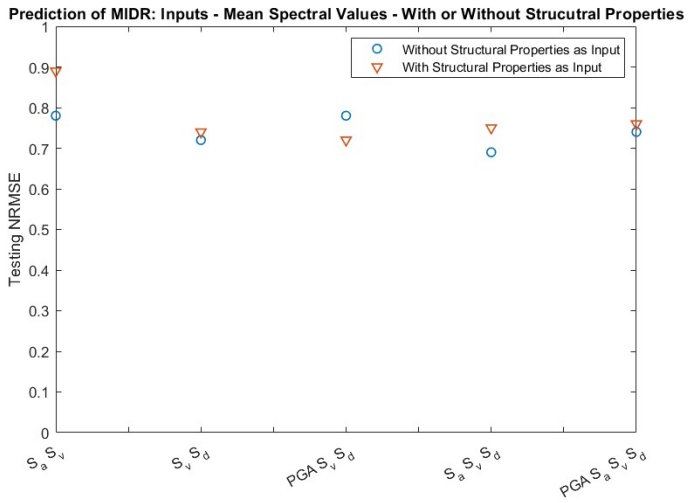
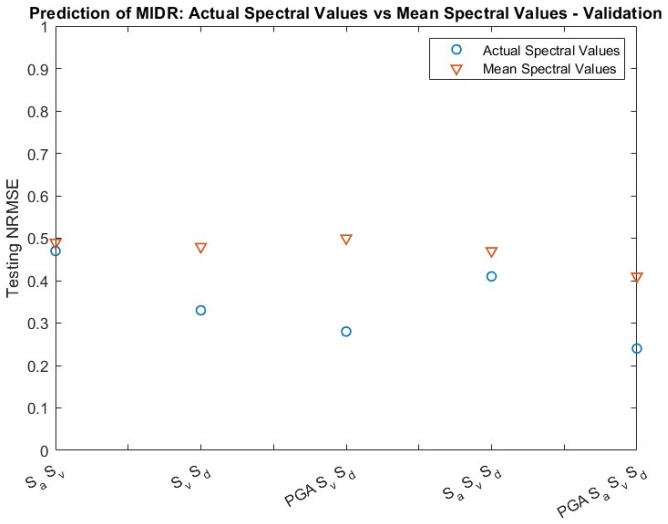


Figure 6.19. Testing Performance: MSDR: Mean Spectral Response Properties - With and Without Structural Properties

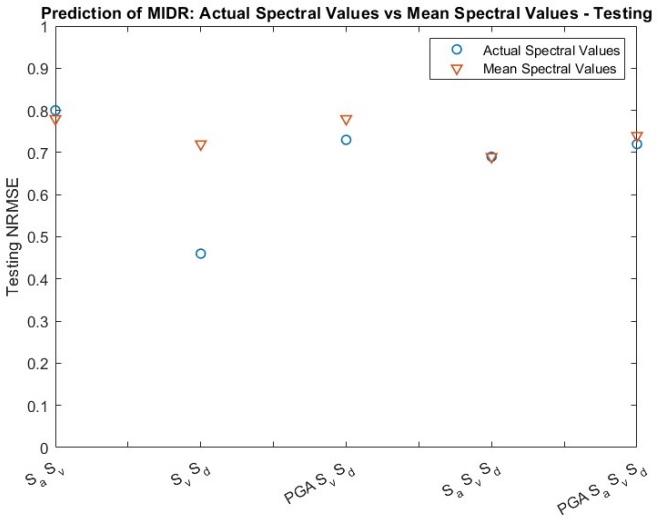
- Using mean spectral response properties provided similar, if not slightly better performance than using actual spectral response properties. This reinforces the conclusion in the previous point that the randomness in structural properties does not have a significant enough influence on the maximum roof drift ratio, therefore, not capturing this randomness does not result in a reduction in the prediction error. Figure 6.20 compares the NRMSE in validation and testing for the prediction of maximum story drift ratio when using actual spectral response properties to when using mean spectral response properties.

- The number of points required in training is significantly increased when randomness in structural properties is incorporated into the problem. In fact, 100 points are not enough to achieve the target performance.

- Once again, it is seen that it is more difficult to predict the maximum story drift ratio than to predict the maximum roof drift ratio. This is expected because there is an additional level of complexity added in the prediction of maximum story drift because this drift can occur in any story as opposed to the maximum roof drift ratio.



(a) Validation NRMSE



(b) Testing NRMSE

Figure 6.20. Performance of Data-Driven Surrogate Model: MSDR: Using Actual Spectral Response Properties vs Using Mean Spectral Response Properties

- For the prediction of maximum story drift it is observed that linear regression does not perform as well as other methods used, implying that although the model is linear (or stepwise linear- by using a lumped mass model, it is assumed that each story deforms

linearly), there is some level of nonlinearity added to the prediction (probably because of the multiple different stories at which maximum story drift may occur).

Findings

This subsection lists the findings from the second preliminary study, where we used pseudo-high-fidelity data while considering randomness in structural properties as well as ground motions.

The findings for the prediction of the maximum roof drift ratio are:

- For both the prediction of the maximum roof drift ratio as well as the prediction of the maximum story drift ratio, providing the structural properties along with either the actual or mean spectral response properties as inputs did not make a significant difference in the prediction performance.
- Similarly, no appreciable differences in prediction performance was found when using the actual spectral response properties as inputs vs using mean spectral response properties as inputs.
- The number of ground motions required for obtaining an accurate surrogate model remained 25 for the prediction of the maximum roof drift ratio, but it increased significantly for the prediction of the maximum story drift ratio.

6.1.3 Full Study

Now that we have completed all the preliminary studies, and have answered the questions we needed to before generating the actual high-fidelity data, we can finally use the actual high-fidelity data and generate data-driven models to predict the peak structural responses. Findings from all the preliminary studies are used for the design of experiments and to select predictor parameters for the data-driven models.

The main objective of this study is to answer the following question:

1. How many high-fidelity data points are required to train data-driven surrogate models for the prediction of MRDR and MSDR with acceptable performance?

2. What is the increase in the number of points when using actual nonlinear high-fidelity data from the number of points when using linear pseudo-high-fidelity data?

3. Is the trend regarding the best predictor variables similar when using nonlinear high-fidelity data as when using linear high-fidelity data?

Design of Experiments

The same points as used for the second preliminary study (subsection 6.1.2) are used in this study. Doing this facilitates a direct comparison between the surrogate models obtained training on linear pseudo-high-fidelity data to the surrogate models obtained training on nonlinear high-fidelity data.

Therefore, the lhsnorm function in MATLAB is used to sample the structural properties using Latin hypercube sampling. 25 ground motions are kept aside for testing. The remaining ground motions are grouped into four training sets, with the first set having 25 ground motions and each subsequent set having 25 ground motions more than the previous set. The number of ground motion sin training set is limited to 100 because the objective of the study in this thesis is to reduce computational time to obtain estimates of seismic responses and generating more than 100 points of high-fidelity data is a computationally expensive task in itself.

High-Fidelity Model

The validated 2D nonlinear Abaqus model presented in Chapter 4 is used to obtain high-fidelity data for these values. The peak structural responses are obtained by carrying out a nonlinear dynamic analysis in Abaqus and then finding the maximum value of the responses.

Data-Driven Modeling Techniques

Similar to the preliminary studies, 4 data-driven options are tried out using MATLAB's regression learner app: 1) Support Vector Machines (SVM), 2) Gaussian Process Regression (GPR), 3) Artificial Neural Networks (ANN), and 4) Linear Regression (LR). Bayesian

optimization with 200 iterations is used to tune the hyperparameters. 5-fold cross-validation is used to validate the model while using all of the points in training and the 25 points testing set is used to test the accuracy and fit of prediction.

Findings from the preliminary studies are used to select the right combinations of predictor parameters for the surrogate models and to estimate the number of points required in training. In the preliminary studies, it was found that using mean spectral response properties provided similar results as using actual spectral response properties. It was also found that providing the structural properties as inputs did not improve the performance of the surrogate models significantly. Therefore, mean spectral response properties are used as predictor variables, and structural properties are not provided as inputs.

Results

The results for individual data-driven surrogate models for various combinations of input parameters for the prediction of the maximum roof drift ratio and the maximum story drift ratio using actual high-fidelity data are provided in Appendix D. A summary of these results, listing the best models for each combination of input parameters, is provided in Table 6.7 for maximum roof drift ratio and in Table 6.8 for story drift ratio.

6.2 Discussion

The discussions regarding the prediction of maximum roof drift ratio using data-driven surrogate models for responses generated using actual high-fidelity data and randomness in both structural properties and ground motions are provided below:

- The number of ground motions required in training to obtain acceptable performance from surrogate models did not increase from 25 on selecting nonlinear high-fidelity data over the linear pseudo-high-fidelity data used in the preliminary studies.
- The trend regarding the best predictor variables followed the same trend seen in the preliminary studies. Spectral displacement is the best intensity measure to condition the maximum roof drift ratio, followed by spectral acceleration.

Table 6.7. Data-Driven Model Results: Varying Structural Properties - Maximum Roof Drift Ratio

Inputs Parameters	Points Req	Best Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
$S_{a\ mean}$	25	ANN	0.10	0.998	0.32	0.98
$S_{d\ mean}$	25	ANN	0.11	0.998	0.25	0.99
PGA, $S_{d\ mean}$	25	SVM	0.16	0.99	0.24	0.99
$S_{a\ mean}$, $S_{d\ mean}$	25	LR	0.10	0.998	0.25	0.99
$S_{a\ mean}$, $S_{v\ mean}$, $S_{d\ mean}$	25	GPR	0.18	0.99	0.31	0.98
PGA, $S_{a\ mean}$, $S_{v\ mean}$, $S_{d\ mean}$	25	SVM	0.09	0.998	0.25	0.99
$S_{d\ act}$	25	ANN	0.12	0.997	0.41	0.96
PGA, $S_{d\ act}$	25	GPR	0.09	0.998	0.38	0.97
$S_{a\ act}$, $S_{d\ act}$	25	SVM	0.30	0.98	0.39	0.96
$S_{v\ act}$, $S_{d\ act}$	25	ANN	0.21	0.99	0.40	0.96
$S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	25	LR	0.17	0.99	0.34	0.97
PGA, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$	25	ANN	0.18	0.99	0.5	0.95

Table 6.8. Data-Driven Model Results: Varying Structural Properties - Maximum story Drift Ratio

Inputs Parameters	Points Req	Best Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
$S_{v\ mean}$	100	ANN	0.75	0.86	0.78	0.89
$S_{a\ mean}$, $S_{v\ mean}$	100	ANN	0.75	0.86	0.81	0.88
$S_{v\ mean}$, $S_{d\ mean}$	100	ANN	0.71	0.87	0.76	0.90
PGA, $S_{v\ mean}$, $S_{d\ mean}$	100	SVM	0.53	0.93	0.70	0.91
$S_{a\ mean}$, $S_{v\ mean}$, $S_{d\ mean}$	100	ANN	0.78	0.85	0.58	0.94
PGA, $S_{a\ mean}$, $S_{v\ mean}$, $S_{d\ mean}$	100	ANN	0.61	0.91	0.75	0.90
$S_{v\ actual}$	100	ANN	0.60	0.91	0.76	0.90
$S_{a\ actual}$, $S_{v\ actual}$	100	SVM	0.58	0.92	0.87	0.86
$S_{v\ actual}$, $S_{d\ actual}$	100	GPR	0.57	0.92	0.82	0.88
PGA, $S_{v\ actual}$, $S_{d\ actual}$	100	SVM	0.34	0.97	1.20	0.74
$S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$	100	LR	0.66	0.89	0.77	0.89
PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$	100	SVM	0.44	0.95	1.40	0.65

- Using mean structural properties provided better performance than using actual spectral response properties. While the surrogate models developed using actual spectral response properties also provide satisfactory performance, it is seen that there is an overfit on the training data in these models, thus reducing the accuracy in testing. Therefore, increasing the number of points in the testing dataset could potentially change this trend.

  The biggest difference between the two approaches is the ability to capture uncertainty in structural properties. When using surrogate models developed from the first approach with mean spectral response properties, the estimates of the maximum roof drift ratios will be the same for a given ground motion, irrespective of the structural properties. When using the surrogate model developed using the actual spectral response properties, the surrogate model will be able to capture the changes in the response due to the randomness in structural properties but would have a higher error in prediction than surrogate models developed using the first approach.

- Unlike the preliminary studies, here the machine learning techniques with the capability to fit nonlinear data performed better than linear regression. This is expected, because the high-fidelity data here is nonlinear, unlike in the preliminary studies.

The discussions regarding the prediction of maximum story drift ratio using data-driven surrogate models for responses generated using actual high-fidelity data and randomness in both structural properties and ground motions are provided below:

- For the prediction of maximum story drift ratio, it is seen that the best performance obtained using the selected data-driven methods is an NRMSE of 0.70 and a R^2 of 0.91 in testing for 100 ground motions in training. While this is still good, it does not meet the performance target that we had set beforehand.

- The trend regarding the best predictor variables for the prediction of maximum story drift ratio follows the same trend as in the preliminary study. Spectral velocity is the best predictor variable, and adding more predictor variables increases the accuracy in general.

- In most cases, for the prediction of the maximum story drift ratio, it is seen that ANN provided the best results. A possible reason for this is that ANNs have the capability of capturing nonlinearity in data without overfitting the training data.
- Similar to the prediction of maximum roof drift ratio, here too it is seen that using the actual spectral response properties causes overfitting of the training data, resulting in poor testing performance.

Findings

This subsection lists the findings from the full study, where we used high-fidelity data while considering randomness in structural properties as well as ground motions.

- 25 training points are enough to obtain data-driven surrogate models with acceptable levels of performance for the prediction of maximum roof drift ratio.
- Spectral displacement is the best parameter for the prediction of maximum roof drift ratio. Spectral acceleration is also a good parameter for this.
- For the prediction of maximum story drift ratio, a training set with 100 points is not enough to obtain targeted performance.
- Spectral velocity is the best parameter for the prediction of maximum story drift ratio.
- Using mean spectral response properties provides better performance over actual spectral response properties.

6.3 Physics-Based Surrogate Techniques

This section describes the physics-based surrogate models developed to predict the peak structural responses for the select building. In this study, the physics-based surrogate models are developed by using a simplified physics-based model to generate low-fidelity data and then by creating a map between the low-fidelity data and the high-fidelity data.

6.3.1 Design of Experiments

Latin hypercube sampling is used to sample the structural properties for developing physics-based surrogate models for the prediction of MRDR and MSDR. A set with 25 points across the structural properties and 25 ground motions is kept aside for testing. Three separate sets with 10, 25, and 100 data points are used for training the physics-based surrogate models. The testing set as well as the training sets with 25 and 100 data points are kept exactly the same as the sets used in the development of data-driven surrogate models presented in Section 6.1 to facilitate direct comparison between the two types of surrogate modeling techniques.

