Title
Evaluating the effectiveness of ground motion intensity measures for structural response simulation using statistical and causal inferencing

Author
Burton, Henry

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Evaluating the effectiveness of ground motion intensity measures for structural response simulation using statistical and causal inferencing

A thesis submitted in partial satisfaction of the

requirements for the degree

Master of Science in Statistics

by

Henry Burton

2022

ABSTRACT OF THE THESIS

Evaluating the effectiveness of ground motion intensity measures for structural response simulation using statistical and causal inferencing

by

Henry Burton

Master of Science in Statistics

University of California, Los Angeles, 2022

Professor Yingnian Wu, Chair

The sufficiency criterion has long been used to evaluate the effectiveness of a ground motion intensity measure (IM) in capturing the link between ground shaking and structural response. However, a typical sufficiency-based evaluation of an IM only tests for the possibility of linear dependency and the interaction among the upstream parameters is not considered. To address these and other limitations, two new IM evaluation methodologies are proposed. The first methodology considers the loss of statistical information when an IM is used to predict the engineering demand parameters ($EDPs$) without including the upstream parameters (i.e., earthquake magnitude, source-to-site distance and epsilon). The best IM is the one that minimizes the loss of predictive performance when it is the only model input relative to when it is used as a predictor together with the upstream parameters. To consider the possible interactive effects, a machine learning model is used when both the IM and upstream parameters are used as inputs. The second methodology uses

a causal inference approach where the effect of the IM on the EDP distribution is quantified while considering the earthquake magnitude, source-to-site distance and epsilon as control variables. The double machine learning approach is implemented for this purpose. The two methodologies are applied to a set of five steel specifical moment resisting frames. The results show that the statistical loss-based and causal inferencing approaches produce results that are more conclusive than the sufficiency-based approach and more consistent with the physical laws that govern the IM-EDP relationship.

The thesis of Henry Burton is approved.

Hongquan Xu

Mark Stephen Handcock

Yingnian Wu, Committee Chair

University of California, Los Angeles

2022

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Structural response simulation or nonlinear response history analysis ($NRHAs$) is a central part of the performance-based earthquake engineering ($PBEE$) procedure (Moehle and Deierlein, 2004). Within the context of $PBEE$, the goal of $NRHAs$ is to quantify the distribution of engineering demand parameters ($EDPs$) as a function of a scalar or vector-valued ground motion intensity measure (IM). The quantitative description of an IM can be exclusively based on the properties of the ground motion (e.g., peak ground acceleration) or can contain information about both the ground motion and the structure of interest (e.g., spectral acceleration averaged over a range of periods). Several criteria have been used to evaluate the effectiveness of an IM in capturing the link between ground shaking and structural response. Efficiency, sufficiency, scaling robustness and predictability are the most widely used criteria for evaluating the optimality of an IM.

An efficient IM is one that minimizes the variability in the $EDPs$ from $NRHAs$. This variability is also related to the number of ground motions used in the analysis (Eads et al. 2015). Numerous studies have been performed to evaluate the relative efficiency of alternative IMs. The earliest ones focused on demonstrating the superior performance of spectral acceleration at the first mode period of the structure (Sa_{T1}) relative to peak ground acceleration (PGA) in predicting displacement-based $EDPs$ (e.g., peak story drifts) (Shome and Cornell, 1999). Later studies however, showed that Sa_{T1} may not be as efficient when predicting the responses of buildings that respond to a range of periods due to higher-mode effects or inelastic behavior (e.g., Luco and

Cornell, 2007; Tothong and Luco, 2007). The efficiency of Sa_{T1} was shown to be enhanced by including a second spectral acceleration parameter conditioned on another period (e.g., 2nd mode, longer period) (e.g., Mehanny and Deierlein, 2000; Cordova et al. 2000). More recent studies have shown that using the geometric mean spectral acceleration over a range of periods (Sa_{avg}) also provided benefits in terms of efficiency (e.g., Eads et al. 2015; Shokrabadi and Burton. 2017).

A key assumption in $NRHAs$ is that, conditioned on the shaking intensity level, the resulting EDP distribution only depends on the chosen IM and not on other "high-level" or "upstream" properties such as the event magnitude (M), the source-to-site distance (R) and epsilon (ε). Also used as a proxy for spectral shape, ε measures the number of logarithmic standard deviations between the observed and predicted shaking intensity (Baker, 2011). A sufficient IM is one that does not violate this assumption. The sufficiency criterion is desirable because it facilitates performing $NRHAs$ using ground motions from a diverse set of events (in terms of M, R and ε) to evaluate the structural response distribution at some predefined intensity level (Eads et al. 2015). Several studies have evaluated the sufficiency of different IMs with respect to M, R and ε. This is typically done as part of an overall comparative assessment of the effectiveness of different IMs where efficiency is also considered (e.g., Shome and Cornell, 1999; Luco and Cornell, 2007; Tothong and Luco, 2007; Eads et al. 2015; Shokrabadi and Burton. 2017; Bradley et al. 2010). Sufficiency has also been used to evaluate the potential bias that is introduced when ground motions are scaled to achieve specific intensity levels (e.g., Bradley et al. 2010).

A key challenge with using sufficiency to evaluate the relative effectiveness of IMs is that the results are often inconclusive. Consider the case where an IM is sufficient with respect to a particular ground motion parameter at some fraction of the considered intensity levels (e.g., Shokrabadi and Burton. 2017). Similarly, when considering the spatial distribution of $EDPs$ over

the height of a structure, an IM may achieve sufficiency at some floor levels (or stories) but not at others (e.g., Bradley et al. 2010). There may also be situations where two IMs being compared are found to be both sufficient (or insufficient) with respect to the same ground motion parameters. Another limitation of the sufficiency-based evaluation is that it is based on linear dependence i.e., nonlinear dependencies are not considered. Also, the sufficiency criterion is able to consider the dependence on individual ground motion parameters (e.g., between the EDP and M or between the EDP and R). However, it does not consider any potential interaction effects between two or more parameters (e.g., between the EDP and the joint distribution of M and R or between the EDP and the joint distribution of M, R and ε).

1.2 Objectives

This study presents two new approaches to evaluating the relative effectiveness of two or more ground motion intensity measures in estimating structural response demands. The first methodology considers the statistical information that is lost when only the IM of choice is used to obtain the EDP distribution relative to when the IM is used as a predictor together with other ground motion parameters. The loss of statistical information is measured as the root mean square difference between the $EDPs$ obtained from the IM-only estimator i.e., $EDP = f(IM)$ and the $EDPs$ obtained from the IM and the other ground motion parameters e.g., $EDP = f(IM, M, R, \varepsilon)$. The IM that minimizes the relative loss of statistical information is deemed the most desirable. To consider the nonlinear and interactive relationships between the ground motion parameters and the EDP distribution, a machine learning model is used to estimate $f(IM, M, R, \varepsilon)$.

The second methodology uses a causal inference approach to evaluate the relative efficacy of different IMs. Considering the IM as the treatment and the EDP as the outcome of interest, the causal effect is evaluated while considering M, R and $, \varepsilon$ as control variables. To avoid any

assumptions about how the controls enter into the regression, a partial linear model is used. More specifically, a double machine learning approach is adopted where separate models that predict the treatment and outcome as a function of the controls are developed. The effect of the IM on the EDP is then obtained through residual-on-residual regression.

The two new methodologies are applied to a set of special steel moment frames ($SMRFs$) designed to modern code-confirming standards.

1.3 Organization

The thesis is comprised of five chapters which are organized as follows:

Chapter 1 outlines the background motivation and objectives of the study.

Chapter 2 describes the considered building cases including the structural modeling and is followed by a summary of the ground motions used in the study.

Chapter 3 presents the new methodology for evaluating the efficacy of IMs used in $NRHAs$ that is based on minimizing the loss of statistical information in the predicted $EDPs$ relative to when other "upstream" seismic parameters are used.

Chapter 4 introduces an approach based on casual inferencing for assessing the adequacy of IMs used in $NRHAs$.

Chapter 5 presents a case study that applies the new (statistical loss minimization and causal inference) and traditional (i.e., efficiency and sufficiency) approaches to assessing the relative effectiveness of ground motion intensity measures.

Chapter 6 discusses the main conclusions and limitations of this study and gives suggestions for future related work.