6.3.2 High-Fidelity Model

The 2D nonlinear Abaqus model described in Chapter 4 is used to generate high-fidelity data points for the peak structural responses in this study. The peak structural responses are obtained by carrying out a nonlinear dynamic analysis in Abaqus and then finding the maximum value of the responses.

6.3.3 Physics-Based Modeling Techniques

The shear-building model with modifications to the mass matrix to account for a portion of the mass of the columns, described in Chapter 5 is used as a low-fidelity model to generate low-fidelity data for MRDR and MSDR. The discrete system of equations of motion for the multi-degree of freedom model is solved using numerical integration (ODE45 in MATLAB). The floor displacements obtained in this manner are then converted to the required drift ratios.

Once the low-fidelity points are obtained for the training set, these points are mapped to high-fidelity data points by using a machine learning technique. This technique is called response correction using space mapping[82]. Four options are used to map the low-fidelity data to high-fidelity data: 1) Support Vector Machines (SVM), 2) Gaussian Process Regression (GPR), 3) Artificial Neural Networks (ANN), and 4) Linear Regression (LR). Bayesian

optimization with 200 iterations is used to tune the hyperparameters. 5-fold cross-validation is used to validate the model while using all of the points in training and the 25 points testing set is used to test the accuracy and fit of prediction.

6.3.4 Results

The results for the performance of the developed physics-based surrogate models for the prediction of the maximum roof drift ratio are provided in Tables 6.9 to 6.14 for increasing number of training points.

Table 6.9. Physics-Based Model Results: MRDR - Inputs: Low-Fidelity Estimate of MRDR - 10 Points

Mapping	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.10	0.998	0.04	0.999
GPR	0.09	0.998	0.03	1.000
ANN	0.12	0.997	0.03	1.000
Linear Regression	0.09	0.998	0.04	1.000

Table 6.10. Physics-Based Model Results: MRDR - Inputs: Low-Fidelity Estimate of MRDR - 25 Points

Mapping	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.25	0.99	0.07	0.999
GPR	0.25	0.97	0.08	0.998
ANN	0.24	0.99	0.06	0.999
Linear Regression	0.18	0.993	0.07	0.999

Table 6.11. Physics-Based Model Results: MRDR - Inputs: Low-Fidelity Estimate of MRDR - 100 Points

Mapping	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.33	0.98	0.17	0.99
GPR	0.36	0.97	0.11	0.998
ANN	0.32	0.98	0.36	0.97
Linear Regression	0.36	0.97	0.03	0.9999

Table 6.12. Physics-Based Model Results: MSDR - Inputs: Low-Fidelity Estimate of MSDR - 10 Points

Mapping	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.10	0.995	0.21	0.98
GPR	0.12	0.994	0.78	0.79
ANN	0.12	0.993	0.18	0.99
Linear Regression	0.13	0.994	0.18	0.99

Table 6.13. Physics-Based Model Results: MSDR - Inputs: Low-Fidelity Estimate of MSDR - 25 Points

Mapping	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.24	0.98	0.23	0.99
GPR	0.29	0.97	0.20	0.99
ANN	0.17	0.99	0.29	0.98
Linear Regression	0.22	0.98	0.21	0.99

Table 6.14. Physics-Based Model Results: MSDR - Inputs: Low-Fidelity Estimate of MSDR - 100 Points

Mapping	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.28	0.98	0.19	0.99
GPR	0.37	0.97	0.25	0.99
ANN	0.34	0.97	0.18	0.99
Linear Regression	0.30	0.98	0.19	0.99

6.3.5 Discussion

From Tables 6.9 and 6.12 the physics-based surrogate models trained on as little as ten high-fidelity data points perform well for the estimation of both the maximum roof drift ratio and the maximum story drift ratio. As the number of points in the training dataset is increased, the error in validation increases marginally. This could be the result of an increase in the variability of ground motions as the number of ground motions increases. It is noticed that the overall mean of the maximum roof and the maximum story drift ratios decreased when moving from set with 25 points to the set with 100 points. This could be a limitation

of selecting ground motions for the training set in random. However, the performance on the testing set remains the same on increasing the number of training data points.

This result shows that the low-fidelity shear building model captures most of the characteristics of the response of the building to ground motion excitation. Given this, it is easy to map the results obtained from the low-fidelity model to the high-fidelity responses. Also, it is seen that all of the four machine learning techniques used to map the results, including linear regression, provide similar results. This shows that it is a simple map to learn and does not require a large training data set to do so.

Since the low-fidelity model captures the effects due to randomness in structural properties, this too is reflected in the prediction. Although this study is conducted on a symmetric building with no variations in geometry, it can be expected that this method would perform similarly on irregular structures with varying geometric as well as structural properties, as the effects of these variations and irregularities can be captured by the low-fidelity model.

6.3.6 Findings

This subsection lists the findings from

- As low as 10 training points are enough to obtain physics-based surrogate models for predicting both, the maximum roof drift ratio and the maximum story drift ratio with an NRMSE of less than 0.10 and a R^2 value of over 0.99 in testing.
- Any machine learning technique can be used to learn the map between the results from the low-fidelity results and the high-fidelity responses.

6.4 Comparison between Data-Driven and Physics-Based Surrogates

This section compares the data-driven and physics-based surrogates developed in this chapter for the prediction of peak structural responses. The performance, computational cost, and capabilities of the two types of surrogate models.

6.4.1 Comparison of Performance

Table 6.15 shows the best data-driven and physics-based surrogate models obtained for the prediction of maximum roof drift ratio. From the table, it can be seen that both the data-driven and physics-based surrogate models for the estimation of maximum roof drift ratio perform similarly when trained on a training set with 25 points. However, the physics-based model trained on a training set with just 10 points also shows a similar performance to the data-driven model trained on 25 points.

Table 6.15. Data-Driven Surrogates vs Physics-Based Surrogates for Prediction of MRDR: Performance

Model Type	Number of Training Points	Validation NRMSE	Validation R^2	Testing NRMSE	Testing R^2
Data-Driven	25	0.09	0.998	0.25	0.99
Physics-Based	10	0.09	0.998	0.03	1.000
Physics-Based	25	0.24	0.99	0.06	0.999

Table 6.17 shows the best data-driven and physics-based surrogate models obtained for the prediction of maximum story drift ratio. For the prediction of maximum story drift ratio, the physics-based surrogate model trained on 100 points of training data performs much better than the data-driven surrogate model trained on the same number of training data points. The physics-based model trained on just 10 training data points provides better estimates for the maximum story drift ratio than the data-driven model trained on 100 training data points. Both of the physics-based surrogate models satisfy the performance target set beforehand, however, the data-driven surrogate model trained on 100 training points does not. Therefore, for the prediction of maximum story drift ratio the physics-based surrogate models perform better than the data-driven surrogate models developed in this study.

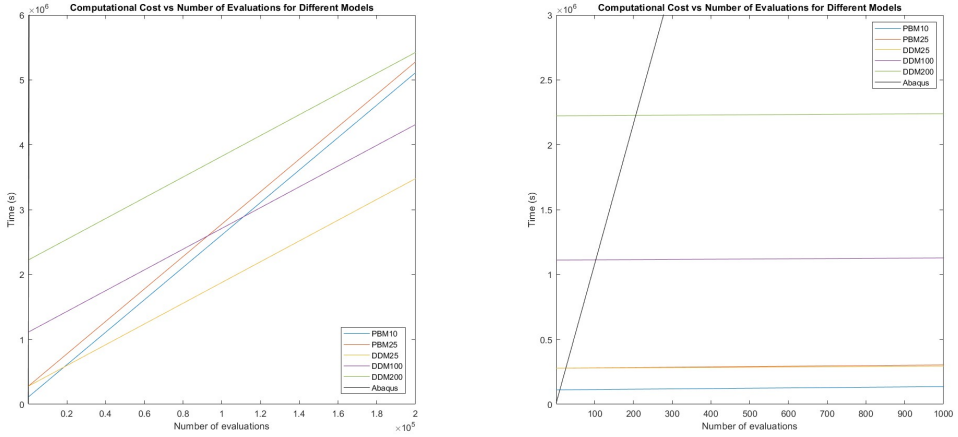
6.4.2 Computational Time

Figure 6.21 compares the computational costs to evaluate the peak responses for different types of models. The computational cost includes the time for training the models followed

Table 6.16. Data-Driven Surrogates vs Physics-Based Surrogates for Prediction of MSDR: Performance

Model Type	Number of Training Points	Validation NRMSE	Validation R^2	Testing NRMSE	Testing R^2
Data-Driven	100	0.53	0.93	0.70	0.91
Physics-Based	10	0.13	0.994	0.18	0.99
Physics-Based	100	0.28	0.98	0.19	0.99

as well as the time for predicting the responses using the trained model. All of these costs are calculated using the same device to allow for direct comparisons between each of these methods. For the Abaqus 2D Nonlinear model (labeled as Abaqus on the graph), there is no training time, and the time for evaluation for each simulation is assumed to be 180 minutes (average of three and four hours). This however, does not include the time required to build and validate the Abaqus model, which is usually much higher than the time required to run the model.



(a) Computational Cost vs Number of Simulations

(b) Computational Cost vs Number of Simulations: Zoomed in

Figure 6.21. Computational Cost vs Number of Simulations for Different Types of Models

For the data-driven surrogate models (labeled DDM on the graph, followed by the number of training points used), the training cost includes the cost to generate high-fidelity data

(which is equal to 180 minutes per training point), the cost to obtain the spectral response properties (which is assumed to be about 16 seconds per training point) and the cost to run the Bayesian optimization and train the models (assumed to be 5 minutes in total). The prediction cost includes the time to obtain the spectral response properties (16 seconds per evaluation) and the time to run the data-driven surrogate to obtain the prediction (which is of the orders of milliseconds and can be neglected).

For the physics-based surrogate models (labeled PBM on the graph, followed by the number of training points used), the training cost includes the cost to generate high-fidelity data (which is equal to 180 minutes per training point), the cost to evaluate the low-fidelity shear building model (assumed to be 25 seconds per training point), and the cost to run the Bayesian optimization and train the models (assumed to be about 16 seconds per training point). The prediction cost includes the time to run the low-fidelity shear building model (25 points per evaluation) and the time to run the mapping between the low-fidelity and the high-fidelity data (which is of the orders of milliseconds and can be neglected).

Table 6.17. Data-Driven Surrogates vs Physics-Based Surrogates for Prediction of MSDR: Performance

Model	**Training Time (s per training point)**	**Prediction Time (s per prediction point)**
Abaqus 2D Nonlinear	-	10800
Data-Driven	11116	16
Physics-Based	11125	25

Figure 6.21b provides a magnified view by limiting the number of evaluations to 1000. From the figure, it can be seen that while the initial computational cost is high for all the surrogate models, the cost required for the Abaqus model goes over the cost for all of the surrogate models when a slightly higher number of evaluations (less than 50 for all except the data-driven surrogate with 200 training points) are required to be conducted. As for the comparison between the data-driven and physics-based models, when the same number of points are used to train each of these types of models, the initial cost is the same for both the data-driven and physics-based models. However, since the cost for each evaluation, after training, is lower for data-driven surrogate models, when the same number of data points

are used to train a physics-based and a data-driven model, the data-driven model has the lower cost as the number of evaluations required increases. When the number of evaluations required is greater than 2000, the data-driven surrogate model trained with 25 training points requires less computational time than the physics-based surrogate model with 10 training points.

Therefore, for the prediction of maximum roof drift ratio, when the number of evaluations required is less than 2000, the physics-based model with 10 points can be used, and when the number of evaluations is over 2000, the data-driven model with 25 training points can be used (since both these model have similar magnitudes of prediction errors, the computational cost is the next parameter to choose which model to use). For the prediction of maximum story drift ratio (if we assume that a data-driven surrogate model trained on 200 data points provides the same level of performance as the physics-based surrogate models trained on either 10 or 25 points), then the physics-based surrogate model (trained on either 10 points or 25 points) requires less computational cost than the data-driven surrogate model for up to over 200000 evaluations. When the number of evaluations required is over 200000 (a little higher than 200000 to be more precise), the data-driven model will be less computationally expensive than the physics-based models for predicting maximum story drift ratio.

6.4.3 Capabilities

In this subsection, the two types of surrogate models are compared in terms of application. Machine learning models are usually applicable to very specific scenarios on which the models are trained. Therefore, the data-driven models developed in this study will apply only to this particular 15-story steel moment-resisting frame building with the structural properties being inside the assumed range of variation. Therefore, the data-driven surrogate models developed for this building should not be used for buildings with different configurations or structural properties that are outside the bound of the variations considered in this study. This is further supported by the fact that the tuned hyperparameters for each of these data-driven surrogate models were somewhat arbitrary. Similarly, the same framework used in

this study to develop data-driven surrogate models for the selected building may not directly apply to other sorts of structures.

Furthermore, if data-driven surrogate models with mean spectral response properties and no structural properties as inputs are used, then the effects of randomness in structural properties are not captured by the data-driven surrogates. Moreover, if the data-driven surrogate is built using a single input, the same response is obtained for all the ground motions that have the same value of this input.

As for the physics-based surrogate models, since the low-fidelity model captures the characteristics of the response of the building, these can be potentially used for structural properties that are outside the range assumed in the study (unless the structural properties are extremely high or low). Since the changes in geometry (with fixed symmetry) can be captured by the changes in the overall stiffness and mass of the low-fidelity model, the same surrogate models can potentially be used considering the randomness in geometric properties as well (although this would have to be validated). Furthermore, the same type of framework used here can be applied to other types of structures with a low-fidelity model capturing the general behavior of the structure to ground motion and a map to modify the low-fidelity results to the high-fidelity responses.

The physics-based surrogate models capture the effects of changes in structural properties and also do not depend on a single ground motion characteristic for the prediction, therefore not necessarily predicting the same values for the response for different ground motions with certain similar characteristics.

7. FRAGILITY ANALYSIS

In this chapter, the surrogate models developed in Chapter 6 are used for fragility analysis of the selected structure.

7.1 Ground Motion Selection

It is assumed that the structure is located in Big Bear Lake City in California. This location is selected because past ground motion records are easily available in databases in this region owing to the California Strong Motion Instrumentation Program. 90 ground motion records from regions near Big Bear Lake and other regions in California are selected for the fragility study.

7.2 Engineering Demand Parameter

Staying consistent with past research, the maximum story drift ratio is selected as the engineering demand parameter for this study. Damage states definitions provided by FEMA-356[44] are used in the study. Table 7.1 provides the damage state definitions from FEMA-356 for steel moment frame buildings. While FEMA-356 recommends not to use damage states based on limiting drift values for absolute acceptance criteria for assessing damage in a building post-earthquake, it does allow us to use the limiting drift values as a guide for the probabilistic assessment of damage states of a building as is done here.

Table 7.1. Damage State Definitions from FEMA-356

Building Performance Levels	Damage	Transient Drift
Operational Overall	Very Little	-
Immediate Occupancy (IO)	Light	0.7%
Life Safety (LS)	Moderate	2.5%
Collapse Prevention (CP)	Severe	5%

7.3 Intensity Measure

In Chapter 6 it is found that spectral velocity is the ground motion intensity parameter that correlates best with the maximum story drift ratio. Therefore, the 1.1% damped spectral velocity at the fundamental frequency of the building is used as the ground motion intensity measure for the fragility analysis.

7.4 Analysis

Incremental dynamic analysis is used for obtaining the fragility points. Each ground motion is scaled to different values of spectral velocities from 200 mm/s to 3000 mm/s at 200 mm/s intervals (these values are spectral velocities for the unscaled structure, these are scaled down by a factor of 5.48 before applying to the scaled building). Monte Carlo simulation with 5040 points is used to sample the structural properties. This results in 56 different structural property values for each ground motion selected.

The physics-based surrogate models developed in Chapter 6 are used to obtain the maximum story drift ratios for each combination of ground motions and structural points. For this, first, the low-fidelity results are obtained using the shear-building model with mass matrix augmentation and then the trained mapping is used to obtain the predictions for maximum story drifts.

Once the story drifts are obtained, these are compared to the limit values to obtain the damage states. Using the damage states, the probability of the building exhibiting a particular damage state is calculated as the ratio of the number of data points showing a particular damage state to the total number of data points for each value of spectral acceleration.

Maximum likelihood analysis is used to estimate the parameters of the best lognormal curve to fit the fragility points. 95% confidence intervals for the estimated parameters are found using the Wald confidence interval approach and plotted on the fragility curve. While the Wald method is less accurate than the likelihood ratio method to determine the confidence intervals, the confidence intervals can be obtained analytically and do not require numerical solutions to be found like when using the likelihood ratio. The biggest

limitation of the Wald method is that it assumes the normal approximation of the binomial distribution, thus resulting in close to no variation at the ends. The convergence of the estimated parameters is studied by varying the number of ground motions used for the Monte Carlo Sampling.

The same process is carried out without varying the structural properties to see the effects of randomness in structural properties as compared to randomness in ground motion properties. For this, the response is obtained for the mean structural properties for each of the 90 ground motion records, and fragility curves are obtained just using these responses.

7.5 Results and Discussions

Figure 7.1 shows the probability of damage states for each damage state at the selected spectral velocity values as well as the lognormal curve fitted through the points.

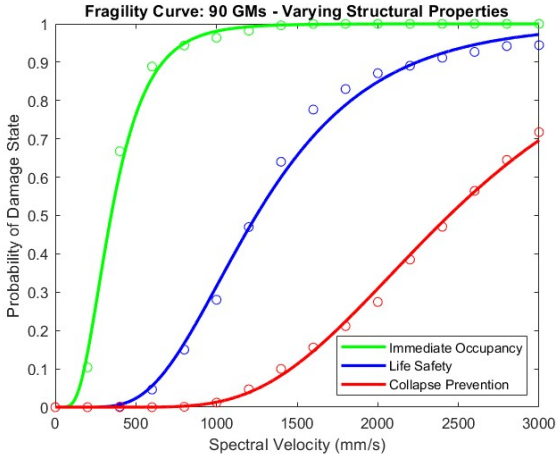


Figure 7.1. Fragility Curves with Determined Fragility Points at Equally Spaced Intervals

Figure 7.2 shows the fragility curve along with the 95% confidence bounds accounting for uncertainties. It can be seen from this graph that the band covering the 95% confidence interval is a narrow one, suggesting that the choice of spectral velocity as the intensity measure to condition the structural response is a good one. The only drawback of selecting spectral velocity as the ground motion intensity measure is the lack of adequate hazard maps

listing the hazards for different regions in spectral velocity values as opposed to spectral acceleration or peak ground acceleration values.

The confidence band is wider at probabilities in the mid-range as compared to the ends where the probability of damage states being exhibited is 0 or 100%. This is partly because, at low ground motion intensities, no ground motion record produces high structural responses, and at high ground motion intensities, all ground motion records produce high structural responses. A part of the reason for such narrow confidence intervals near the ends is due to the limitation of the Wald method used to obtain the confidence interval. However, at medium values of ground motion intensities, some of the scaled ground motions drive the structure into a certain damage state, while some others do not, causing uncertainty based on the ground motion set selected.

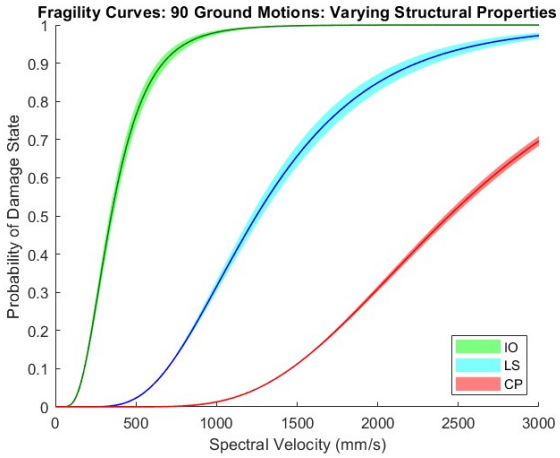


Figure 7.2. Fragility Curves with Estimated Confidence Intervals

Figure 7.3 shows the convergence of the mean of the lognormal distribution estimated through maximum likelihood estimate as the number of ground motions used in the Monte Carlo simulation is increased. The mean of the lognormal distribution for a damage state is the measure of the log of spectral velocity at which the probability of the structural response exceeding that particular damage state is 0.5. Table 7.2 shows the values of spectral velocity (rounded to the nearest multiple of 50 mm/s) at which the mean probability of exceeding each damage state is 0.5 when using 90 ground motions in the Monte Carlo simulation.

Finally, it can be seen that the value of the mean of the lognormal distribution reaches a constant value within 65 ground motions.

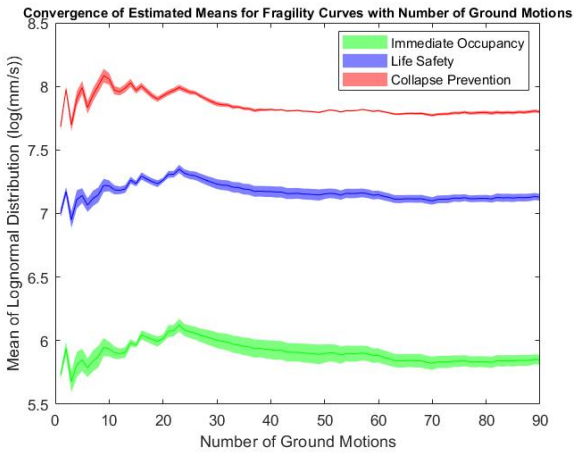


Figure 7.3. Convergence of Means of Fragility Curve with Number of Ground Motions

Table 7.2. Values of Spectral Velocity for Mean Probability of Exceeding Damage State of 0.5 Using 90 Ground Motions in the MCS

Building Performance Levels	Spectral Velocity (mm/s)
Immediate Occupancy (IO)	350
Life Safety (LS)	1250
Collapse Prevention (CP)	2450

Similarly, Figure 7.4 shows the convergence of the standard deviation estimate of the fragility curves obtained using maximum likelihood analysis. The standard deviation of the lognormal distribution is a measure of the steepness of the fragility curve. A lower standard deviation means a steeper curve. Similar to the mean of the lognormal distribution, the standard deviation reaches a near-constant value of around 65 ground motions. For both the mean, as well as standard deviation estimates, it is seen that the uncertainty is higher for Immediate Occupancy and Life Safety limit states than it is for Collapse Prevention limit states in the range of spectral velocity selected.

Figure 7.5 shows the fragility curves obtained when the randomness in structural properties is ignored and the fragility is calculated at the mean values of structural properties.

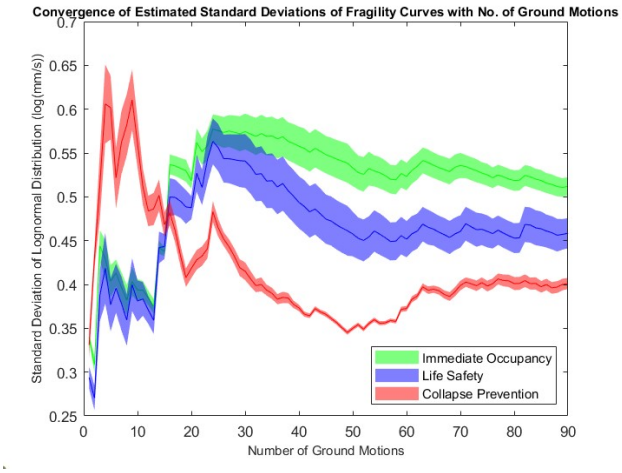


Figure 7.4. Convergence of Standard Deviations of Fragility Curve with Number of Ground Motions

It can be seen that there is almost no change in the fragility or the bounds of estimated parameters on removing the variation in structural properties. This confirms that the effect of ground motion variation is much more pronounced than the effect of randomness in structural parameters on the fragility of buildings.

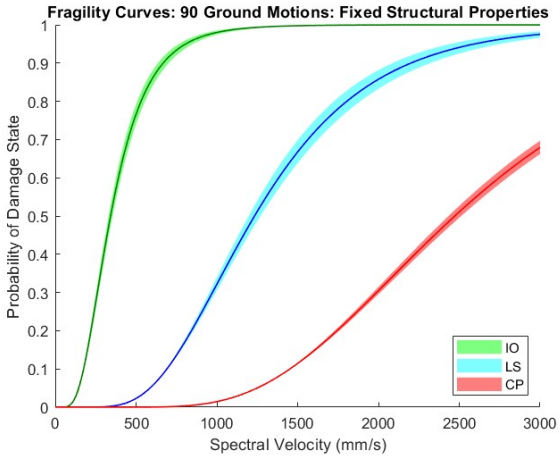


Figure 7.5. Fragility Curves for Building with Fixed Structural Properties

Figures 7.6 and 7.7 show the convergence of the estimated mean and standard deviation values for the study with fixed structural properties. It can be seen from these graphs that

here also the convergence is obtained when around 65 ground motions are used in the Monte Carlo simulation. The estimates for the mean obtained using constant structural properties match closely the estimates obtained when considering the randomness in structural properties. This reaffirms that the effects due to variation in ground motions trump the effects due to variation in structural properties.

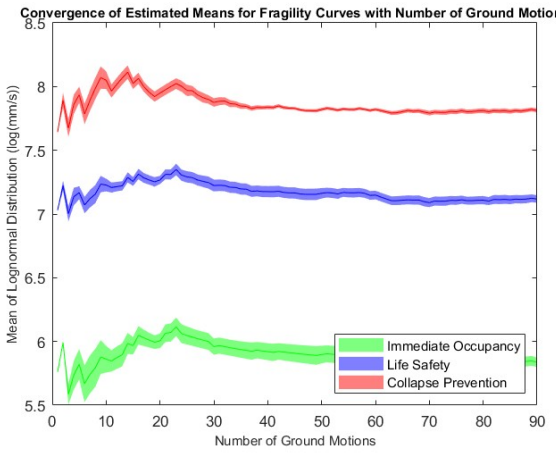


Figure 7.6. Convergence of Means of Fragility Curve with Number of Ground Motions for Constant Structural Properties

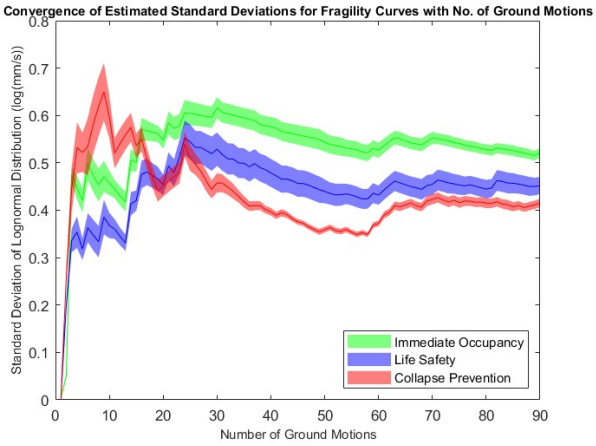


Figure 7.7. Convergence of Standard Deviations of Fragility Curve with Number of Ground Motions for Constant Structural Properties

7.6 Findings

The findings from the study are listed in this section:

- Spectral velocity is a good choice for the ground motion intensity measure when using maximum story drift ratio as the engineering demand parameter.
- The effects due to randomness in ground motion characteristics on the fragility curve are significantly higher than corresponding effects due to randomness in structural properties.
- For the building and location selected (15-story steel moment resisting frame building in California), around 65 ground motions are required for the convergence of parameters for the lognormal distribution of fragility.

8. CONCLUSIONS AND FUTURE WORK

This chapter lists the conclusions drawn from this study and lists a set of possible future work for the study conducted in this thesis.

8.1 Conclusions

In this study, the process of developing data-driven and physics-based surrogate models for the prediction of responses of buildings to ground motion excitations is studied. The two types of surrogate models are compared and the limitations and capabilities of both of these types of surrogate models are learned.

When using data-driven surrogate models for the selected building, it is found to be easier to predict the maximum roof drift ratio than the maximum story drift ratio. Just 25 training points are seen to be sufficient for the prediction of the maximum roof drift ratio using data-driven surrogate models. The number of training data points required to predict the maximum roof drift ratio when varying the structural properties does not increase much compared to when structural properties are fixed. On the other hand, when predicting the maximum story drift ratio using data-driven surrogate models, the target performance is not obtained with 100 data points. Furthermore, there is a significant increase in the number of data points required in testing when accounting for variations in structural properties as compared to when keeping the structural properties fixed. One of the reasons for the higher difficulty when predicting the maximum story drift ratio is that, unlike the maximum roof drift ratio, the maximum story drift ratio may occur in any of the 15 stories of the building. This makes it difficult for a data-driven technique to learn the pattern with a limited number of training points.

Meanwhile, while using physics-based surrogate models for the selected building, the relative ease of predicting the maximum roof drift ratio and the maximum story drift ratio is the same. It is also found that physics-based models require a lower amount of training data points to achieve similar levels of accuracy as data-driven surrogate models with a higher number of training data points. Therefore, while the time required for prediction is lower for data-driven surrogate models, the initial time to train a model with low magnitudes

of errors is significantly lesser for physics-based surrogate models. One of the reasons for both these observations is that the low-fidelity model captures the physical characteristics of the building's response. Therefore, the map used to match the low-fidelity data to the high-fidelity data just has to learn the differences in the two types of data at various levels of responses.