Chapter 2

Building Cases, Structural Models and Ground Motions

This chapter describes the buildings, structural models and ground motions used in the nonlinear response history analyses.

2.1 Description of Building Cases

The buildings used in the current study are part of the database developed by Guan et al. (2020). The overall database is comprised of 621 steel reduced-beam section (RBS) $SMRF$ buildings designed in accordance with the ASCE 7-16 standard (ASCE, 2016) and the provisions specified in AISC 341-16 (AISC, 2016a), AISC 360-16 (AISC, 2016b) and AISC 358-16 (AISC, 2016c). Five of those buildings are used to evaluate the effectiveness of alternative IMs for estimating EDP distributions via nonlinear response history analyses. The buildings contain 1, 5, 9, 14 and 19 stories with two 5-bay $SMFs$ in each direction and typical bay widths and story heights of 30'-0" (9.14 m) and 13'-0" (3.96 m) , respectively. The overall plan configuration and structural framing layout is based on the archetypes used in the ATC-123 project (FEMA, 2018). The typical floor dead load is 80 psf (3.83 kN/m^2) and the roof dead load is 67.5 psf (3.23 kN/m^2). The typical floor and roof live loads are 50 psf (2.39 kN/m^2) and 20 psf (0.96 kN/m^2), respectively.

The designs are based on a location in Los Angeles, California $(34.008°\text{N}, 118.152°\text{W})$ with site class D and spectral intensity parameters $S_s = 2.25g$ and $S_1 = 0.6g$. The seismic weights and periods obtained from the equation specified in Chapter 12 of ASCE 7-16 are summarized in Table 1. Assuming that the buildings are used for office space with Risk Category II and Seismic Design Category D, the equivalent lateral force procedure is used to design all

buildings with a seismic response coefficient $R = 8$, importance factor $I = 1.0$, drift amplification factor $C_D = 5.5$ and overstrength factor $\Omega_0 = 3$. A design drift limit of 2% is considered and the minimum strong-column-weak-beam ($SCWB$) ratio is taken to be 1.0.

Table 2.1 Number of stories, seismic weight and ASCE 7-16 and 1st mode period (from eigen value analyses) for the five buildings

Number of Stories	Seismic Weight, kips (kN)	ASCE 7-16 Period (s)	1st Mode Period (s)
1	1800 (8007)	0.21	0.41
5	8719 (38782)	0.79	1.21
9	15919 (70810)	1.26	1.41
14	24919 (110844)	1.80	1.78
19	33919 (150878)	2.30	2.20

2.2 Structural Models

Two-dimensional (2D) nonlinear structural models constructed in the Open System of Earthquake Engineering (OpenSees) (Mazzoni et al. 2006) are also included as part of the Guan et al. (2020) database. Beams and columns are modeled using linear elastic beam-column elements with inelastic flexural springs at the ends. The modified Ibarra-Medina-Krawinkler (IMK) material model is applied to the zero-length flexural hinges to capture both in-cycle and cyclic deterioration. The 24 parameters that define the IMK material model are computed using the empirical equations developed and advanced by Ibarra et al. (2005), Lignos (2008) and Lignos and Krawinkler (2010). The shear behavior of the panel zones are based on the modeling approach developed by Krawinkler (1978). To account for the P-Δ effects caused by loads on the gravity framing not explicitly modeled, a truss element is used as a leaning column. A Rayleigh damping ratio of 2% is applied to the first and third modal periods of the structure.

2.3 Ground Motions

The 240 unscaled ground motions assembled by Miranda (1999) are used to analyze the nonlinear structural models. All ground motions in the set are recorded in California at sites that contain rock or firm soil where the shear wave velocity in the upper 30 m (V_{s30}) of the site profile exceeds 600 ft/s (180 m/s). The event magnitudes range from **M** 6.0 to **M** 7.0 with an average of **M** 6.7. The PGA range is from 0.03 to 0.61g. The individual and median ground motion spectra for the record-set is shown in Figure 2.1. Histograms showing the empirical distribution of PGA and PGV are presented in Figure 2.2. The PGA range is from 0.029 to 0.776 g with a mean value of 0.175 g. The minimum and maximum PGV is 1.0 cm/s and 113.44 cm/s, respectively and the mean value is 17.15 cm/s. Figure 2.3 plots histograms with the empirical distributions of the event magnitude, Boore-Joyner distance (R_{jb}) and epsilon. The event magnitude range is from **M** 5.8 to **M** 7.7 with a mean of 6.7. The minimum R_{jb} is 0.6 km and the maximum is 156 km. The mean R_{jb} for the record-set is 38.2 km. As shown in Figure 2.3c, the ϵ are approximately normally distributed with a near-zero (-0.31) mean value. The lower and upper limit for ϵ is -3.03 and 2.96, respectively.

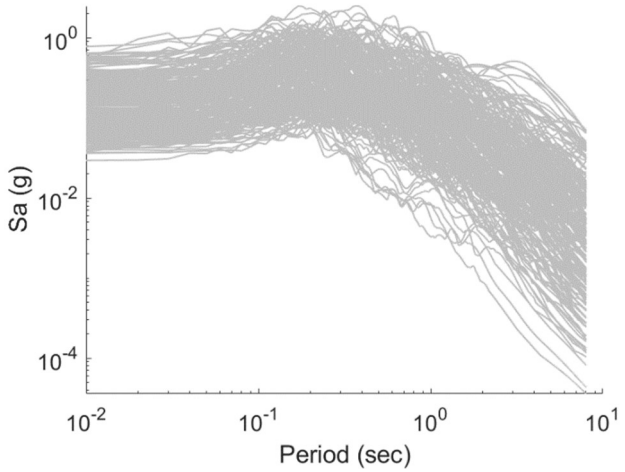

Figure 2.1 – Response spectra for the 240 unscaled ground motions used in the nonlinear response history analyses

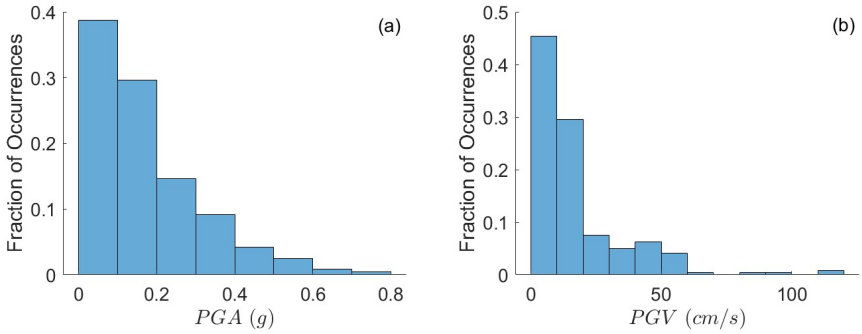

Figure 2.2 – Histograms showing the empirical distribution of (a) PGA and (b) PGV for the 240 unscaled ground motions

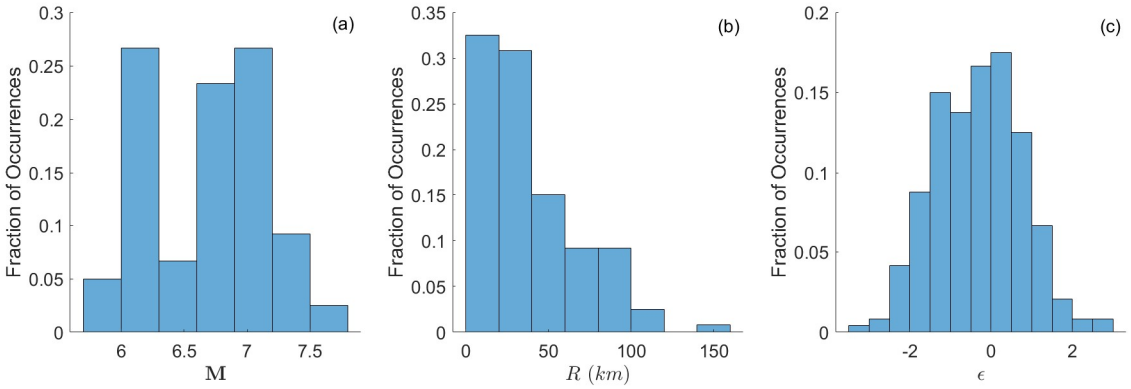

Figure 2.3 – Histograms showing the empirical distribution of (a) M, (b) R_{jb} and (c) ϵ for the 240 unscaled ground motions