Furthermore, the framework used for developing physics-based surrogate models herein can be applied to other structures as well as long as a low-fidelity model can be built, capturing the variation in mass and stiffness. The low-fidelity models could also have the capability of capturing the effects of changes in the geometry and topology of buildings (although this will need to be validated).

Finally, in this study, it is found that among the ground motion characteristics selected, the spectral displacement has the best correlation with the maximum roof drift ratio, and the spectral velocity has the best correlation with the maximum story drift ratio. Fragility curves developed using these parameters as intensity measures while using the maximum roof drift ratio and the maximum story drift ratio as engineering demand parameters respectively will contain less variability.

The surrogate models developed in this study can be used for fragility analysis, parametric studies, and for exploring the design space during the conceptual design stage. In this study, the surrogate models are used to conduct fragility analysis for the selected building using a maximum likelihood estimate to obtain the parameters along with confidence bounds for the fragility curve. The findings listed previously are used when developing the fragility curves. The maximum story drift ratio is selected as the engineering demand parameter, and the spectral velocity is used as the ground motion intensity measure.

8.2 Future Work

This section lists a set of potential future work for the study conducted in this thesis.

- Variation in building geometry: In this study, it was assumed that the building geometry remains the same and only the material properties vary. A study can be conducted to vary the building geometry and topology as well and study how different surrogate

modeling techniques will have to be adapted to incorporate this change. This could potentially facilitate the use of surrogate models to make early design decisions.

- Irregularity in the building: In this study, the building was assumed to be a symmetric moment-resisting frame. A study can be conducted modeling the irregularities in the geometric or the material properties in different parts of the building to study how the surrogate modeling techniques will have to be adapted to incorporate this change.

- Include damage and imperfections in joints in the high-fidelity model, increasing the complexity of the problem and studying how the surrogate modeling techniques have to be adapted based on this change.

- Compare the fragility curves developed in this study with the fragility curves developed using the capacity spectrum method. A comparison regarding the reliability of the fragility curves obtained using the two methods and the computational time required to obtain these curves can help determine which method would be more useful based on the application.

- Develop knowledge-based surrogate models using observed data for similar structures from previous earthquakes or by using expert judgement. However, the biggest challenge here is that this data is difficult to come by.

REFERENCES

[1] ATC, *Guidelines for Performance-based Seismic Design of Buildings*. FEMA, 2018.

[2] H. Sun, H. V. Burton, and H. Huang, "Machine learning applications for building structural design and performance assessment: State-of-the-art review," *Journal of Building Engineering*, vol. 33, p. 101 816, 2021.

[3] S. Vaseghiamiri, M. Mahsuli, M. A. Ghannad, and F. Zareian, "Surrogate sdof models for probabilistic performance assessment of multistory buildings: Methodology and application for steel special moment frames," *Engineering Structures*, vol. 212, p. 110 276, 2020.

[4] M. Shinozuka, M. Q. Feng, J. Lee, and T. Naganuma, "Statistical analysis of fragility curves," *Journal of engineering mechanics*, vol. 126, no. 12, pp. 1224–1231, 2000.

[5] M. Z. Esteghamati and M. M. Flint, "Do all roads lead to rome? a comparison of knowledge-based, data-driven, and physics-based surrogate models for performance-based early design," *Engineering Structures*, vol. 286, p. 116 098, 2023.

[6] C. A. Kircher, R. V. Whitman, and W. T. Holmes, "Hazus earthquake loss estimation methods," *Natural Hazards Review*, vol. 7, no. 2, pp. 45–59, 2006.

[7] K. Bakalis and D. Vamvatsikos, "Seismic fragility functions via nonlinear response history analysis," *Journal of structural engineering*, vol. 144, no. 10, p. 04 018 181, 2018.

[8] A. Muntasir Billah and M. Shahria Alam, "Seismic fragility assessment of highway bridges: A state-of-the-art review," *Structure and infrastructure engineering*, vol. 11, no. 6, pp. 804–832, 2015.

[9] O.-S. Kwon and A. Elnashai, "The effect of material and ground motion uncertainty on the seismic vulnerability curves of rc structure," *Engineering structures*, vol. 28, no. 2, pp. 289–303, 2006.

[10] M. A. Hariri-Ardebili and V. E. Saouma, "Seismic fragility analysis of concrete dams: A state-of-the-art review," *Engineering structures*, vol. 128, pp. 374–399, 2016.

[11] P. B. Tekie and B. R. Ellingwood, "Seismic fragility assessment of concrete gravity dams," *Earthquake engineering & structural dynamics*, vol. 32, no. 14, pp. 2221–2240, 2003.

[12] R. Pang, B. Xu, X. Kong, and D. Zou, "Seismic fragility for high cfrds based on deformation and damage index through incremental dynamic analysis," *Soil Dynamics and Earthquake Engineering*, vol. 104, pp. 432–436, 2018.

[13] T. K. Mandal, S. Ghosh, and N. N. Pujari, "Seismic fragility analysis of a typical Indian PHWR containment: Comparison of fragility models," *Structural Safety*, vol. 58, pp. 11–19, 2016.

[14] S. Argyroudis and K. Pitilakis, "Seismic fragility curves of shallow tunnels in alluvial deposits," *Soil Dynamics and Earthquake Engineering*, vol. 35, pp. 1–12, 2012.

[15] G. Tsinidis, F. de Silva, I. Anastasopoulos, *et al.*, "Seismic behaviour of tunnels: From experiments to analysis," *Tunnelling and underground space technology*, vol. 99, p. 103 334, 2020.

[16] N. D. Lagaros, Y. Tsompanakis, P. N. Psarropoulos, and E. C. Georgopoulos, "Computationally efficient seismic fragility analysis of geostructures," *Computers & Structures*, vol. 87, no. 19-20, pp. 1195–1203, 2009.

[17] D. H. Kim, S. G. Lee, and I. K. Lee, "Seismic fragility analysis of 5 mw offshore wind turbine," *Renewable energy*, vol. 65, pp. 250–256, 2014.

[18] S. Soroushian, A. E. Zaghi, M. Maragakis, A. Echevarria, Y. Tian, and A. Filiatrault, "Analytical seismic fragility analyses of fire sprinkler piping systems with threaded joints," *Earthquake Spectra*, vol. 31, no. 2, pp. 1125–1155, 2015.

[19] Y. Pang, X. Wu, G. Shen, and W. Yuan, "Seismic fragility analysis of cable-stayed bridges considering different sources of uncertainties," *Journal of Bridge Engineering*, vol. 19, no. 4, p. 04 013 015, 2014.

[20] M. Polese, G. M. Verderame, C. Mariniello, I. Iervolino, and G. Manfredi, "Vulnerability analysis for gravity load designed rc buildings in naples–italy," *Journal of Earthquake Engineering*, vol. 12, no. S2, pp. 234–245, 2008.

[21] T. Rossetto and A. Elnashai, "Derivation of vulnerability functions for european-type rc structures based on observational data," *Engineering structures*, vol. 25, no. 10, pp. 1241–1263, 2003.

[22] S. Glaister and R. Pinho, "Development of a simplified deformation-based method for seismic vulnerability assessment," *Journal of Earthquake Engineering*, vol. 7, no. spec01, pp. 107–140, 2003.

[23] J. Seo, L. Dueñas-Osorio, J. I. Craig, and B. J. Goodno, "Metamodel-based regional vulnerability estimate of irregular steel moment-frame structures subjected to earthquake events," *Engineering Structures*, vol. 45, pp. 585–597, 2012.

[24] Y. Pan, A. K. Agrawal, and M. Ghosn, "Seismic fragility of continuous steel highway bridges in new york state," *Journal of Bridge Engineering*, vol. 12, no. 6, pp. 689–699, 2007.

[25] I. Iervolino, G. Fabbrocino, and G. Manfredi, "Fragility of standard industrial structures by a response surface based method," *Journal of Earthquake Engineering*, vol. 8, no. 06, pp. 927–945, 2004.

[26] B. G. Nielson and R. DesRoches, "Analytical seismic fragility curves for typical bridges in the central and southeastern united states," *Earthquake spectra*, vol. 23, no. 3, pp. 615–633, 2007.

[27] E. Choi, R. DesRoches, and B. Nielson, "Seismic fragility of typical bridges in moderate seismic zones," *Engineering structures*, vol. 26, no. 2, pp. 187–199, 2004.

[28] T. Rossetto and A. Elnashai, "A new analytical procedure for the derivation of displacement-based vulnerability curves for populations of rc structures," *Engineering structures*, vol. 27, no. 3, pp. 397–409, 2005.

[29] G. Lupoi, P. Franchin, A. Lupoi, and P. E. Pinto, "Seismic fragility analysis of structural systems," *Journal of Engineering Mechanics*, vol. 132, no. 4, pp. 385–395, 2006.

[30] A. M. Reinhorn, R. Barron-Corverra, and A. G. Ayala, "Global spectral evaluation of seismic fragility of structures," in *Proc., 7th US national conference on earthquake engineering*, 2002.

[31] K. M. Mosalam, G. Ayala, R. N. White, and C. Roth, "Seismic fragility of lrc frames with and without masonry infill walls," *Journal of Earthquake Engineering*, vol. 1, no. 04, pp. 693–720, 1997.

[32] C. Haselton, A. Whittaker, A. Hortacsu, J. Baker, J. Bray, and D. Grant, "Selecting and scaling earthquake ground motions for performing response-history analyses," in *Proceedings of the 15th world conference on earthquake engineering*, Earthquake Engineering Research Institute Oakland, CA, USA, 2012, pp. 4207–4217.

[33] C. A. Cornell, "On earthquake record selection for nonlinear dynamic analysis," in *The Esteva Symposium*, 2005.

[34] Y. Pan, A. K. Agrawal, M. Ghosn, and S. Alampalli, "Seismic fragility of multispan simply supported steel highway bridges in new york state. ii: Fragility analysis, fragility curves, and fragility surfaces," *Journal of Bridge Engineering*, vol. 15, no. 5, pp. 462–472, 2010.

[35] S. Lagomarsino and S. Giovinazzi, "Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings," *Bulletin of Earthquake Engineering*, vol. 4, pp. 415–443, 2006.

[36] D. Seyedi, P. Gehl, J. Douglas, L. Davenne, N. Mezher, and S. Ghavamian, "Development of seismic fragility surfaces for reinforced concrete buildings by means of nonlinear time-history analysis," *Earthquake Engineering & Structural Dynamics*, vol. 39, no. 1, pp. 91–108, 2010.

[37] F. Peña, I. Bilionis, S. J. Dyke, Y. Cao, and G. P. Mavroeidis, "Efficient seismic fragility functions through sequential selection," *Structural Safety*, vol. 87, p. 101 977, 2020.

[38] J. W. Baker and C. Allin Cornell, "A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon," *Earthquake Engineering & Structural Dynamics*, vol. 34, no. 10, pp. 1193–1217, 2005.

[39] K. R. Mackie and B. Stojadinovi, *Fragility basis for California highway overpass bridge seismic decision making*. Pacific Earthquake Engineering Research Center, College of Engineering , 2005.

[40] J. Kiani and S. Pezeshk, "Sensitivity analysis of the seismic demands of rc moment resisting frames to different aspects of ground motions," *Earthquake engineering & structural dynamics*, vol. 46, no. 15, pp. 2739–2755, 2017.

[41] G. W. Housner, "Intensity of ground motion during strong earthquakes," 1952.

[42] F. Khosravikia, M. Mahsuli, and M. A. Ghannad, "The effect of soil–structure interaction on the seismic risk to buildings," *Bulletin of Earthquake Engineering*, vol. 16, pp. 3653–3673, 2018.

[43] B. G. Nielson and R. DesRoches, "Seismic fragility methodology for highway bridges using a component level approach," *Earthquake engineering & structural dynamics*, vol. 36, no. 6, pp. 823–839, 2007.

[44] F. E. FEMA 356 *et al.*, "Prestandard and commentary for the seismic rehabilitation of buildings," *Federal Emergency Management Agency: Washington, DC, USA*, 2000.

[45] C. Seminar, "Nehrp guidelines for the seismic rehabilitation of buildings (fema 273)," *Building Seismic Safety Council: Washington, DC, USA*, 1997.

[46] E. Miranda and C. J. Reyes, "Approximate lateral drift demands in multistory buildings with nonuniform stiffness," *Journal of Structural Engineering*, vol. 128, no. 7, pp. 840–849, 2002.

[47] Y.-Y. Lin and E. Miranda, "Estimation of maximum roof displacement demands in regular multistory buildings," *Journal of engineering mechanics*, vol. 136, no. 1, pp. 1–11, 2010.

[48] S. A. Freeman, "The capacity spectrum method as a tool for seismic design," in *Proceedings of the 11th European conference on earthquake engineering*, Citeseer, 1998, pp. 6–11.

[49] A. K. Chopra and R. K. Goel, "A modal pushover analysis procedure for estimating seismic demands for buildings," *Earthquake engineering & structural dynamics*, vol. 31, no. 3, pp. 561–582, 2002.

[50] J. Park, P. Towashiraporn, J. I. Craig, and B. J. Goodno, "Seismic fragility analysis of low-rise unreinforced masonry structures," *Engineering Structures*, vol. 31, no. 1, pp. 125–137, 2009.

[51] D. Straub and A. Der Kiureghian, "Improved seismic fragility modeling from empirical data," *Structural safety*, vol. 30, no. 4, pp. 320–336, 2008.

[52] M. Mahsuli and T. Haukaas, "Seismic risk analysis with reliability methods, part i: Models," *Structural Safety*, vol. 42, pp. 54–62, 2013.

[53] M. Mahsuli and T. Haukaas, "Seismic risk analysis with reliability methods, part ii: Analysis," *Structural Safety*, vol. 42, pp. 63–74, 2013.

[54] X. Guan, H. Burton, M. Shokrabadi, and Z. Yi, "Seismic drift demand estimation for steel moment frame buildings: From mechanics-based to data-driven models," *journal of structural engineering*, vol. 147, no. 6, p. 04 021 058, 2021.

[55] J. Kiani, C. Camp, and S. Pezeshk, "On the application of machine learning techniques to derive seismic fragility curves," *Computers & Structures*, vol. 218, pp. 108–122, 2019.

[56] Z. Wang, N. Pedroni, I. Zentner, and E. Zio, "Seismic fragility analysis with artificial neural networks: Application to nuclear power plant equipment," *Engineering Structures*, vol. 162, pp. 213–225, 2018.

[57] C. C. Mitropoulou and M. Papadrakakis, "Developing fragility curves based on neural network ida predictions," *Engineering Structures*, vol. 33, no. 12, pp. 3409–3421, 2011.

[58] I. Gidaris, A. A. Taflanidis, and G. P. Mavroeidis, "Kriging metamodeling in seismic risk assessment based on stochastic ground motion models," *Earthquake Engineering & Structural Dynamics*, vol. 44, no. 14, pp. 2377–2399, 2015.

[59] L. Micheli, A. Alipour, and S. Laflamme, "Multiple-surrogate models for probabilistic performance assessment of wind-excited tall buildings under uncertainties," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, vol. 6, no. 4, p. 04 020 042, 2020.

[60] M. Z. Esteghamati and M. M. Flint, "Developing data-driven surrogate models for holistic performance-based assessment of mid-rise rc frame buildings at early design," *Engineering Structures*, vol. 245, p. 112 971, 2021.

[61] J. Ghosh, J. E. Padgett, and L. Dueñas-Osorio, "Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges," *Probabilistic Engineering Mechanics*, vol. 34, pp. 189–199, 2013.

[62] C.-S. Huang, S.-L. Hung, C. Wen, and T. Tu, "A neural network approach for structural identification and diagnosis of a building from seismic response data," *Earthquake engineering & structural dynamics*, vol. 32, no. 2, pp. 187–206, 2003.

[63] B. K. Oh, Y. Park, and H. S. Park, "Seismic response prediction method for building structures using convolutional neural network," *Structural Control and Health Monitoring*, vol. 27, no. 5, e2519, 2020.

[64] R. Zhang, Y. Liu, and H. Sun, "Physics-guided convolutional neural network (phy-cnn) for data-driven seismic response modeling," *Engineering Structures*, vol. 215, p. 110 704, 2020.

[65] M. Zaker Esteghamati, J. Lee, M. Musetich, and M. M. Flint, "Inssept: An open-source relational database of seismic performance estimation to aid with early design of buildings," *Earthquake Spectra*, vol. 36, no. 4, pp. 2177–2197, 2020.

[66] N. Buratti, P. J. Stafford, and J. J. Bommer, "Earthquake accelerogram selection and scaling procedures for estimating the distribution of drift response," *Journal of Structural Engineering*, vol. 137, no. 3, pp. 345–357, 2011.

[67] K. Lang and H. Bachmann, "On the seismic vulnerability of existing buildings: A case study of the city of basel," *Earthquake Spectra*, vol. 20, no. 1, pp. 43–66, 2004.

[68] C. A. Kircher, A. A. Nassar, O. Kustu, and W. T. Holmes, "Development of building damage functions for earthquake loss estimation," *Earthquake Spectra*, vol. 13, no. 4, pp. 663–682, 1997.

[69] S. H. R. Tabatabaiefar, B. Fatahi, and B. Samali, "Numerical and experimental investigations on seismic response of building frames under influence of soil-structure interaction," *Advances in structural Engineering*, vol. 17, no. 1, pp. 109–130, 2014.

[70] M. P. González and J. L. Zapico, "Seismic damage identification in buildings using neural networks and modal data," *Computers & structures*, vol. 86, no. 3-5, pp. 416–426, 2008.

[71] B. Chiou, R. Darragh, N. Gregor, and W. Silva, "Nga project strong-motion database," *Earthquake Spectra*, vol. 24, no. 1, pp. 23–44, 2008.

[72] T. D. Ancheta, R. B. Darragh, J. P. Stewart, *et al.*, "Nga-west2 database," *Earthquake Spectra*, vol. 30, no. 3, pp. 989–1005, 2014.

[73] C. A. Goulet, T. Kishida, T. D. Ancheta, *et al.*, "Peer nga-east database," *Earthquake Spectra*, vol. 37, no. 1_suppl, pp. 1331–1353, 2021.

[74] H. Haddadi, A. Shakal, C. Stephens, *et al.*, "Center for engineering strong-motion data (cesmd)," in *Proceedings of the 14th World Conference on Earthquake Engineering, Beijing, October*, 2008, pp. 12–17.

[75] T. Fushiki, "Estimation of prediction error by using k-fold cross-validation," *Statistics and Computing*, vol. 21, pp. 137–146, 2011.

[76] Dassault Systèmes. "Abaqus analysis user's guide - version 6.14." (2014), [Online]. Available: http://130.149.89.49:2080/v6.14/books/usb/default.htm.

[77] Z.-F. Fu and J. He, *Modal analysis*. Elsevier, 2001.

[78] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in neural information processing systems*, vol. 25, 2012.

[79] N. A. F. Senan, "A brief introduction to using ode45 in matlab," *University of California at Berkeley, USA*, 2007.

[80] M. Stein, "Large sample properties of simulations using latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.

[81] T. M. Inc., *Matlab version: 9.14.0 (r2023a)*, Natick, Massachusetts, United States, 2023. [Online]. Available: https://www.mathworks.com.

[82] S. Koziel and A. Pietrenko-Dabrowska, *Performance-driven surrogate modeling of high-frequency structures*. Springer, 2020, vol. 765.

A. APPENDIX 1: RESULTS FOR DATA-DRIVEN MODELS FOR THE PREDICTION OF PEAK RESPONSES IN PRELIMINARY STUDY 1

Table A.1. Data-Driven Model Results: MRDR - Inputs: PGA - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	2.74	0.19	1.75	0.49
GPR	2.83	0.14	2.10	0.27
ANN	2.62	0.26	1.98	0.35
Linear Regression	2.90	0.10	2.17	0.22

Table A.2. Data-Driven Model Results: MRDR - Inputs: PGA - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	1.94	0.35	3.00	-0.49
GPR	2.03	0.29	3.46	-0.99
ANN	1.99	3.94	1.83	0.44
Linear Regression	2.14	0.20	2.06	0.30

Table A.3. Data-Driven Model Results: MRDR - Inputs: PGA - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	1.94	0.29	1.93	0.39
GPR	1.84	0.36	2.72	-0.22
ANN	1.87	0.34	1.83	0.44
Linear Regression	1.94	0.30	2.08	0.28

Table A.4. Data-Driven Model Results: MRDR - Inputs: PGA - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	1.74	0.31	1.97	0.36
GPR	1.75	0.30	2.41	0.04
ANN	1.72	0.32	1.85	0.43
Linear Regression	1.79	3.19	2.08	0.28

Table A.5. Data-Driven Model Results: MRDR - Inputs: S_a - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.19	0.996	0.35	0.98
GPR	0.25	0.993	0.35	0.98
ANN	0.23	0.995	0.34	0.98
Linear Regression	0.31	0.99	0.38	0.98

Table A.6. Data-Driven Model Results: MRDR - Inputs: S_v - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.61	0.96	1.58	0.59
GPR	0.72	0.95	1.09	0.80
ANN	0.86	0.92	0.43	0.97
Linear Regression	0.81	0.93	0.47	0.96

Table A.7. Data-Driven Model Results: MRDR - Inputs: S_v - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.42	0.97	2.02	0.33
GPR	0.55	0.94	0.47	0.96
ANN	0.53	0.95	0.64	0.93
Linear Regression	0.53	0.95	0.51	0.96

Table A.8. Data-Driven Model Results: MRDR - Inputs: S_v - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.55	0.94	1.52	0.62
GPR	0.50	0.95	1.32	0.71
ANN	0.51	0.95	0.80	0.89
Linear Regression	0.60	0.93	0.47	0.96

Table A.9. Data-Driven Model Results: MRDR - Inputs: S_v - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.55	0.93	0.51	0.96
GPR	0.49	0.94	0.47	0.96
ANN	0.52	0.94	0.53	0.95
Linear Regression	0.53	0.94	0.47	0.96

Table A.10. Data-Driven Model Results: MRDR - Inputs: S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.18	0.997	0.34	0.98
GPR	0.21	0.996	0.43	0.97
ANN	0.22	0.995	0.42	0.97
Linear Regression	0.26	0.99	0.38	0.98

Table A.11. Data-Driven Model Results: MRDR - Inputs: PGA, S_a - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.11	0.999	0.36	0.98
GPR	0.16	0.997	0.14	0.997
ANN	0.12	0.998	0.15	0.996
Linear Regression	0.32	0.99	0.24	0.99

Table A.12. Data-Driven Model Results: MRDR - Inputs: PGA, S_v - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.30	0.99	0.68	0.92
GPR	0.40	0.98	0.85	0.88
ANN	0.34	0.99	0.73	0.91
Linear Regression	0.63	0.96	0.46	0.97

Table A.13. Data-Driven Model Results: MRDR - Inputs: PGA, S_v - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.59	0.94	0.39	0.98
GPR	0.41	0.97	0.72	0.91
ANN	0.47	0.96	0.68	0.92
Linear Regression	0.54	0.95	0.40	0.97

Table A.14. Data-Driven Model Results: MRDR - Inputs: PGA, S_v - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.48	0.96	0.38	0.98
GPR	0.47	0.96	0.42	0.97
ANN	0.45	0.96	0.42	0.97
Linear Regression	0.47	0.96	0.42	0.97

Table A.15. Data-Driven Model Results: MRDR - Inputs: PGA, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.13	0.998	0.22	0.99
GPR	0.39	0.98	0.28	0.99
ANN	0.29	0.99	0.29	0.99
Linear Regression	0.36	0.99	0.24	0.99

Table A.16. Data-Driven Model Results: MRDR - Inputs: S_a, S_v - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.12	0.999	0.33	0.98
GPR	0.13	0.998	0.28	0.99
ANN	0.12	0.999	0.25	0.99
Linear Regression	0.13	0.998	0.27	0.99

Table A.17. Data-Driven Model Results: MRDR - Inputs: S_a, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.25	0.99	0.33	0.98
GPR	0.28	0.99	0.41	0.97
ANN	0.21	0.996	0.37	0.98
Linear Regression	0.24	0.994	0.32	0.98

Table A.18. Data-Driven Model Results: MRDR - Inputs: S_v, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.16	0.997	0.27	0.99
GPR	0.14	0.998	0.20	0.99
ANN	0.13	0.998	0.26	0.99
Linear Regression	0.15	0.998	0.28	0.99

Table A.19. Data-Driven Model Results: MRDR - Inputs: PGA, S_a, S_v - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.31	0.99	0.29	0.99
GPR	0.46	0.98	0.29	0.99
ANN	0.21	0.996	0.32	0.98
Linear Regression	0.46	0.98	0.29	0.99

Table A.20. Data-Driven Model Results: MRDR - Inputs: PGA, S_a, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.13	0.998	0.42	0.97
GPR	0.34	0.99	0.25	0.99
ANN	0.28	0.99	0.27	0.99
Linear Regression	0.22	0.996	0.19	0.99

Table A.21. Data-Driven Model Results: MRDR - Inputs: PGA, S_v, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.27	0.99	0.28	0.99
GPR	0.17	0.997	0.28	0.99
ANN	0.14	0.998	0.28	0.99
Linear Regression	0.20	0.996	0.29	0.99

Table A.22. Data-Driven Model Results: MRDR - Inputs: S_a, S_v, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.15	0.998	0.27	0.99
GPR	0.31	0.99	0.34	0.98
ANN	0.11	0.999	0.25	0.99
Linear Regression	0.22	0.995	0.29	0.99

Table A.23. Data-Driven Model Results: MRDR - Inputs: PGA, S_a, S_v, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.13	0.998	0.29	0.99
GPR	0.13	0.998	0.34	0.98
ANN	0.12	0.998	0.32	0.98
Linear Regression	0.26	0.99	0.40	0.97

Table A.24. Data-Driven Model Results: MSDR - Inputs: PGA - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	1.66	0.55	1.25	0.70
GPR	1.72	0.51	2.44	-0.15
ANN	1.61	0.57	1.15	0.74
Linear Regression	1.83	0.45	1.31	0.67

Table A.25. Data-Driven Model Results: MSDR - Inputs: PGA - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.64	0.58	1.13	0.75
GPR	1.25	0.61	1.83	0.36
ANN	1.21	0.63	1.02	0.80
Linear Regression	1.38	0.52	1.29	0.68

Table A.26. Data-Driven Model Results: MSDR - Inputs: PGA - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	1.27	0.57	1.10	0.77
GPR	1.30	0.55	1.10	0.78
ANN	1.27	0.57	1.02	0.80
Linear Regression	1.35	0.51	1.28	0.68

Table A.27. Data-Driven Model Results: MSDR - Inputs: PGA - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	1.18	0.56	1.11	0.76
GPR	1.18	0.55	1.09	0.77
ANN	1.16	0.57	1.11	0.76
Linear Regression	1.24	0.50	1.28	0.68

Table A.28. Data-Driven Model Results: MSDR - Inputs: S_a - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.85	0.89	1.23	0.71
GPR	1.01	0.84	1.50	0.57
ANN	0.93	0.87	1.05	0.79
Linear Regression	1.54	0.64	1.26	0.70

Table A.29. Data-Driven Model Results: MSDR - Inputs: S_a - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.47	0.94	1.03	0.80
GPR	0.50	0.93	1.18	0.73
ANN	0.46	0.95	1.13	0.75
Linear Regression	0.81	0.83	1.28	0.69

Table A.30. Data-Driven Model Results: MSDR - Inputs: S_a - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.61	0.90	1.19	0.73
GPR	0.58	0.91	0.96	0.82
ANN	0.55	0.92	1.01	0.80
Linear Regression	0.75	0.57	1.27	0.69

Table A.31. Data-Driven Model Results: MSDR - Inputs: S_a - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.56	0.90	1.18	0.73
GPR	0.57	0.90	1.05	0.79
ANN	0.55	0.90	1.02	0.80
Linear Regression	0.66	0.86	1.27	0.69

Table A.32. Data-Driven Model Results: MSDR - Inputs: S_v - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.53	0.95	0.75	0.89
GPR	0.62	0.94	0.64	0.92
ANN	0.47	0.96	0.62	0.93
Linear Regression	0.77	0.90	0.82	0.87

Table A.33. Data-Driven Model Results: MSDR - Inputs: S_v - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.33	0.97	0.79	0.88
GPR	0.46	0.95	0.63	0.92
ANN	0.39	0.96	0.60	0.93
Linear Regression	0.51	0.94	0.82	0.87

Table A.34. Data-Driven Model Results: MSDR - Inputs: S_v - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.35	0.97	0.86	0.86
GPR	0.38	0.96	0.87	0.85
ANN	0.40	0.96	0.61	0.93
Linear Regression	0.43	0.95	0.81	0.87

Table A.35. Data-Driven Model Results: MSDR - Inputs: S_v - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.33	0.96	0.82	0.87
GPR	0.34	0.96	0.76	0.89
ANN	0.39	0.95	0.62	0.93
Linear Regression	0.39	0.95	0.80	0.88

Table A.36. Data-Driven Model Results: MSDR - Inputs: S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.68	0.92	1.25	0.70
GPR	0.81	0.66	1.23	0.71
ANN	0.74	0.54	1.1	0.77
Linear Regression	1.17	0.77	1.26	0.67

Table A.37. Data-Driven Model Results: MSDR - Inputs: S_d - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.56	0.92	1.21	0.72
GPR	0.59	0.91	1.31	0.67
ANN	0.65	0.89	1.16	0.74
Linear Regression	0.87	0.75	1.28	0.68