Chapter 3

A Statistical Loss Minimization Approach to Evaluating the Effectiveness of Ground Motion Intensity Measures

Ground motion intensity measures are computed as a function of the earthquake magnitude, fault distance, and other seismicity parameters (e.g., fault type). As noted earlier, the epsilon parameter is derived from the adopted ground motion model and is used as a measure of spectral shape. When performing $NRHAs$, it is desirable that, conditioned on that IM, the EDP distribution is independent of these high-level parameters (M, R, ε). The sufficiency criteria achieves that goal by using linear regression to evaluate this independence property for individual parameters.

$$E[\ln(EDP)|X_i, IM = im] = \beta_0 + \beta_1 X_i \tag{3.1}$$

where X_i is the high-level parameter of interest (M, R, ε) and β_1 is the regression coefficient that is used to evaluate the independence property. Under the null hypothesis that $\beta_1 = 0$, a p-value above some predefined threshold (usually 5%) means that $\ln(EDP)$ is independent of X_i conditioned on the IM level. This "one-parameter-at-a-time" approach to evaluating the effectiveness of an IM using the sufficiency criterion is useful because it provides insight into whether the resulting EDP distribution will be biased against the considered parameter. However, aside from relying on the linearity assumption for the relationship between $\ln(EDP)$ and X_i, the joint distribution or interaction effects between the X_i's is not considered.

The conditional independence property between the EDP and the joint distribution of X_i's can also be expressed as the desire to have $f(IM) \cong f(IM, \mathbf{X})$ where $f(IM)$ is the EDP conditional expectation estimator that based only on the IM and $f(IM, \mathbf{X})$ is the EDP conditional

expectation estimator based on the IM in addition to all high-level parameters i.e., $[\mathbf{X}] = \lfloor M \quad R \quad \varepsilon \rfloor$. Then, the most effective IM in terms of the conditional independence property (IM_*) is taken as the one that minimizes the loss of predictive performance in $f(IM)$ relative to $f(IM, \mathbf{X})$. The relative loss of predictive performance can be computed as the root mean square error ($RMSE$) between $f(IM)$ and $f(IM, \mathbf{X})$ or, more formally,

$$IM_* = \underset{IM_j}{\operatorname{argmin}} \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(f\left(IM_j, M, R, \varepsilon\right) - f\left(IM_j\right)\right)^2} \tag{3.2}$$

where IM_j is the intensity measure being considered and N is the number of ground motion records, $f\left(IM_j\right)$ is the EDP expectation estimator conditioned only on IM_j and $f\left(IM_j, M, R, \varepsilon\right)$ is the expectation estimator when IM_j and all high-level parameters are considered. The $RMSE$ is chosen as the loss metric because its units will be the same as the EDP that is being estimated. The well-established linear regression model considering the log-log transformed variables (derived from a power law) can be used as the EDP expectation estimator conditioned only the IM i.e., $\hat{f}(IM)$ (Cornell et al. 2002).

$$\ln \widehat{EDP} = \ln a + \mathrm{b} \ln IM \tag{3.3}$$

where $\widehat{EDP}$ is the median EDP that is predicted by the estimator and a and b are parameters obtained from the regression. To avoid any assumptions about the relationship between the predictors and response variable and to consider the interactive effect across predictors, a machine learning model is used to as the conditional expectation estimator for the case where the IM and all high-level parameters are considered i.e., $\hat{f}(IM, M, R, \varepsilon)$. The random forest ($RF$) (Breiman, 2001) algorithm is used for this purpose. Note that while several other machine learning models could also be used here, the relative performance of different algorithms is less important than the performance of $\hat{f}(IM)$ relative to $\hat{f}(IM, M, R, \varepsilon)$.

RF has been used in several studies within the structural/earthquake engineering community (e.g., Mangalathu and Jeon, 2018; Mangalathu and Jeon, 2019; Xie et al. 2020) where detailed descriptions of the algorithm have been provided. As such, only a brief summary is provided here. RF is an ensemble learning model that aggregates multiple estimators to improve predictions in both classification and regression problems. RF operates by bootstrapping multiple subsets of the same training data set, which are used to construct a pre-specified number of classification and regression trees $(CART)$. The $CART$ algorithm (Breiman et al., 2017) was originally developed for classification purposes but has since been extended to solve regression problems. Given a data subset, a single tree is constructed by recursively dividing the sample space into child nodes based on the observed values of one of the predictors and a chosen split point. The predictor and split point are determined using the greedy algorithm, which maximizes the purity of the generated sub-samples based on the Gini Index (Breiman et al., 2017). This process is repeated until a pre-defined stopping criterion is met. The minimum number of sub-samples in a given node, the maximum tree depth and the maximum number of leaf (terminal) nodes are examples of stopping criteria that are used in the $CART$ development. The uniqueness of RF relative to other $CART$ algorithms stems from the fact that a random subset of the original predictors is considered at each split point to avoid highly correlated trees. For new samples, the weighted average value of the response variable from the individual trees is used as the final prediction.

Chapter 4

Evaluating the Effectiveness of Ground Motion Intensity Measures using Causal Inference

4.1 Overview

At its core, nonlinear response history analysis ($NRHA$) seeks to quantify the causal effect of ground shaking due to earthquakes (as measured by the ground motion intensity measure i.e., the IM) on structural response. As noted earlier, these IMs are computed as a function of a set of upstream parameters that include the earthquake magnitude (M), source-to-site distance (R) and epsilon (ε) where the latter represents the number of standard deviations between the predicted and measured IM. Figure 4.1 shows the directed acyclic graph (DAG) that represents the causal relationship between the IM, engineering demand parameter (EDP) and the upstream seismic parameters (M, R, ε). A DAG is a graph that only contains directed edges and has no cycles (i.e., there is no path from a single node that leads back to itself) (Geiger and Pearl, 1990). These relationships can also be represented using a structural causal model (SCM), which contains a set of endogenous and exogenous variables that are linked by a set of functions that capture the causal relationships between the variables. The SCM corresponding to the DAG in Figure 4.1 is shown Equation 4.1.

$$M = f_M(U_M) \tag{4.1a}$$

$$R = f_R(U_R) \tag{4.1b}$$

$$\varepsilon = f_\varepsilon(U_\varepsilon) \tag{4.1c}$$

$$IM = f_{IM}(M, R, \varepsilon, U_{IM}) \tag{4.1d}$$

$$EDP = f_{EDP}(IM, M, R, \varepsilon, U_{EDP}) \tag{4.1d}$$

where $f_{\cdot}(\cdot)$ is the function representing the relationship between variables and the U's are the unobserved exogenous variables.

Within the $NRHAs$ and from the causal perspective shown in Figure 4.1, the IM can be viewed as the treatment (D), the $EDPs$ represent the outcome (Y) and M, R and ε are the set of confounders $(\mathbf{X})$. Another perspective is that the effects of M, R and ε are partially mediated by the IM. In other words, the IM represents the indirect effect of M, R and ε on the $EDPs$, however, there is also a direct effect i.e., directly from M, R and ε to the $EDPs$ instead of through the IM. In the case where only the IM is considered in $NRHAs$, the one that maximizes the mediated causal effect i.e., the effect of the IM on the EDP.

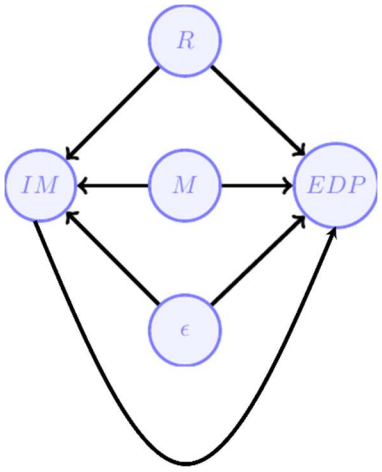

Figure 4.1 – DAG showing the causal relationship between the IM, EDP, M, R and ε

4.2 Using causal machine learning to evaluate IM performance

As described in the previous section, the effectiveness of an IM can be evaluated by quantifying its causal effect on the EDP of interest while considering the earthquake magnitude, source-to-site distance and epsilon as confounders. In other words, the best IM (from causal perspective) is the one that maximizes the effect on the EDP. This study quantifies this effect by integrating partial

linear regression (Robinson, 1988) with machine learning, where the latter is used avoid any assumptions about the way that the confounding variables (M, R and ε) are entered into the regression (Athey et al. 2019; Chernozhukov et al. 2018).

An unbiased quantification of the causal effect (θ) of a treatment, D (IM in this study) on the outcome, Y (here, the $EDPs$) requires correct specification of the control variables, [**X**]. However, since correct specification is never known, assumptions about how **X** is considered in the inference can lead to "model-dependent" or biased results (King and Zeng, 2006). To address this challenge, a semi-parametric or partially linear model (i.e., Y is linear and additive in D) is used where the way that the control variables enter the regression is unspecified (Athey et al. 2019; Chernozhukov et al. 2018; Ratkovic et al. 2021; Robinson, 1998).

$$y_i = \theta d_i + f(\boldsymbol{x}_i) + e_i \tag{4.2}$$

$$d_i = g(\boldsymbol{x}_i) + v_i \tag{4.3}$$

where y_i, d_i and $\boldsymbol{x}_i$ are the outcome, treatment and control (or confounding) variables, respectively, for observation $i \in \{1,2,\dots,n\}$. $f(\boldsymbol{x}_i)$ is a flexible function that represents the effect of the control variables on the outcome. Similarly, $g(\boldsymbol{x}_i)$ represents the relationship between the treatment and controls. e_i and v_i are error terms associated with the estimation of y_i and d_i, respectively, where $\mathbb{E}(e_i|\boldsymbol{x}_i, d_i) = 0$ and $\mathbb{E}(v_i|\boldsymbol{x}_i) = 0$. Since $f(\cdot)$ and $g(\cdot)$ are unknown, they are estimated using machine learning models denoted as $\hat{f}(\cdot)$ and $\hat{g}(\cdot)$, respectively. [**X**] is an $n \times p$ matrix where n is the number of observations and p is the number of control variables or confounders.

The "double machine learning" method developed by Chernozhukov et al. (2018) is used to obtain an estimate of the causal effect $(\hat{\theta})$. The procedure begins by equally splitting the set of observations into a "main" $(\mathcal{S}_1)$ and an "auxiliary" part $(\mathcal{S}_2)$. The latter is used to fit the machine

learning models to obtain $\hat{f}(\cdot)$ and $\hat{g}(\cdot)$. The main part of the dataset, $\mathcal{S}_1$, is used to perform the inference by regressing $y_i - \hat{f}(\boldsymbol{x}_i)$ on $d_i - \hat{g}(\boldsymbol{x}_i)$. The procedure is repeated with the $\mathcal{S}_1$ and $\mathcal{S}_2$ switched (i.e., cross-fitting) and the final causal effect is taken as the average of the two cases. Finally, through random splitting and repeated cross-fitting, an empirical distribution can be obtained for $\hat{\theta}$ where the median is taken as the point estimate and a standard error is also computed.

Chapter 5

Case study: Application to Special Steel Moment Resisting Frame Buildings

The effectiveness of four intensity measures for predicting the peak story drift ratio ($PSDR$) and peak floor acceleration (PFA) in the $SMRF$ structures are evaluated. The considered IMs include peak ground velocity (PGV), PGA, Sa_{T1} and Sa_{avg}. Based on the results of a sensitivity analysis, a period range of $0.2T_1$ to $1.5T_1$ is used to compute Sa_{avg}. The relative loss minimization and causal inference approaches are used in the IM evaluations along with the sufficiency and efficiency criteria.

5.1 Efficiency of Ground Motion Intensity Measures

To evaluate the efficiency of the four IMs, a probabilistic seismic demand model ($PSDM$) is developed using the functional form shown in Equation 3. Nonlinear response history analyses are performed on each of the five $SMRF$ archetypes using the 240 unscaled ground motions. For a given structure, the $PSDR$ and PFA is recorded for each ground motion. Figures 5.1 and 5.2 show log-log scatter plots of each IM versus the $PSDR$ and PFA, respectively, for the 9-story $SMRF$. A clear linear trend is observed in all cases which confirms the validity of the $PSDM$ functional form. However, there are observable differences in the amount of "scatter" for the different IM-EDP pairs. The dispersion (β) in the EDP, which is used as the measure of efficiency, is computed as

$$\beta = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}\left(EDP - \widehat{EDP}\right)^2} \tag{5.1}$$

where EDP and $\widehat{EDP}$ are the recorded (in the $NRHAs$) and estimated (using the $PSDM$) values of the structural response demand, respectively, and N is the number of ground motions.

The β values for the five $SMRF$ archetypes are summarized in Table 5.1. With the exception of the 1-story building, the $PSDR$ dispersion for Sa_{avg} is consistently lower than the others (Table 5.1a). This result is generally consistent with prior studies which showed that Sa_{avg} is more efficient compared to Sa_{T1} both in terms of $PSDR$ estimation and collapse performance assessment (Shokrabadi and Burton. 2017; Eads et al. 2015; Tsantaki et al. 2012). Across all buildings, the $PSDR$ dispersion is highest when PGA is used as the ground motion intensity measure. This result is also consistent with previous studies comparing the efficiency of PGA and Sa_{T1} when estimating the distribution of peak story drift ratios (e.g., Shome and Cornell, 1999). Unlike $PSDR$, the most efficient intensity measure for estimating PFA varies significantly based on the building case (Table 5.1b). For the 1-story and 5-story buildings, the most efficient IMs are Sa_{T1} and Sa_{avg}, respectively. For the remaining buildings PGV has the lowest PFA dispersion. If the average dispersion over all buildings is used as the overall measure of efficiency, Sa_{avg}, which has a mean β of 0.294 for PFA, would be the most efficient.

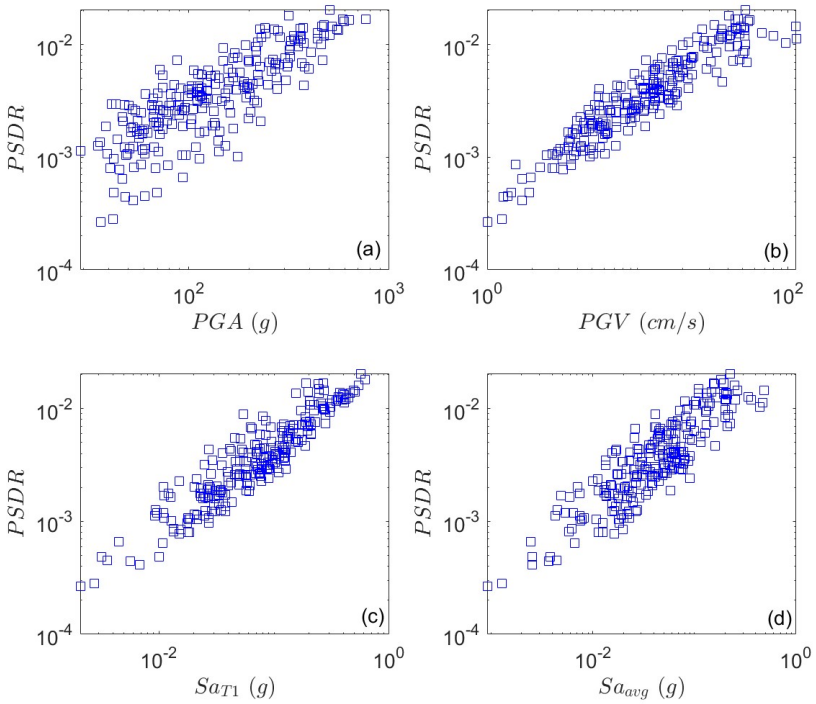

Figure 5.1 – Scatter plots showing $PSDR$ versus (a) PGA, (b) PGV, (c) Sa_{T1} and (d) Sa_{avg} in log-log space for the 9-story $SMRF$