Table A.38. Data-Driven Model Results: MSDR - Inputs: S_d - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.60	0.91	1.21	0.72
GPR	0.67	0.89	0.99	0.81
ANN	0.60	0.91	1.04	0.79
Linear Regression	0.78	0.84	1.28	0.69

Table A.39. Data-Driven Model Results: MSDR - Inputs: S_d - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.56	0.90	1.19	0.73
GPR	0.56	0.90	0.99	0.81
ANN	0.53	0.91	1.06	0.78
Linear Regression	0.74	0.83	1.27	0.69

Table A.40. Data-Driven Model Results: MSDR - Inputs: PGA, S_a - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.36	0.98	0.61	0.93
GPR	0.40	0.98	0.53	0.95
ANN	0.49	0.97	0.51	0.95
Linear Regression	0.50	0.97	0.55	0.94

Table A.41. Data-Driven Model Results: MSDR - Inputs: PGA, S_a - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.37	0.96	0.64	0.92
GPR	0.31	0.97	0.61	0.93
ANN	0.31	0.97	0.57	0.94
Linear Regression	0.38	0.96	0.53	0.95

Table A.42. Data-Driven Model Results: MSDR - Inputs: PGA, S_a - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.40	0.96	0.82	0.87
GPR	0.48	0.94	0.54	0.94
ANN	0.40	0.96	0.56	0.94
Linear Regression	0.42	0.95	0.53	0.94

Table A.43. Data-Driven Model Results: MSDR - Inputs: PGA, S_a - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.34	0.96	0.82	0.87
GPR	0.41	0.95	0.52	0.95
ANN	0.33	0.96	0.55	0.94
Linear Regression	0.36	0.96	0.3	0.95

Table A.44. Data-Driven Model Results: MSDR - Inputs: PGA, S_v - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.73	0.92	0.56	0.94
GPR	0.49	0.96	0.58	0.94
ANN	0.46	0.97	0.53	0.95
Linear Regression	0.66	0.93	0.52	0.95

Table A.45. Data-Driven Model Results: MSDR - Inputs: PGA, S_v - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.40	0.96	0.63	0.92
GPR	0.45	0.95	0.62	0.93
ANN	0.46	0.95	0.57	0.94
Linear Regression	0.51	0.93	0.51	0.95

Table A.46. Data-Driven Model Results: MSDR - Inputs: PGA, S_v - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.42	0.95	0.56	0.94
GPR	0.42	0.95	0.56	0.94
ANN	0.38	0.96	0.58	0.93
Linear Regression	0.45	0.95	0.64	0.92

Table A.47. Data-Driven Model Results: MSDR - Inputs: PGA, S_v - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.33	0.97	0.65	0.92
GPR	0.36	0.96	0.63	0.92
ANN	0.31	0.97	0.58	0.94
Linear Regression	0.44	0.94	0.57	0.94

Table A.48. Data-Driven Model Results: MSDR - Inputs: PGA, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.46	0.97	0.58	0.93
GPR	0.39	0.97	0.53	0.95
ANN	0.53	0.95	0.50	0.95
Linear Regression	0.50	0.96	0.55	0.94

Table A.49. Data-Driven Model Results: MSDR - Inputs: PGA, S_d - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.37	0.97	0.65	0.92
GPR	0.35	0.97	0.57	0.94
ANN	0.33	0.97	0.50	0.95
Linear Regression	0.36	0.97	0.53	0.95

Table A.50. Data-Driven Model Results: MSDR - Inputs: S_a, S_v - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.26	0.99	0.40	0.97
GPR	0.55	0.95	0.58	0.93
ANN	0.51	0.96	0.64	0.92
Linear Regression	0.92	0.87	0.65	0.92

Table A.51. Data-Driven Model Results: MSDR - Inputs: S_a, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.83	0.90	1.22	0.72
GPR	0.83	0.90	1.02	0.80
ANN	0.98	0.86	0.77	0.88
Linear Regression	0.64	0.94	0.89	0.85

Table A.52. Data-Driven Model Results: MSDR - Inputs: S_a, S_d - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.47	0.94	1.11	0.76
GPR	0.50	0.93	1.07	0.78
ANN	0.53	0.93	1.05	0.79
Linear Regression	0.66	0.88	0.84	0.86

Table A.53. Data-Driven Model Results: MSDR - Inputs: S_a, S_d - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.51	0.93	1.18	0.73
GPR	0.49	0.94	0.93	0.83
ANN	0.46	0.94	0.81	0.87
Linear Regression	0.41	0.96	0.82	0.87

Table A.54. Data-Driven Model Results: MSDR - Inputs: S_a, S_d - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.56	0.91	1.21	0.72
GPR	0.52	0.92	0.96	0.82
ANN	0.45	0.94	0.86	0.86
Linear Regression	0.46	0.94	0.81	0.87

Table A.55. Data-Driven Model Results: MSDR - Inputs: S_v, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.29	0.99	0.45	0.96
GPR	0.57	0.95	0.75	0.89
ANN	0.45	0.97	0.63	0.92
Linear Regression	0.83	0.89	0.65	0.92

Table A.56. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_v - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.49	0.96	0.60	0.93
GPR	0.52	0.96	0.56	0.94
ANN	0.36	0.98	0.55	0.94
Linear Regression	0.92	0.86	0.57	0.94

Table A.57. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_v - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.41	0.96	0.60	0.93
GPR	0.34	0.97	0.61	0.93
ANN	0.35	0.97	0.51	0.95
Linear Regression	0.53	0.93	0.49	0.95

Table A.58. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_v - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.43	0.95	0.57	0.94
GPR	0.43	0.95	0.54	0.94
ANN	0.37	0.96	0.57	0.94
Linear Regression	0.40	0.96	0.64	0.92

Table A.59. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_v - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.35	0.96	0.61	0.93
GPR	0.36	0.96	0.63	0.92
ANN	0.33	0.97	0.53	0.95
Linear Regression	0.45	0.94	0.55	0.94

Table A.60. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.55	0.95	0.59	0.93
GPR	0.51	0.96	0.57	0.94
ANN	0.32	0.98	0.51	0.95
Linear Regression	0.54	0.82	0.66	0.92

Table A.61. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_d - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.35	0.97	0.65	0.92
GPR	0.35	0.97	0.59	0.93
ANN	0.30	0.98	0.54	0.94
Linear Regression	0.61	0.90	0.62	0.93

Table A.62. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_d - 75 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.38	0.96	0.59	0.93
GPR	0.41	0.96	0.66	0.92
ANN	0.36	0.97	0.52	0.95
Linear Regression	0.41	0.96	0.66	0.92

Table A.63. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_d - 100 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.41	0.94	0.66	0.92
GPR	0.40	0.95	0.62	0.93
ANN	0.34	0.96	0.55	0.94
Linear Regression	0.43	0.94	0.58	0.93

Table A.64. Data-Driven Model Results: MSDR - Inputs: PGA, S_v, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.84	0.71	0.59	0.93
GPR	0.61	0.94	0.57	0.94
ANN	0.36	0.98	0.52	0.95
Linear Regression	1.23	0.77	0.59	0.93

Table A.65. Data-Driven Model Results: MSDR - Inputs: PGA, S_v, S_d - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.36	0.97	0.60	0.93
GPR	0.32	0.97	0.50	0.95
ANN	0.28	0.98	0.56	0.94
Linear Regression	0.70	0.87	0.49	0.95

Table A.66. Data-Driven Model Results: MSDR - Inputs: S_a, S_v, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.42	0.97	0.35	0.98
GPR	0.65	0.93	0.58	0.94
ANN	0.54	0.95	0.62	0.93
Linear Regression	0.87	0.88	0.65	0.92

Table A.67. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_v, S_d - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.68	0.93	0.63	0.92
GPR	0.57	0.95	0.50	0.95
ANN	0.40	0.98	0.59	0.93
Linear Regression	1.14	0.80	0.53	0.95

Table A.68. Data-Driven Model Results: MSDR - Inputs: PGA, S_a, S_v, S_d - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.45	0.95	0.61	0.93
GPR	0.38	0.96	0.56	0.94
ANN	0.30	0.98	0.50	0.95
Linear Regression	0.71	0.87	0.57	0.94

B. APPENDIX 2: RESULTS FOR DATA-DRIVEN MODELS FOR THE PREDICTION OF PEAK RESPONSES IN PRELIMINARY STUDY 2: ACTUAL SPECTRAL RESPONSE PROPERTIES

Table B.1. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{a\ actual}$ - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.15	0.99	0.59	0.92
GPR	0.22	0.99	0.54	0.93
ANN	0.12	0.996	0.53	0.93
Linear Regression	0.32	0.97	0.59	0.92

Table B.2. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{a\ actual}$ - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.30	0.98	0.51	0.94
GPR	0.39	0.97	0.54	0.93
ANN	0.34	0.97	0.49	0.94
Linear Regression	0.40	0.97	0.43	0.95

Table B.3. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, $S_{a\ actual}$ - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.26	0.98	0.61	0.91
GPR	0.27	0.98	0.58	0.92
ANN	0.16	0.99	0.57	0.92
Linear Regression	0.33	0.97	0.51	0.92

Table B.4. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, $S_{a\ actual}$ - 50 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.19	0.99	0.32	0.97
GPR	0.22	0.99	0.57	0.92
ANN	0.17	0.99	0.43	0.95
Linear Regression	0.32	0.98	0.41	0.96

Table B.5. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{d\ actual}$ - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.08	0.998	0.62	0.91
GPR	0.14	0.995	0.58	0.92
ANN	0.24	0.99	0.45	0.95
Linear Regression	0.23	0.99	0.39	0.96

Table B.6. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{d\ actual}$ - 25 Points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.08	0.998	0.62	0.91
GPR	0.14	0.995	0.58	0.92
ANN	0.24	0.99	0.45	0.95
Linear Regression	0.23	0.99	0.39	0.96

Table B.7. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.29	0.98	0.58	0.92
GPR	0.24	0.99	0.48	0.95
ANN	0.13	0.996	0.52	0.94
Linear Regression	0.27	0.95	0.45	0.95

Table B.8. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.17	0.99	0.63	0.90
GPR	0.21	0.99	0.51	0.94
ANN	0.18	0.99	0.54	0.93
Linear Regression	0.15	0.995	0.51	0.94

Table B.9. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.26	0.99	0.73	0.87
GPR	0.21	0.99	0.84	0.83
ANN	0.31	0.98	0.56	0.92
Linear Regression	0.30	0.98	0.52	0.93

Table B.10. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.26	0.98	0.71	0.89
GPR	0.28	0.98	0.36	0.93
ANN	0.40	0.96	0.58	0.93
Linear Regression	0.14	0.97	0.56	0.93

Table B.11. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.25	0.99	0.87	0.87
GPR	0.23	0.99	0.88	0.87
ANN	0.13	0.97	0.62	0.94
Linear Regression	0.09	0.98	0.57	0.95

Table B.12. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, $S_{a\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.16	0.99	0.63	0.90
GPR	0.29	0.98	0.54	0.93
ANN	0.15	0.995	0.62	0.91
Linear Regression	0.24	0.99	0.50	0.94

Table B.13. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, $S_{a\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.27	0.98	0.52	0.93
GPR	0.27	0.98	0.60	0.91
ANN	0.27	0.98	0.39	0.96
Linear Regression	0.30	0.98	0.52	0.93

Table B.14. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.32	0.98	0.36	0.97
GPR	0.18	0.99	0.37	0.97
ANN	0.10	0.998	0.39	0.96
Linear Regression	0.18	0.99	0.35	0.97

Table B.15. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.18	0.99	0.4	0.96
GPR	0.23	0.99	0.38	0.97
ANN	0.08	0.998	0.39	0.96
Linear Regression	0.19	0.99	0.33	0.97

Table B.16. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.20	0.99	0.68	0.89
GPR	0.26	0.98	0.57	0.92
ANN	0.13	0.996	0.52	0.94
Linear Regression	0.25	0.98	0.50	0.94

Table B.17. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.20	0.99	0.41	0.96
GPR	0.19	0.99	0.39	0.96
ANN	0.20	0.99	0.41	0.96
Linear Regression	0.19	0.99	0.39	0.96

Table B.18. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.15	0.99	0.62	0.91
GPR	0.29	0.98	0.53	0.93
ANN	0.35	0.97	0.53	0.93
Linear Regression	0.20	0.99	0.50	0.94

Table B.19. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.16	0.99	0.36	0.97
GPR	0.24	0.99	1.00	0.75
ANN	0.16	0.99	0.31	0.98
Linear Regression	0.20	0.99	0.22	0.99

Table B.20. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{a\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.09	0.998	0.52	0.94
GPR	0.12	0.996	0.56	0.92
ANN	0.13	0.996	0.54	0.93
Linear Regression	0.21	0.99	0.33	0.97

Table B.21. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.34	0.97	0.46	0.95
GPR	0.19	0.99	0.56	0.93
ANN	0.13	0.996	0.52	0.94
Linear Regression	0.27	0.98	0.37	0.97

Table B.22. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.24	0.99	0.33	0.97
GPR	0.17	0.99	0.59	0.92
ANN	0.26	0.98	0.39	0.96
Linear Regression	0.20	0.99	0.46	0.95

Table B.23. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.22	0.99	0.43	0.96
GPR	0.18	0.99	0.51	0.94
ANN	0.25	0.99	0.51	0.96
Linear Regression	0.19	0.99	0.25	0.99

Table B.24. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ actual}$, $S_{v\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.28	0.98	0.64	0.90
GPR	0.13	0.996	0.65	0.90
ANN	0.08	0.999	0.52	0.94
Linear Regression	0.24	0.99	0.54	0.93

Table B.25. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ actual}$, $S_{v\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.17	0.99	0.55	0.93
GPR	0.24	0.99	0.74	0.87
ANN	0.34	0.97	0.42	0.96
Linear Regression	0.31	0.98	0.42	0.96

Table B.26. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, $S_{a\ actual}$, $S_{v\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.24	0.99	1.39	0.54
GPR	0.23	0.99	0.64	0.90
ANN	0.21	0.99	0.56	0.92
Linear Regression	0.19	0.99	0.58	0.92

Table B.27. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, $S_{a\ actual}$, $S_{v\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.21	0.99	0.36	0.97
GPR	0.25	0.99	0.40	0.96
ANN	0.16	0.99	0.43	0.95
Linear Regression	0.29	0.98	0.33	0.98

Table B.28. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.20	0.99	0.40	0.96
GPR	0.16	0.99	0.59	0.92
ANN	0.44	0.95	0.40	0.96
Linear Regression	0.14	0.995	0.36	0.97

Table B.29. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, $S_{a\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.19	0.99	0.51	0.94
GPR	0.20	0.99	0.42	0.96
ANN	0.14	0.995	0.46	0.95
Linear Regression	0.18	0.99	0.36	0.97