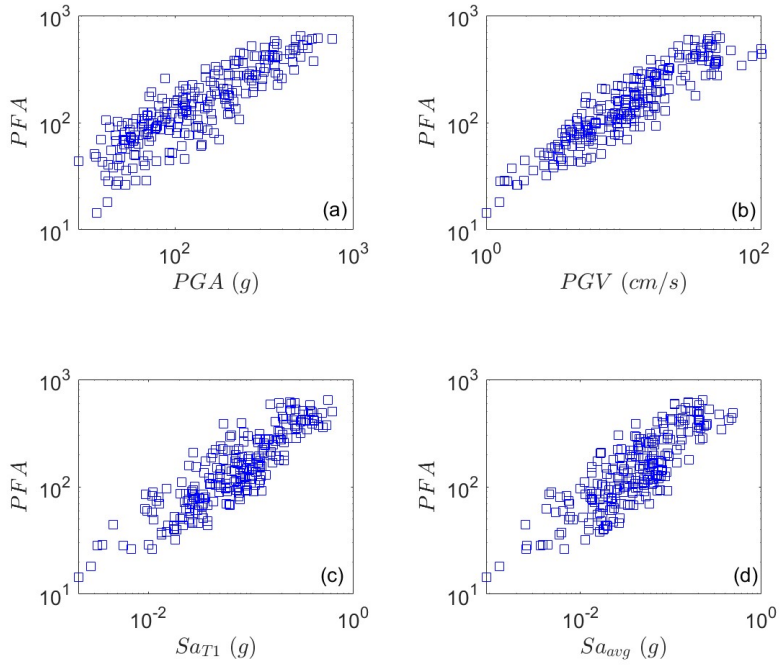

Figure 5.2 – Scatter plots showing PFA versus (a) PGA, (b) PGV, (c) Sa_{T1} and (d) Sa_{avg} in log-log space for the 9-story $SMRF$

Table 5.1 – (a) $PSDR$ and (b) PFA dispersions (β) for all SMF archetypes

(a)

Building	$PSDR$ dispersions (β)			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	0.490	0.453	0.170	0.285
5-story	0.572	0.358	0.281	0.223
9-story	0.512	0.316	0.340	0.279
14-story	0.582	0.336	0.382	0.316
19-story	0.523	0.300	0.438	0.365

(b)

Building	PFA dispersions (β)			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	0.406	0.404	0.188	0.249
5-story	0.374	0.330	0.361	0.297
9-story	0.395	0.320	0.389	0.332
14-story	0.384	0.364	0.471	0.419
19-story	0.396	0.315	0.471	0.410

5.2 Sufficiency of Ground Motion Intensity Measures

This section evaluates the sufficiency of the four IMs with respect to magnitude, distance and epsilon for estimating both peak story drift ratio and peak floor accelerations. Recall that a set of 240 unscaled ground motions are used to perform $NRHAs$ on the five $SMRF$ buildings and the $PSDR$ and PFA is recorded. To facilitate the sufficiency-based evaluation, the ground motion set is placed in six bins and the median value within each bin is used as the relevant intensity level. The number of ground motions in each bin and the corresponding median PGA and PGV values are summarized in Table 5.2. The period/building dependent Sa_{T1} and Sa_{avg} are not included in Table 5.2.

The sufficiency of the four IMs with respect to M, R, and ε is first examined for the 9-

story $SMRF$. For each IM and EDP, Figures 5.3 and 5.4 shows how the *p-value* varies with the ground motion intensity level. The assumed sufficiency threshold of 0.05 is also shown on the plots. Recall that an IM is deemed sufficient with respect to the high-level variable of interest if the *p-value* from the linear regression exceeds this threshold. The first key observation is that for both $PSDR$ (Figure 5.3) and PFA (Figure 5.4) and the three high-level parameters, the sufficiency of some IMs varies based on the intensity level. For example, Sa_{avg} is sufficient with respect to the event magnitude when estimating $PSDR$ for four of the six intensity levels. For the same EDP, Sa_{avg} is sufficient at two-of-six and five-of-six intensity levels for R and ε, respectively. This dependency of the sufficiency of an IM with the considered intensity level has been observed in prior studies (e.g., Shokrabadi and Burton, 2017). Other studies that considered the spatial distribution of $EDPs$ (not considered in the present study) have also found variations in the sufficiency of IMs across different floor and story levels (e.g., Bradley et al. 2010).

Table 5.2 – Ground motion count and median PGA and PGV in each bin

Intensity Level No.	**No. of Records**	**Median PGA (g)**	**Median PGV (cm/s)**
1	10	0.0398	1.50
2	30	0.0527	3.59
3	40	0.0718	5.59
4	40	0.1067	9.26
5	40	0.1533	13.24
6	80	0.3099	31.74

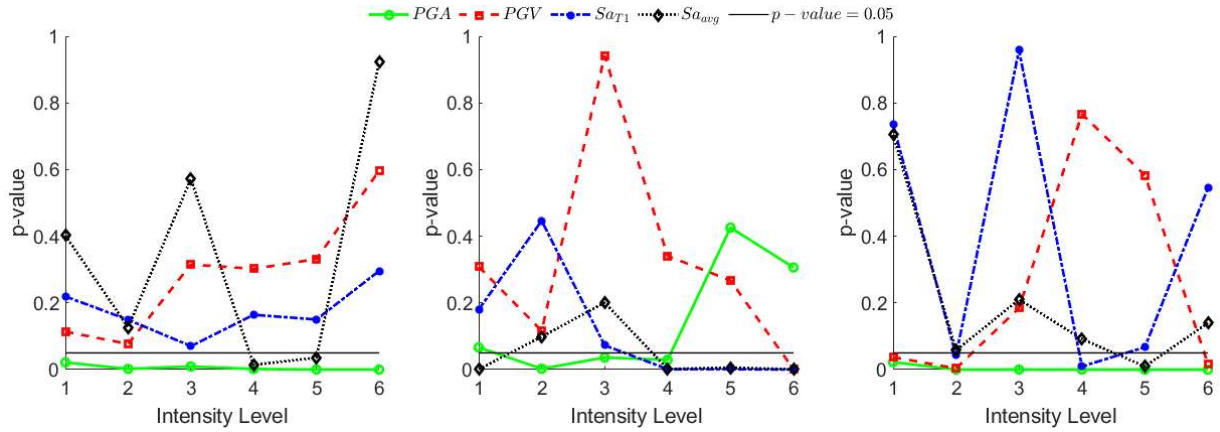

Figure 5.3 – Variation of *p-values* with intensity level for sufficiency of IMs with respect to (a) M, (b) R, and (c) ε when estimating $PSDR$ in 9-story building

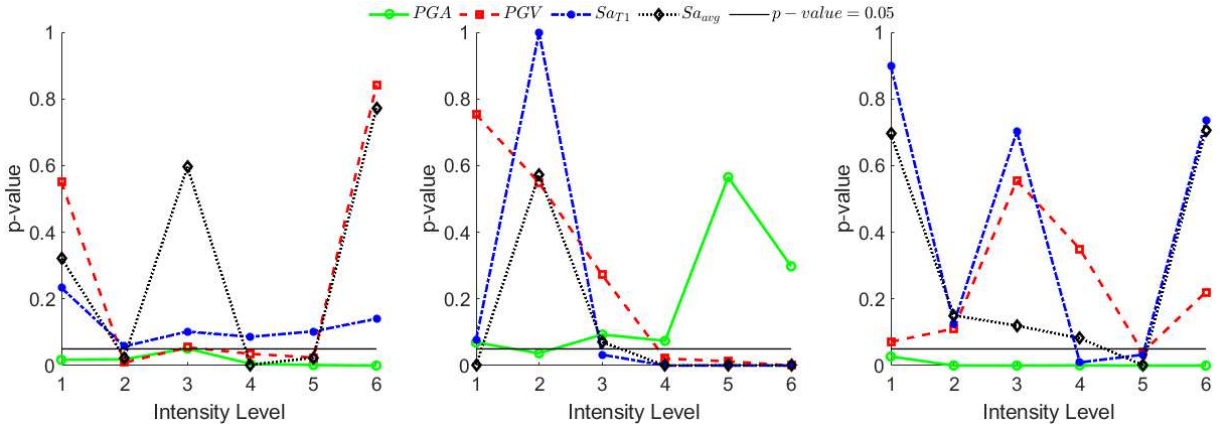

Figure 5.4 – Variation of *p-values* with intensity level for sufficiency of IMs with respect to (a) M, (b) R, and (c) ε when estimating PFA

For a more holistic assessment of the sufficiency for all buildings, Tables 5.3, 5.4 and 5.5 summarize the fraction of intensity levels for which each IM is found to be sufficient. For sufficiency with respect to event magnitude in estimating $PSDR$, PGV and Sa_{avg} have similar overall performance and PGA performs the worse. Also, while PGA shows more overall sufficiency for PFA (compared to $PSDR$), it is still generally outperformed by the other three IMs. For sufficiency with respect to source-to-site distance, PGA outperforms the other IMs when estimating PFA and PGV is the most sufficient for $PSDR$. PGA performs poorly when estimating

both PFA and $PSDR$ when sufficiency with respect to epsilon is considered. Sa_{avg} has the best overall performance in this regard when considering the two $EDPs$. These results highlight the difficulty in choosing "the best" IM based on the sufficiency criteria because of the variation in performance across the different high-level parameters (i.e., M, R and ε), intensity level and type of EDP.

Table 5.3 – Sufficiency of IMs with respect to magnitude for (a) $PSDR$ and (b) PFA

(a)

Building	**Fraction of Intensity Levels where IM is Sufficient (p-value > 0.05)**			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	4/6	6/6	6/6	6/6
5-story	1/6	6/6	6/6	6/6
9-story	0/6	6/6	6/6	4/6
14-story	0/6	5/6	5/6	6/6
19-story	0/6	5/6	1/6	4/6

(b)

Building	**Fraction of Intensity Levels where IM is Sufficient (p-value > 0.05)**			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	4/6	4/6	6/6	6/6
5-story	2/6	6/6	6/6	4/6
9-story	1/6	3/6	6/6	3/6
14-story	3/6	3/6	5/6	3/6
19-story	1/6	5/6	1/6	4/6

Table 5.4 – Sufficiency of IMs with respect to source-to-site distance for (a) $PSDR$ and (b) PFA

(a)

Building	Fraction of Intensity Levels where IM is Sufficient (p-value > 0.05)			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	5/6	3/6	4/6	3/6
5-story	4/6	5/6	2/6	3/6
9-story	3/6	5/6	3/6	2/6
14-story	2/6	4/6	2/6	2/6
19-story	1/6	4/6	3/6	1/6

(b)

Building	Fraction of Intensity Levels where IM is Sufficient (p-value > 0.05)			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	5/6	4/6	5/6	5/6
5-story	4/6	3/6	1/6	2/6
9-story	5/6	3/6	2/6	2/6
14-story	4/6	4/6	1/6	2/6
19-story	3/6	3/6	2/6	1/6

Table 5.5 – Sufficiency of IMs with respect to epsilon for (a) $PSDR$ and (b) PFA

(a)

Building	Fraction of Intensity Levels where IM is Sufficient (p-value > 0.05)			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	1/6	5/6	6/6	5/6
5-story	0/6	4/6	3/6	3/6
9-story	0/6	3/6	4/6	5/6
14-story	0/6	2/6	5/6	6/6
19-story	0/6	4/6	6/6	6/6

(b)

Building	Fraction of Intensity Levels where IM is Sufficient (p-value > 0.05)			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	1/6	5/6	6/6	5/6
5-story	0/6	5/6	4/6	4/6
9-story	0/6	5/6	4/6	5/6
14-story	0/6	3/6	5/6	5/6

19-story	0/6	4/6	6/6	5/6

5.3 Evaluating Ground Motion Intensity Measures using Statistical Loss Minimization Appraoch

This section evaluates the four IMs using the relative loss minimization approach. Recall that two types of estimation models are developed for each EDP. The first type of model, which is denoted as $\hat{f}(IM)$, is based on linear regression of the IM against the relevant EDP in log-log space. The second model $\hat{f}(IM, M, R, \epsilon)$ estimates the EDP using the RF algorithm where the IM is adopted as a feature along with M, R and ϵ. The $RMSE$ of the estimated EDP is then used as a proxy for the predictive performance of $\hat{f}(IM)$ relative to $\hat{f}(IM, M, R, \epsilon)$. The results of the relative loss minimization assessment is summarized in Table 5.6, where it is observed that Sa_{avg} has the lowest $RMSE$ value for $PSDR$ (Table 5.6a) across all buildings. Sa_{T1} has the second-best overall performance for the same EDP with a lower $RMSE$ value for four of the five buildings. These two findings are generally consistent with the results from the efficiency-based assessment as well as prior studies. Also noteworthy is the fact that, across the five buildings, the 1-story case has the lowest relative loss when PGA is used to estimate $PSDR$. In contrast to $PSDR$, PGA is the best performing IM when estimating PFA, producing the lowest $RMSE$ in four of the five buildings. Again, this result is consistent with that of other studies which have shown that PGA performs well when estimating acceleration or force-based $EDPs$. The next best is Sa_{avg} which has the lowest $RMSE$ in one of the five buildings and values comparable to that of PGA in three of the five buildings. Compared to the sufficiency criteria, the relative loss minimization approach provides more conclusive findings that are consistent with prior studies and the efficiency-based assessment.

Table 5.6 – $RMSE$ values corresponding loss of predictive performance of $\hat{f}(IM)$ relative to $\hat{f}(IM, M, R, \varepsilon)$: (a) $PSDR$ and (b) PFA

(a)

Building	Relative Loss ($RMSE$)			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	9.11E-04	1.44E-03	9.95E-04	7.29E-04
5-story	2.06E-03	2.24E-03	1.24E-03	9.94E-04
9-story	1.52E-03	1.73E-03	1.13E-03	9.26E-04
14-story	1.42E-03	1.49E-03	1.05E-03	9.67E-04
19-story	1.21E-03	1.00E-03	1.11E-03	9.57E-04

(b)

Building	Relative Loss ($RMSE$)			
	PGA	PGV	Sa_{T1}	Sa_{avg}
1-story	0.033	0.044	0.044	0.034
5-story	0.039	0.045	0.037	0.032
9-story	0.045	0.061	0.056	0.044
14-story	0.042	0.062	0.066	0.056
19-story	0.041	0.055	0.065	0.056

5.4 Evaluating Ground Motion Intensity Measures using Causal Inferencing

The four IMs are evaluated using the casual inference approach in this section. Recall that the goal is to evaluate the causal effect (θ) of each IM on the $EDPs$ of interest while controlling for the earthquake magnitude, source-to-site distance and epsilon i.e., M, R and ϵ. The double machine learning method is implemented where the Random Forests algorithm is used as the learning model for both the $EDPs$ $\left(y_i - \hat{f}(\boldsymbol{x}_i)\right)$ (outcome) and IMs (treatment) $\left(d_i - \hat{g}(\boldsymbol{x}_i)\right)$. The residuals from those two models are plotted against each other in Figure 5.5 and Figure 5.6. A linear trend is generally observed for both the $PSDR$ (Figure 5.5) and PFA (Figure 5.6) residuals.

However, as expected, there are variations in the amount of scatter across the different IMs. Recall that θ is obtained from regressing the EDP residuals against the IMs residuals. Cross-fitting is applied to a single random sample split because minimal variance in the results from the two samples is obtained.

The results for all five buildings are summarized in Tables 5.7 and 5.8. Focusing first on $PSDR$, Table 5.7a shows that, for the 1-story building, PGA is the best performing IM (from a causal perspective), closely followed by Sa_{avg}. However, for the remaining buildings (i.e., 5-, 9-, 14- and 19-story), Sa_{avg} has the highest θ and is therefore the most effective. This finding is generally consistent with that of the evaluation based on relative loss of statistical information. For the same set of buildings, Sa_{T1} has the second-best performance with PGV having orders of magnitude lower θ values than the other three IMs. The overall results are similar for PFA with PGA having the highest θ value for the 1-story building followed by Sa_{avg} then Sa_{T1}. Whereas, for the remaining buildings, Sa_{avg} has the highest θ value and is therefore the most effective of the IMs. Like $PSDR$, PGV performs much worse than the other three IMs across all five buildings.