Table B.30. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.24	0.99	0.33	0.97
GPR	0.14	0.996	0.40	0.96
ANN	0.31	0.98	0.39	0.96
Linear Regression	0.21	0.99	0.32	0.98

Table B.31. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.15	0.99	0.62	0.91
GPR	0.26	0.98	0.47	0.95
ANN	0.29	0.98	0.47	0.95
Linear Regression	0.27	0.98	0.26	0.98

Table B.32. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.31	0.98	0.34	0.97
GPR	0.15	0.99	0.72	0.88
ANN	0.28	0.98	0.41	0.96
Linear Regression	0.25	0.98	0.26	0.98

Table B.33. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.16	0.99	0.51	0.94
GPR	0.17	0.99	0.44	0.95
ANN	0.15	0.995	0.41	0.96
Linear Regression	0.21	0.99	0.35	0.97

Table B.34. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.17	0.99	0.33	0.97
GPR	0.15	0.99	0.62	0.91
ANN	0.30	0.98	0.39	0.96
Linear Regression	0.21	0.99	0.32	0.98

Table B.35. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.19	0.99	0.43	0.96
GPR	0.22	0.99	0.46	0.95
ANN	0.14	0.995	0.52	0.94
Linear Regression	0.20	0.99	0.36	0.97

Table B.36. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.63	0.87	1.41	0.48
GPR	0.67	0.86	0.76	0.85
ANN	0.68	0.85	0.78	0.84
Linear Regression	0.52	0.92	0.73	0.86

Table B.37. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.54	0.93	0.73	0.88
GPR	0.64	0.90	0.72	0.88
ANN	0.57	0.92	0.70	0.89
Linear Regression	0.57	0.92	0.72	0.88

Table B.38. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.33	0.97	0.89	0.83
GPR	0.33	0.97	0.75	0.88
ANN	0.36	0.96	0.62	0.92
Linear Regression	0.39	0.96	0.72	0.89

Table B.39. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.35	0.97	0.84	0.88
GPR	0.42	0.95	0.85	0.88
ANN	0.47	0.94	0.80	0.91
Linear Regression	0.40	0.96	0.87	0.87

Table B.40. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.49	0.92	0.73	0.86
GPR	0.71	0.84	0.76	0.85
ANN	0.58	0.89	0.87	0.81
Linear Regression	0.54	0.90	0.68	0.88

Table B.41. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.74	0.87	0.74	0.87
GPR	0.69	0.89	0.73	0.88
ANN	0.69	0.89	0.64	0.91
Linear Regression	0.73	0.88	0.74	0.87

Table B.42. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.42	0.95	0.82	0.86
GPR	0.45	0.94	0.79	0.87
ANN	0.40	0.95	0.77	0.87
Linear Regression	0.43	0.95	0.80	0.86

Table B.43. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.36	0.97	0.89	0.86
GPR	0.40	0.96	0.82	0.88
ANN	0.35	0.97	0.85	0.87
Linear Regression	0.43	0.95	0.83	0.88

Table B.44. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{v\ actual}$, $S_{d\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.73	0.87	0.73	0.88
GPR	0.65	0.90	1.387	0.56
ANN	0.73	0.88	0.72	0.88
Linear Regression	0.71	0.88	0.72	0.88

Table B.45. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{v\ actual}$, $S_{d\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.30	0.97	0.99	0.79
GPR	0.35	0.97	0.83	0.85
ANN	0.35	0.96	0.71	0.89
Linear Regression	0.40	0.95	0.74	0.89

Table B.46. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{v\ actual}$, $S_{d\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.34	0.97	0.75	0.90
GPR	0.40	0.96	0.88	0.87
ANN	0.33	0.97	0.46	0.96
Linear Regression	0.39	0.96	0.82	0.88

Table B.47. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{v\ actual}$, $S_{d\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.53	0.93	0.73	0.88
GPR	0.61	0.91	0.72	0.88
ANN	0.50	0.94	0.80	0.85
Linear Regression	0.54	0.93	0.72	0.88

Table B.48. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{v\ actual}$, $S_{d\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.44	0.94	0.79	0.87
GPR	0.42	0.95	0.85	0.85
ANN	0.48	0.93	0.62	0.92
Linear Regression	0.48	0.93	0.75	0.88

Table B.49. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{v\ actual}$, $S_{d\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.33	0.97	1.00	0.83
GPR	0.38	0.96	0.99	0.83
ANN	0.41	0.96	0.56	0.90
Linear Regression	0.40	0.96	0.81	0.89

Table B.50. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, $S_{v\ actual}$, $S_{d\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.35	0.97	0.56	0.93
GPR	0.43	0.96	0.67	0.90
ANN	0.39	0.97	0.63	0.91
Linear Regression	0.44	0.96	0.63	0.91

Table B.51. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, $S_{v\ actual}$, $S_{d\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.31	0.97	0.67	0.90
GPR	0.33	0.97	0.67	0.91
ANN	0.31	0.97	0.68	0.90
Linear Regression	0.32	0.97	0.67	0.91

Table B.52. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, $S_{v\ actual}$, $S_{d\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.24	0.99	0.74	0.91
GPR	0.26	0.98	0.78	0.90
ANN	0.28	0.98	0.73	0.91
Linear Regression	0.29	0.98	0.73	0.91

Table B.53. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, $S_{v\ actual}$, $S_{d\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	2.09	0	2.08	0
GPR	0.74	0.87	0.63	0.91
ANN	0.53	0.93	0.65	0.90
Linear Regression	0.62	0.86	0.63	0.91

Table B.54. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, $S_{v\ actual}$, $S_{d\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.41	0.95	0.67	0.90
GPR	0.33	0.97	0.67	0.91
ANN	0.31	0.97	0.68	0.90
Linear Regression	0.35	0.96	0.68	0.90

Table B.55. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, $S_{v\ actual}$, $S_{d\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.29	0.98	0.73	0.91
GPR	0.25	0.98	0.74	0.91
ANN	0.28	0.98	0.74	0.90
Linear Regression	0.26	0.98	0.74	0.91

Table B.56. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.85	0.78	1.21	0.62
GPR	0.63	0.88	0.77	0.85
ANN	0.60	0.88	0.77	0.84
Linear Regression	0.71	0.84	0.87	0.81

Table B.57. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.59	0.92	0.71	0.88
GPR	0.79	0.85	0.75	0.87
ANN	0.62	0.91	0.77	0.86
Linear Regression	0.76	0.87	0.73	0.88

Table B.58. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.30	0.97	0.82	0.86
GPR	0.37	0.96	0.82	0.86
ANN	0.29	0.98	0.58	0.93
Linear Regression	0.50	0.93	0.81	0.86

Table B.59. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.40	0.96	0.93	0.85
GPR	0.42	0.95	0.85	0.88
ANN	0.41	0.96	0.69	0.92
Linear Regression	0.45	0.95	0.71	0.91

Table B.60. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.85	0.78	1.21	0.62
GPR	0.73	0.83	0.75	0.85
ANN	0.78	0.80	0.66	0.89
Linear Regression	0.81	0.79	0.72	0.87

Table B.61. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.47	0.95	2.74	-0.73
GPR	0.70	0.89	1.21	0.66
ANN	0.62	0.91	0.83	0.84
Linear Regression	0.97	0.78	0.73	0.88

Table B.62. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.42	0.95	0.82	0.86
GPR	0.34	0.97	0.93	0.82
ANN	0.38	0.96	0.80	0.87
Linear Regression	0.45	0.94	0.78	0.87

Table B.63. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.36	0.97	0.98	0.83
GPR	0.38	0.71	0.84	0.88
ANN	0.32	0.93	0.86	0.87
Linear Regression	0.43	0.95	0.84	0.88

Table B.64. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.49	0.95	0.68	0.89
GPR	0.42	0.96	0.70	0.89
ANN	0.44	0.96	0.65	0.90
Linear Regression	0.77	0.87	0.69	0.89

Table B.65. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.30	0.98	0.67	0.91
GPR	0.37	0.96	0.71	0.89
ANN	0.31	0.97	0.67	0.90
Linear Regression	0.51	0.93	0.70	0.90

Table B.66. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.24	0.98	0.72	0.91
GPR	0.25	0.98	0.75	0.90
ANN	0.26	0.98	0.73	0.91
Linear Regression	0.27	0.98	0.75	0.90

Table B.67. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.60	0.92	0.69	0.89
GPR	0.77	0.86	0.59	0.92
ANN	0.59	0.92	0.65	0.90
Linear Regression	0.79	0.86	0.70	0.89

Table B.68. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 75 points

Model	**Val NRMSE**	**Val R^2**	**Test NRMSE**	**Test R^2**
SVM	0.27	0.98	0.67	0.91
GPR	0.31	0.97	0.70	0.90
ANN	0.31	0.97	0.68	0.90
Linear Regression	0.46	0.94	0.70	0.90

Table B.69. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, $S_{a\ actual}$, $S_{v\ actual}$, $S_{d\ actual}$ - 100 points

Model	**Val NRMSE**	**Val R^2**	**Test NRMSE**	**Test R^2**
SVM	0.28	0.98	0.74	0.90
GPR	0.30	0.98	0.76	0.90
ANN	0.29	0.98	0.72	0.91
Linear Regression	0.32	0.97	0.75	0.90

C. APPENDIX 3: RESULTS FOR DATA-DRIVEN MODELS FOR THE PREDICTION OF PEAK RESPONSES IN PRELIMINARY STUDY 2: MEAN SPECTRAL RESPONSE PROPERTIES

Table C.1. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: S_a - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.15	0.995	0.60	0.91
GPR	0.25	0.99	0.77	0.86
ANN	0.15	0.995	0.47	0.95
Linear Regression	0.45	0.95	0.30	0.98

Table C.2. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, S_a - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.22	0.99	0.85	0.83
GPR	0.12	0.996	0.77	0.86
ANN	0.43	0.96	0.39	0.96
Linear Regression	0.43	0.95	0.31	0.98

Table C.3. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.27	0.98	0.31	0.98
GPR	0.39	0.97	0.25	0.99
ANN	0.20	0.99	0.30	0.98
Linear Regression	0.40	0.97	0.30	0.98

Table C.4. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.30	0.98	0.65	0.90
GPR	0.21	0.99	0.86	0.83
ANN	0.23	0.99	0.49	0.94
Linear Regression	0.39	0.96	0.31	0.98

Table C.5. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, S_a - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.10	0.998	0.20	0.99
GPR	0.26	0.98	0.57	0.92
ANN	0.28	0.98	0.24	0.99
Linear Regression	0.27	0.98	0.29	0.98

Table C.6. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, S_a - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.27	0.98	0.30	0.98
GPR	0.18	0.99	0.79	0.85
ANN	0.30	0.98	0.35	0.97
Linear Regression	0.29	0.98	0.30	0.98

Table C.7. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	1.25	0.66	1.95	0.10
GPR	0.33	0.98	0.33	0.97
ANN	0.25	0.99	0.22	0.99
Linear Regression	0.33	0.98	0.31	0.98

Table C.8. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.31	0.98	0.31	0.98
GPR	0.21	0.99	0.79	0.85
ANN	0.34	0.97	0.26	0.98
Linear Regression	0.28	0.98	0.30	0.98

Table C.9. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: S_a, S_v - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.40	0.96	0.47	0.95
GPR	0.24	0.99	0.85	0.83
ANN	0.15	0.995	0.45	0.95
Linear Regression	0.32	0.98	0.32	0.98

Table C.10. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, S_a, S_v - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.25	0.98	0.52	0.93
GPR	0.37	0.97	0.47	0.95
ANN	0.26	0.98	0.51	0.94
Linear Regression	0.32	0.97	0.32	0.98

Table C.11. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: S_a, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.11	0.997	0.62	0.91
GPR	0.22	0.99	1.37	0.55
ANN	0.21	0.99	0.42	0.96
Linear Regression	0.38	0.98	0.35	0.97

Table C.12. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, S_a, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.25	0.99	0.59	0.92
GPR	0.12	0.996	1.34	0.57
ANN	0.29	0.98	0.31	0.98
Linear Regression	0.36	0.97	0.35	0.97

Table C.13. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: S_v, S_a - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.25	0.98	0.97	0.78
GPR	0.19	0.99	0.85	0.83
ANN	0.46	0.95	0.24	0.99
Linear Regression	0.33	0.97	0.32	0.98

Table C.14. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.21	0.99	0.47	0.95
GPR	0.23	0.99	0.81	0.84
ANN	0.29	0.98	0.47	0.95
Linear Regression	0.32	0.98	0.32	0.98

Table C.15. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, S_a, S_v - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.70	0.88	2.77	-0.82
GPR	0.26	0.98	0.57	0.92
ANN	0.23	0.99	0.56	0.93
Linear Regression	0.30	0.98	0.29	0.98

Table C.16. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, S_a, S_v - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.56	0.93	1.77	0.26
GPR	0.25	0.99	0.26	0.98
ANN	0.24	0.99	0.31	0.98
Linear Regression	0.42	0.96	0.41	0.96

Table C.17. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, S_a, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.23	0.99	0.20	0.99
GPR	0.23	0.99	0.25	0.98
ANN	0.30	0.98	0.26	0.98
Linear Regression	0.29	0.98	0.29	0.98

Table C.18. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, S_a, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.21	0.99	0.35	0.97
GPR	0.23	0.99	0.75	0.86
ANN	0.32	0.97	0.28	0.98
Linear Regression	0.31	0.98	0.29	0.98

Table C.19. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.21	0.99	0.38	0.97
GPR	0.13	0.996	0.61	0.91
ANN	0.30	0.98	0.36	0.97
Linear Regression	0.19	0.99	0.33	0.98

Table C.20. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.29	0.98	1.11	0.71
GPR	0.22	0.99	0.76	0.86
ANN	0.28	0.98	0.32	0.98
Linear Regression	0.31	0.98	0.29	0.98

Table C.21. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: S_a, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.23	0.99	0.86	0.82
GPR	0.23	0.99	1.00	0.76
ANN	0.40	0.96	0.25	0.99
Linear Regression	0.38	0.96	0.29	0.98

Table C.22. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, S_a, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.26	0.98	0.59	0.92
GPR	0.39	0.96	0.45	0.95
ANN	0.27	0.98	0.48	0.95
Linear Regression	0.39	0.96	0.29	0.98

Table C.23. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: PGA, S_a, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.31	0.98	0.29	0.98
GPR	0.24	0.99	0.50	0.94
ANN	0.32	0.98	0.25	0.99
Linear Regression	0.28	0.98	0.31	0.98

Table C.24. Data-Driven Model Results: Varying Structural Properties: MRDR - Inputs: E, ρ, PGA, S_a, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.28	0.98	0.34	0.97
GPR	0.24	0.98	0.57	0.92
ANN	0.35	0.97	0.34	0.97
Linear Regression	0.36	0.97	0.46	0.95