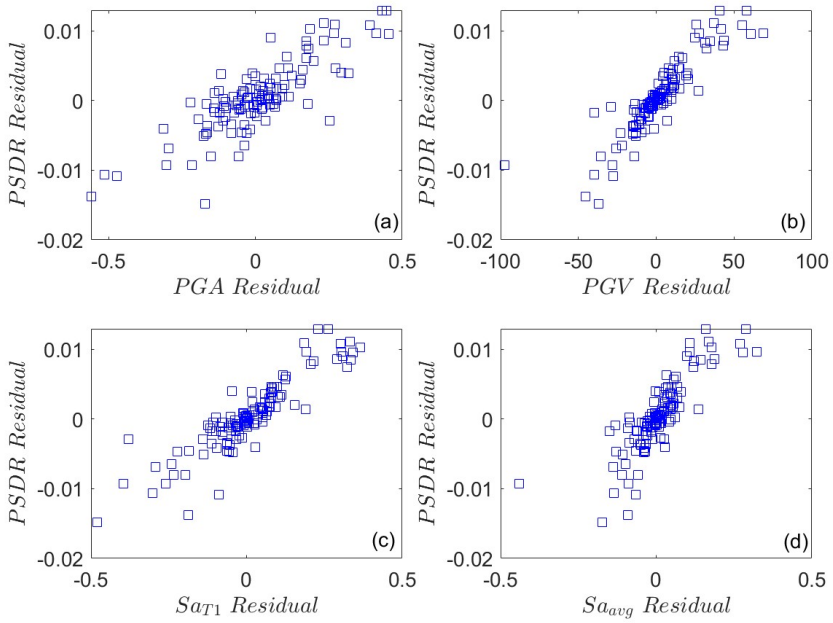

Figure 5.5 – Plot showing $PSDR\left(y_i - \hat{f}(\boldsymbol{x}_i)\right)$ versus $IM\left(d_i - \hat{g}(\boldsymbol{x}_i)\right)$ residuals for the 9-story $SMRF$: (a) PGA, (b) PGV, (c) Sa_{T1} and (d) Sa_{avg}

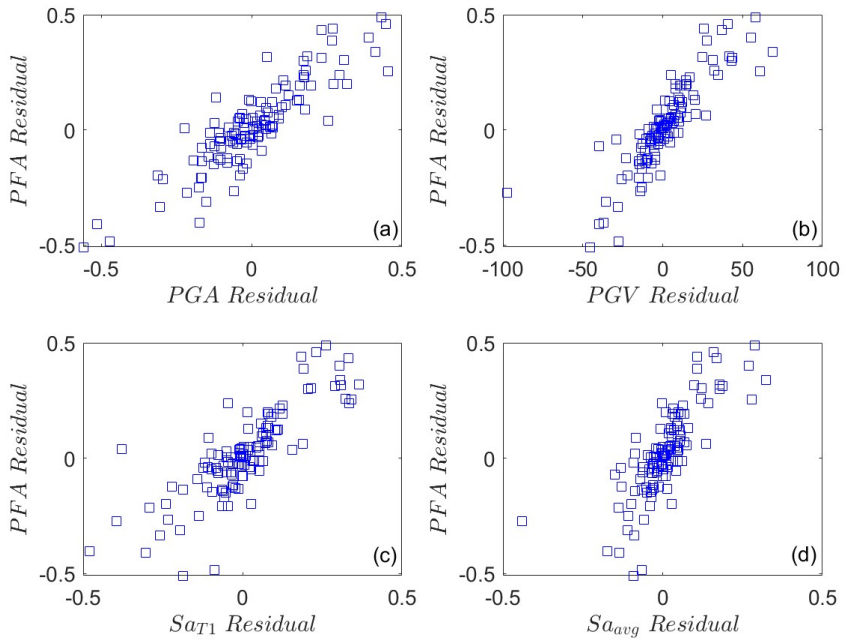

Figure 5.6 – Plot showing $PFA\left(y_i - \hat{f}(\boldsymbol{x}_i)\right)$ versus $IM\left(d_i - \hat{g}(\boldsymbol{x}_i)\right)$ residuals for the 9-story $SMRF$: (a) PGA, (b) PGV, (c) Sa_{T1} and (d) Sa_{avg}

Table 5.7 – Causal effect (θ) values for all ground motion intensity measures and buildings: (a) $PSDR$ and (b) PFA.

(a)

Building	Causal Effect ($\boldsymbol{\theta}$)			
	$\boldsymbol{PGA}$	$\boldsymbol{PGV}$	$\boldsymbol{Sa_{T1}}$	$\boldsymbol{Sa_{avg}}$
1-story	2.11E-2	1.44E-4	1.03E-2	1.58E-2
5-story	3.09E-2	2.41E-4	3.53E-2	4.95E-2
9-story	2.50E-2	1.91E-4	3.16E-2	4.19E-2
14-story	2.11E-2	1.61E-4	3.43E-2	4.06E-2
19-story	1.92E-2	1.53E-4	3.34E-2	4.58E-2

(b)

Building	Causal Effect ($\boldsymbol{\theta}$)			
	$\boldsymbol{PGA}$	$\boldsymbol{PGV}$	$\boldsymbol{Sa_{T1}}$	$\boldsymbol{Sa_{avg}}$
1-story	0.65	0.0046	0.31	0.49
5-story	0.69	0.0050	0.70	1.00
9-story	0.93	0.0068	1.05	1.44
14-story	1.00	0.0069	1.33	1.66
19-story	0.92	0.0066	1.36	1.87

Table 5.6 – Standard error for causal effect (σ_θ) for all ground motion intensity measures and buildings: (a) $PSDR$ and (b) PFA.

(a)

Building	Standard Error ($\boldsymbol{\theta_\sigma}$)			
	$\boldsymbol{PGA}$	$\boldsymbol{PGV}$	$\boldsymbol{Sa_{T1}}$	$\boldsymbol{Sa_{avg}}$
1-story	0.00182	0.00302	0.00131	0.00138
5-story	0.00326	0.00346	0.00254	0.00191
9-story	0.00236	0.00278	0.00235	0.00185
14-story	0.00222	0.00251	0.00261	0.00208
19-story	0.00185	0.00202	0.00259	0.00214

(b)

Building	Standard Error ($\boldsymbol{\theta_\sigma}$)			
	$\boldsymbol{PGA}$	$\boldsymbol{PGV}$	$\boldsymbol{Sa_{T1}}$	$\boldsymbol{Sa_{avg}}$
1-story	0.0676	0.0937	0.0641	0.0585
5-story	0.0627	0.0852	0.0814	0.0640
9-story	0.0773	0.1152	0.1080	0.0880
14-story	0.0749	0.1256	0.1424	0.1203
19-story	0.0741	0.1140	0.1415	0.1229

Chapter 6

Conclusions

Several criteria have been used to evaluate the efficacy of different ground motion intensity measures (IM) in predicting the seismic response of structures, the most common being efficiency and sufficiency. An efficient IM is one that minimizes the dispersion in the predicted engineering demand parameters (EDP) and the sufficiency criterion seeks to limit or eliminate the dependency of the $EDPs$ (conditioned on some IM level) on "higher-level" or "upstream" seismic parameters such as the event magnitude (M), source-to-site distance (R) and epsilon (ε). This study proposes two new approaches to IM evaluation. For the first approach, the most suitable IM (from the perspective of limiting the dependency on upstream parameters) is the one that minimizes the loss

of statistical information when only the IM is used to estimate the EDP distribution $\hat{f}(IM)$, relative to when the IM is used along with other upstream parameters $\hat{f}(IM, M, R, \epsilon)$. More specifically, the best IM is the one that minimizes the root mean square error $(RMSE)$ of the $EDPs$ estimated using $\hat{f}(IM)$ relative to when $\hat{f}(IM, M, R, \epsilon)$ is used. The second approach evaluates the causal effect (θ) of the IM on the EDP of interest while controlling for M, R and ϵ. The IM with the highest θ is deemed the most effective.

For the statistical loss minimization approach, the well-established probabilistic seismic demand model $(PSDM)$ that is based on a linear relationship between the log-transformed IM and EDP is used for $\hat{f}(IM)$. Also, to avoid any assumptions about the type of dependency between the EDP and upstream parameters, a machine learning model is used for $\hat{f}(IM, M, R, \epsilon)$. The machine learning-based approach also captures any joint dependencies or interactions among the upstream variables. The double machine learning strategy is implemented in the causal inference approach where machine learning models are used for estimating the EDP (outcome) and IM (treatment) as a function of the control variables i.e., $\hat{f}(M, R, \varepsilon)$ and $\hat{g}(M, R, \varepsilon)$, respectively. The causal effect is then taken as the coefficient from the linear regression of the EDP residuals against the IM residuals.

The two IM evaluation methodologies are applied to a set of special moment resisting frames $(SMFs)$ using peak story drift ratios $(PSDR)$ and peak floor acceleration (PFA) as the $EDPs$ of interest. The considered IMs include peak ground acceleration (PGA), peak ground velocity (PGV), spectral acceleration at the first mode period (Sa_{T1}) and the spectral acceleration averaged over a range of periods $\left(Sa_{avg}\right)$. The sufficiency and efficiency criteria were also applied and the findings from the four types of assessments were compared. In general, $Sa_{avg}$was found to be the

most efficient IM for estimating both $PSDR$ and PFA, although, more so for the former. The results from the sufficiency-based assessment were much less conclusive. The best IM in terms of sufficiency was found to be highly dependent on the building, ground motion intensity level and the type EDP being estimated. In contrast, the relative loss minimization and causal inference approaches provided clear findings that were generally consistent with the efficiency-based assessment and results from prior studies on multi-criteria evaluation of alternative IMs. Sa_{avg} had the best overall performance (i.e., lowest $RMSE$) for estimating $PSDR$ followed by Sa_{T1}. Although, it is worth noting that PGA performed best in both methods in the 1-story building. For PFA, PGA had the lowest $RMSE$ for four of the five building cases. However, in the causal inference assessment, it only performed best for the 1-story building. In summary, the results from the relative loss minimization and causal inference assessments show that Sa_{avg} is the best IM for estimating displacement-based $EDPs$ (with the possible exception of 1-story buildings) while PGA is more appropriate for acceleration or force-based $EDPs$. This overall finding is consistent with prior studies on assessing the efficacy of alternative IM.

The proposed methodologies were only applied to one structure-type (i.e., $SMFs$) and high-rise buildings were not considered. Also, the IM evaluations were based on $EDPs$ generated using unscaled ground motions. Responses from incremental dynamic analyses ($IDAs$) or multiple stripe analyses ($MSAs$) were not included in the case study. To generalize the findings in terms of the robustness of the relative loss minimization and causal inference approaches, additional studies are needed where IMs are evaluated using other structure types (e.g., other lateral force resisting systems and building heights) and dynamic analysis procedures (e.g., $IDAs$ and $MSAs$).

References

American Institute of Steel Construction (AISC) (2016a). Seismic provisions for structural steel buildings. ANSI/AISC 341-16. American Institute of Steel Construction: Chicago, IL 2016.

American Institute of Steel Construction (AISC) (2016b). Specification for structural steel buildings. ANSI/AISC 360-16. American Institute of Steel Construction: Chicago, IL 2016.

American Institute of Steel Construction (AISC) (2016c). Prequalified connections for special and intermediate steel moment frames for seismic applications. ANSI/AISC 358-16. American Institute of Steel Construction: Chicago, IL 2016.

ASCE 7-16. Minimum design loads and associated criteria for buildings and other structures. Reston, VA: 2016.

Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized Random Forests. *Annals of Statistics 47(2)*, 1148-1178.

Baker, J. W. (2011). Conditional mean spectrum: Tool for ground-motion selection. *Journal of Structural Engineering*, 137(3), 322-331.

Bradley, B. A., Dhakal, R. P., MacRae, G. A., and Cubrinovski, M. (2010). Prediction of spatially distributed seismic demands in specific structures: Ground motion and structural response. *Earthquake engineering & structural dynamics*, 39(5), 501-520.

Breiman, L., 2001. Random forests. Machine learning 45, 5-32.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., (2017). Classification and regression trees. Routledge.

Chernozhukov, V., Chetverikov, D., Demirer, E., Mertand, D., Hansen, C., Newey, W. and Robins, J. (2018). Double/Debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal* .

Cordova, P. P., Deierlein, G. G., Mehanny, S. S. F., and Cornell, C. A. (2000). Development of a two-parameter seismic intensity measure and probabilistic assessment procedure. Proceedings of the Second U.S.-Japan Workshop on Performance-Based Earthquake Engineering Methodology for Reinforced Concrete Building Structures, Sapporo, Hokkaido, Japan, 187-206.

Cornell, C. A., Jalayer, F., Hamburger, R. O., and Foutch, D. A. (2002). Probabilistic basis for 2000 SAC federal emergency management agency steel moment frame guidelines. *Journal of structural engineering*, 128(4), 526-533.

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

FEMA (2018) Assessing Seismic Performance of Buildings with Configuration Irregularities. Applied Technology Council, Redwood City, CA.

Geiger, D., & Pearl, J. (1990). On the logic of causal models. In Machine Intelligence and Pattern Recognition (Vol. 9, pp. 3-14). North-Holland.

Guan M. EERI, X., Burton M. EERI, H., & Shokrabadi, M. (2021). A database of seismic designs, nonlinear models, and seismic responses for steel moment-resisting frame buildings. *Earthquake Spectra*, 37(2), 1199-1222.

Ibarra, L. F., Medina, R. A. and Krawinkler, H. (2005). Hysteretic models that incorporate strength and stiffness deterioration, *Earthquake engineering & structural dynamics* 34, (12),1489-1511.

King, G. and Zeng, L. (2006). The Dangers of Extreme Counterfactuals. Political Analysis 14(2), 131-159.

Krawinkler H (1978) Shear in beam-column joints in seismic design of steel frames. *Engineering Journal* 15, 82–91.

Lignos D (2008) Sidesway collapse of deteriorating structural systems under seismic excitations.

PhD Thesis, Stanford University, Stanford, CA.

Lignos DG and Krawinkler H (2010) Deterioration modeling of steel components in support of collapse prediction of steel moment frames under earthquake loading. *Journal of Structural Engineering* 137(11), 291–1302.

Luco, N., and Cornell, C. A. (2007). Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions." *Earthquake Spectra*, 23(2),357-392.

Mangalathu, S., and Jeon, J. S. (2018). Classification of failure mode and prediction of shear strength for reinforced concrete beam-column joints using machine learning techniques. *Engineering Structures*, 160, 85-94.

Mangalathu, S., and Jeon, J. S. (2019). Machine learning–based failure mode recognition of circular reinforced concrete bridge columns: Comparative study. *Journal of Structural Engineering*, 145(10), 04019104.

Mazzoni, S., McKenna, F., Scott, M. H. and Fenves, G. L. (2006). OpenSees command language manual, Pacific Earthquake Engineering Research (PEER) Center, 264, 137-158.

Mehanny, S. S. F., and Deierlein, G. G. (2000). Modeling and assessment of seismic performance of composite frames with reinforced concrete columns and steel beams. Report No. 135, The John A. Blume Earthquake Engineering Center, Stanford University, Stanford, CA.

Miranda E (1999) Approximate seismic lateral deformation demands in multistory buildings. *Journal of Structural Engineering* 125(4): 417–425.

Moehle, J., and Deierlein, G. G. (2004). A framework methodology for performance-based earthquake engineering. In 13th world conference on earthquake engineering (Vol. 679).

Ratkovic, M. (2020). Relaxing Assumptions, Improving Inference: Utilizing Machine Learning for Valid Causal Inference.

Robinson, Peter. (1988). Root-N Consistent Semiparametric Regression. *Econometrica* 56(4), 931-954.

Shokrabadi, M., & Burton, H. V. (2017). Ground motion intensity measures for rocking building systems. *Earthquake Spectra*, 33(4), 1533-1554.

Shome, N., and Cornell, C. A. (1999). Probabilistic seismic demand analysis of nonlinear structures. Report No. RMS-35, RMS Program, Stanford University, Stanford, CA.

Tothong, P., and Luco, N. (2007). Probabilistic seismic demand analysis using advanced ground motion intensity measures. Earthquake Engineering and Structural Dynamics, 36(13), 1837-1860.

Tsantaki, S., Jäger, C., and Adam, C. (2012). Improved seismic collapse prediction of inelastic

simple systems vulnerable to the P-delta effect based on average spectral acceleration. 15th World Conference on Earthquake Engineering, Lisbon, Portugal.

United States Geological Survey. (2019). Earthquake Hazards Program, Unified Hazards Tool. https://earthquake.usgs.gov/hazards/interactive/

Xie, Y., Ebad Sichani, M., Padgett, J. E., and DesRoches, R. (2020). The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. *Earthquake Spectra*, 36(4), 1769-1801.