Table C.25. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_a, S_v - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.45	0.93	0.87	0.81
GPR	0.60	0.87	0.81	0.83
ANN	0.60	0.87	0.85	0.78
Linear Regression	0.60	0.87	0.81	0.83

Table C.26. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_a, S_v - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.37	0.97	1.04	0.75
GPR	0.36	0.97	1.51	0.48
ANN	0.46	0.95	0.55	0.93
Linear Regression	0.84	0.84	0.86	0.83

Table C.27. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_a, S_v - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.40	0.95	0.77	0.87
GPR	0.51	0.92	0.83	0.86
ANN	0.40	0.95	0.60	0.92
Linear Regression	0.51	0.92	0.84	0.85

Table C.28. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_a, S_v - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.49	0.94	0.78	0.90
GPR	0.53	0.93	1.00	0.83
ANN	0.58	0.91	0.81	0.89
Linear Regression	0.64	0.89	0.82	0.88

Table C.29. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_a, S_v - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.67	0.86	0.87	0.81
GPR	0.63	0.88	0.82	0.83
ANN	0.63	0.88	0.82	0.83
Linear Regression	0.77	0.82	0.84	0.82

Table C.30. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_a, S_v - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.62	0.91	0.76	0.87
GPR	0.49	0.94	0.98	0.78
ANN	0.54	0.93	0.95	0.79
Linear Regression	0.75	0.87	0.75	0.87

Table C.31. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_a, S_v - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.44	0.94	0.81	0.86
GPR	0.40	0.95	0.90	0.83
ANN	0.37	0.96	0.71	0.89
Linear Regression	0.51	0.92	0.81	0.86

Table C.32. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_a, S_v - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.53	0.92	0.96	0.83
GPR	0.54	0.92	0.98	0.83
ANN	0.58	0.91	0.75	0.90
Linear Regression	0.63	0.90	0.89	0.86

Table C.33. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_v, S_d - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.35	0.97	1.00	0.77
GPR	0.34	0.97	1.21	0.66
ANN	0.47	0.95	0.55	0.93
Linear Regression	0.76	0.87	0.78	0.86

Table C.34. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_v, S_d - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.37	0.96	0.87	0.84
GPR	0.42	0.95	0.83	0.86
ANN	0.30	0.97	0.51	0.94
Linear Regression	0.44	0.94	0.84	0.85

Table C.35. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_v, S_d - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.48	0.94	0.72	0.91
GPR	0.58	0.91	0.84	0.88
ANN	0.60	0.91	0.73	0.91
Linear Regression	0.63	0.90	0.82	0.88

Table C.36. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_v, S_d - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.94	0.89	0.84	0.84
GPR	0.49	0.94	1.52	0.47
ANN	0.82	0.84	0.67	0.90
Linear Regression	0.92	0.80	0.84	0.84

Table C.37. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_v, S_d - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.42	0.95	0.84	0.85
GPR	0.46	0.94	1.19	0.70
ANN	0.39	0.96	0.72	0.89
Linear Regression	0.47	0.93	0.88	0.84

Table C.38. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_v, S_d - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.55	0.92	0.74	0.90
GPR	0.53	0.93	0.93	0.85
ANN	0.71	0.88	0.46	0.96
Linear Regression	0.87	0.81	0.87	0.87

Table C.39. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, S_v, S_d - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.61	0.91	0.66	0.90
GPR	0.43	0.96	0.69	0.89
ANN	0.54	0.93	0.43	0.96
Linear Regression	0.73	0.87	0.69	0.89

Table C.40. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, S_v, S_d - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.32	0.97	0.69	0.90
GPR	0.33	0.97	0.69	0.90
ANN	0.30	0.97	0.68	0.90
Linear Regression	0.33	0.97	0.69	0.90

Table C.41. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, S_v, S_d - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.50	0.94	0.78	0.90
GPR	0.40	0.96	0.90	0.86
ANN	0.45	0.95	0.79	0.89
Linear Regression	0.46	0.95	0.79	0.89

Table C.42. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, S_v, S_d - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.51	0.94	0.68	0.89
GPR	0.69	0.47	0.63	0.91
ANN	0.51	0.94	0.75	0.87
Linear Regression	0.76	0.87	0.68	0.89

Table C.43. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, S_v, S_d - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.33	0.97	0.69	0.90
GPR	0.33	0.97	0.67	0.91
ANN	0.38	0.96	0.69	0.90
Linear Regression	0.35	0.96	0.63	0.90

Table C.44. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, S_v, S_d - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.37	0.97	0.76	0.90
GPR	0.37	0.97	0.72	0.91
ANN	0.52	0.94	0.79	0.89
Linear Regression	0.50	0.94	0.94	0.89

Table C.45. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_a, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.34	0.96	1.26	0.59
GPR	0.74	0.81	1.03	0.73
ANN	0.56	0.89	0.86	0.81
Linear Regression	0.75	0.81	0.87	0.81

Table C.46. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_a, S_v, S_d - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.35	0.97	1.00	0.77
GPR	0.36	0.97	0.85	0.83
ANN	0.42	0.96	0.99	0.77
Linear Regression	0.77	0.86	0.81	0.85

Table C.47. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_a, S_v, S_d - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.38	0.96	0.89	0.83
GPR	0.43	0.94	0.87	0.84
ANN	0.40	0.95	0.53	0.94
Linear Regression	0.67	0.87	0.81	0.86

Table C.48. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: S_a, S_v, S_d - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.46	0.95	1.13	0.78
GPR	0.48	0.94	0.92	0.85
ANN	0.47	0.94	0.69	0.92
Linear Regression	0.61	0.90	0.89	0.86

Table C.49. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_a, S_v, S_d - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.59	0.89	0.87	0.81
GPR	0.70	0.84	0.84	0.82
ANN	0.62	0.87	0.74	0.86
Linear Regression	0.78	0.80	0.87	0.80

Table C.50. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_a, S_v, S_d - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.69	0.89	0.79	0.86
GPR	0.44	0.95	1.02	0.76
ANN	0.56	0.93	0.90	0.81
Linear Regression	0.79	0.86	0.81	0.85

Table C.51. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_a, S_v, S_d - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.43	0.95	1.06	0.80
GPR	0.38	0.96	0.95	0.84
ANN	0.59	0.90	0.95	0.84
Linear Regression	0.66	0.87	0.98	0.83

Table C.52. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, S_a, S_v, S_d - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.58	0.91	0.75	0.90
GPR	0.61	0.90	0.96	0.84
ANN	0.59	0.91	0.71	0.91
Linear Regression	0.65	0.89	0.93	0.85

Table C.53. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, S_a, S_v, S_d - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.51	0.94	0.70	0.89
GPR	0.46	0.95	0.92	0.81
ANN	0.56	0.93	0.55	0.93
Linear Regression	0.74	0.87	0.51	0.94

Table C.54. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, S_a, S_v, S_d - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.32	0.97	0.67	0.91
GPR	0.37	0.96	0.71	0.89
ANN	0.31	0.97	0.68	0.90
Linear Regression	0.53	0.92	0.73	0.89

Table C.55. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: PGA, S_a, S_v, S_d - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.40	0.96	0.77	0.90
GPR	0.49	0.94	0.84	0.88
ANN	0.41	0.96	0.74	0.90
Linear Regression	0.50	0.93	0.79	0.89

Table C.56. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, S_a, S_v, S_d - 50 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.78	0.89	0.79	0.86
GPR	0.44	0.95	1.02	0.76
ANN	0.56	0.93	0.90	0.81
Linear Regression	0.79	0.86	0.81	0.85

Table C.57. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, S_a, S_v, S_d - 75 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.34	0.97	0.68	0.90
GPR	0.33	0.97	0.72	0.89
ANN	0.32	0.97	0.68	0.90
Linear Regression	0.42	0.95	0.75	0.88

Table C.58. Data-Driven Model Results: Varying Structural Properties: MSDR - Inputs: E, ρ, PGA, S_a, S_v, S_d - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.40	0.96	0.76	0.90
GPR	0.38	0.96	0.86	0.87
ANN	0.21	0.96	0.76	0.90
Linear Regression	0.50	0.93	0.79	0.89

D. APPENDIX 4: RESULTS FOR DATA-DRIVEN MODELS FOR THE PREDICTION OF PEAK RESPONSES: FULL STUDY

Table D.1. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{a\ mean}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.12	0.997	0.32	0.98
GPR	0.19	0.99	0.37	0.97
ANN	0.10	0.998	0.32	0.98
Linear Regression	0.13	0.996	0.30	0.98

Table D.2. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{d\ mean}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.11	0.998	0.27	0.98
GPR	0.12	0.997	0.28	0.98
ANN	0.11	0.998	0.25	0.99
Linear Regression	0.11	0.998	0.29	0.98

Table D.3. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: PGA, $S_{d\ mean}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.16	0.99	0.24	0.99
GPR	0.16	0.99	0.25	0.99
ANN	0.20	0.99	0.24	0.99
Linear Regression	0.14	0.996	0.25	0.99

Table D.4. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{a\ mean}$, $S_{d\ mean}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.10	0.998	0.28	0.98
GPR	0.11	0.997	0.25	0.99
ANN	0.08	0.998	0.26	0.98
Linear Regression	0.10	0.998	0.25	0.99

Table D.5. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{v\ mean}$, $S_{d\ mean}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.13	0.996	0.35	0.97
GPR	0.12	0.996	0.26	0.98
ANN	0.11	0.997	0.24	0.99
Linear Regression	0.13	0.996	0.24	0.99

Table D.6. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{a\ mean}$, $S_{v\ mean}$, $S_{d\ mean}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.22	0.99	0.31	0.98
GPR	0.25	0.99	0.27	0.98
ANN	0.18	0.99	0.31	0.98
Linear Regression	0.25	0.98	0.27	0.98

Table D.7. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ mean}$, $S_{v\ mean}$, $S_{d\ mean}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.09	0.998	0.25	0.99
GPR	0.07	0.999	0.31	0.98
ANN	0.08	0.998	0.26	0.98
Linear Regression	0.09	0.998	0.25	0.99

Table D.8. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{d\ act}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.10	0.998	0.43	0.96
GPR	0.21	0.99	0.41	0.96
ANN	0.12	0.997	0.41	0.96
Linear Regression	0.11	0.997	0.41	0.96

Table D.9. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: PGA, $S_{d\ act}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.10	0.997	0.38	0.97
GPR	0.09	0.998	0.38	0.97
ANN	0.12	0.996	0.38	0.97
Linear Regression	0.10	0.998	0.36	0.97

Table D.10. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{a\ act}$, $S_{d\ act}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.30	0.98	0.39	0.96
GPR	0.11	0.997	0.46	0.95
ANN	0.11	0.997	0.49	0.95
Linear Regression	0.14	0.996	0.46	0.95

Table D.11. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{v\ act}$, $S_{d\ act}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.17	0.99	0.38	0.97
GPR	0.10	0.998	0.38	0.97
ANN	0.21	0.99	0.32	0.98
Linear Regression	0.17	0.99	0.40	0.96

Table D.12. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.18	0.99	0.56	0.93
GPR	0.18	0.99	0.43	0.96
ANN	0.16	0.99	0.45	0.95
Linear Regression	0.17	0.99	0.34	0.97

Table D.13. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MRDR - Inputs: PGA, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$ - 25 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.19	0.99	0.46	0.95
GPR	0.17	0.99	0.49	0.94
ANN	0.18	0.99	0.45	0.95
Linear Regression	0.16	0.99	0.46	0.95

Table D.14. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: $S_{v\ mean}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.73	0.87	0.89	0.86
GPR	0.83	0.82	0.89	0.86
ANN	0.75	0.86	0.78	0.89
Linear Regression	0.84	0.82	0.96	0.84

Table D.15. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: $S_{a\ mean}$, $S_{v\ mean}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.68	0.88	0.90	0.85
GPR	0.75	0.86	1.14	0.77
ANN	0.75	0.86	0.81	0.88
Linear Regression	0.88	0.80	0.84	0.87

Table D.16. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: $S_{v\ mean}$, $S_{d\ mean}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.71	0.87	0.83	0.88
GPR	0.78	0.84	0.99	0.83
ANN	0.71	0.87	0.76	0.90
Linear Regression	0.81	0.83	0.84	0.87

Table D.17. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: PGA, $S_{v\ mean}$, $S_{d\ mean}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.55	0.92	0.69	0.92
GPR	0.53	0.93	0.70	0.91
ANN	0.53	0.93	0.70	0.91
Linear Regression	0.53	0.93	0.70	0.91

Table D.18. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: $S_{a\ mean}$, $S_{v\ mean}$, $S_{d\ mean}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.68	0.88	1.31	0.69
GPR	0.76	0.85	1.01	0.82
ANN	0.78	0.85	0.58	0.94
Linear Regression	1.00	0.74	0.94	0.84

Table D.19. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: PGA, $S_{a\ mean}$, $S_{v\ mean}$, $S_{d\ mean}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.61	0.91	1.40	0.65
GPR	0.63	0.90	0.80	0.89
ANN	0.65	0.89	0.64	0.93
Linear Regression	0.61	0.91	0.75	0.90

Table D.20. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: $S_{v\ act}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.62	0.90	0.88	0.86
GPR	0.60	0.91	0.82	0.88
ANN	0.60	0.91	0.76	0.90
Linear Regression	0.57	0.91	0.82	0.88

Table D.21. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: $S_{a\ act}$, $S_{v\ act}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.58	0.92	0.87	0.86
GPR	0.75	0.86	0.90	0.86
ANN	0.64	0.90	0.88	0.86
Linear Regression	0.71	0.87	0.84	0.87

Table D.22. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: $S_{v\ act}$, $S_{d\ act}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.58	0.91	0.95	0.84
GPR	0.57	0.92	0.82	0.88
ANN	0.60	0.91	0.85	0.87
Linear Regression	0.54	0.93	0.84	0.87

Table D.23. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: PGA, $S_{v\ act}$, $S_{d\ act}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.34	0.97	1.20	0.74
GPR	0.40	0.96	1.41	0.65
ANN	0.44	0.95	1.35	0.67
Linear Regression	0.57	0.92	1.15	0.77

Table D.24. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.58	0.92	0.90	0.86
GPR	0.58	0.91	0.95	0.84
ANN	0.58	0.91	0.86	0.87
Linear Regression	0.66	0.89	0.77	0.89

Table D.25. Data-Driven Model Results: High-Fidelity Data: Varying Structural Properties: MSDR - Inputs: PGA, $S_{a\ act}$, $S_{v\ act}$, $S_{d\ act}$ - 100 points

Model	Val NRMSE	Val R^2	Test NRMSE	Test R^2
SVM	0.44	0.95	1.40	0.65
GPR	0.45	0.95	1.57	0.56
ANN	0.45	0.95	1.42	0.64
Linear Regression	0.60	0.91	1.43	0.63