Ground Motion Intensity Measures for Post-Earthquake Evaluation using Strong Motion Building Response Data

Eusef Abdelmalek-Lee[1*] and Henry Burton[1]

[1*]Civil and Environmental Engineering, University of California, Los Angeles, 5731 Boelter Hall, Los Angeles, 90095, California, U.S.A.

*Corresponding author(s). E-mail(s): aeusef@g.ucla.edu; Contributing authors: hvburton@ucla.edu;

Abstract

Recent efforts have sought to apply performance-based earthquake engineering (PBEE) principles to post-event decision-support systems that utilize strong motion building response data in near real time. The relative effectiveness of different ground motion intensity measures (IMs) in structural response prediction is a core issue within the PBEE paradigm. This problem has been well-studied in the context of numerical simulations. However, the effectiveness of alternative IMs for reconstructing structural responses in the post-earthquake environment has not been investigated. This study evaluates the performance of several IMs in estimating displacement and acceleration-based responses using building strong motion data. In addition to the efficiency and sufficiency criteria, the IMs are evaluated using causal inference principles. The data set of strong motion structural responses includes recordings from 150 buildings and 35 earthquakes. The study finds that peak ground velocity (PGV) and peak ground acceleration (PGA) outperform the other IMs in terms of their efficiency in estimating peak story drift ratios (PSDR) and peak floor accelerations (PFA), respectively. PGV, cumulative absolute velocity (CAV) and arias intensity (Ia) are sufficient with respect to magnitude (M) and distance (R) for PSDR, whereas for PFA, only the spectral IMs are sufficient with respect to M. The causal analysis finds that when a functional form is not assumed, the IM-EDP relationships are generally linear (in logarithmic space) with reduced effects for some IMs at the tail ends of their distributions. When a linear relationship is assumed, CAV outperforms the other IMs when estimating both PSDR and PFA.

Keywords: ground motion intensity measures, structural health monitoring, causal inference, performance-based earthquake engineering, structural response reconstruction

1 Introduction

Structural health monitoring has been shown to be useful for informing rapid damage assessments of infrastructure following an earthquake. For individual facilities, strong motion response measurements can be incorporated into the decision-making process that determines whether an affected building is safe to occupy. For building portfolios, in addition to informing damage (or functionality) state determination, response measurements can also be used to guide the sequencing and timing of post-earthquake inspections. Numerous studies have been conducted to develop methodologies that use strong motion response measurements to assess post-earthquake structural integrity. While the earliest studies used metrics related to the change in the physical state of the structure [1–4], more recent efforts have sought to leverage the performance-based earthquake engineering (PBEE) framework to assess immediate impacts using decision variables such as monetary losses and/or downtime [5–8].

An important aspect of the PBEE methodology is determining the most suitable ground motion intensity measure (IM) for structural response estimation. In simulation-based PBEE assessments (i.e., without the use of strong motion data), structural responses or engineering demand parameters (EDPs) are estimated using nonlinear response history analyses (NRHAs). In this context, numerous studies have been carried out to appraise the suitability of different IMs for structural response estimation. The basic formulation of such studies includes constructing a nonlinear structural model, performing NRHAs using a set of either site-specific or site-agnostic ground motions, and using the empirical distributions of the generated EDPs to evaluate alternative IMs using different criteria. Efficiency and sufficiency are the most commonly used criteria, however, IMs have also been evaluated based on their robustness to scaling and predictability.

The efficiency criterion evaluates an IM based on its ability to minimize the dispersion in the generated EDPs. The earliest studies focused on efficiency-based IM evaluations date back to the early 1990's where the primary IMs of interest were the peak ground acceleration (PGA) and the spectral acceleration at the first mode period (Sa_{T1}). The latter was found to be more suitable for estimating displacement-based EDPs such as peak story drift ratio [9]. Later studies considered other spectral acceleration-based parameters that were conditioned on two or more periods. Studies by [10] and [11] showed that the performance of Sa_{T1} could be enhanced by including additional modal periods (e.g., 1st and 2nd mode). This idea has since been extended by using the geometric mean spectral acceleration over a range of periods (Sa_{avg}), which has been shown to be very efficient for structures with highly nonlinear response or whose behavior is influenced by multiple modes (e.g., taller buildings) [12, 13]. Peak ground velocity (PGV) has also been shown to be effective (from an efficiency standpoint) for estimating peak story drift responses, especially in tall structures [14] while the filtered incremental velocity has been demonstrated to be suitable for collapse performance estimation [15]. In general, PGA has been shown to be the IM that is most suitable for estimating acceleration-based EDPs (e.g., peak floor acceleration) [13, 14].

The sufficiency criterion evaluates an IM based on the conditional independence assumption that is embedded in the PBEE triple integral. Specifically, this assumption implies that, conditioned on a given IM, the EDPs are independent of the "upstream" parameters used in ground motion models (e.g., earthquake magnitude, source-to-site distance) [16]. There are alternative approaches to assessing the sufficiency of an IM. One uses a two-stage regression, the first step of which develops a linear model that predicts the EDP as a function of the IM. The second step regresses the residuals from the first model against the upstream parameter of interest. The statistical significance of the coefficient from the second regression determines the sufficiency of the IM with respect to the considered upstream parameter. A p-value that exceeds some predefined significance threshold (usually 5%) implies that the upstream parameter is statistically independent of the EDP after conditioning on the IM and is therefore sufficient. An alternative approach to the two-stage regression is to use scaling to condition the ground motions on a predefined IM level. The EDPs generated by the scaled ground motions are then regressed against the upstream parameter and the resulting p-value is interpreted as described earlier. Like efficiency, the sufficiency criterion has been widely used in prior simulation-based IM evaluation studies [9, 12, 13, 15–17]. However, a recent study by [14] showed that the results of the sufficiency-based assessment varies depending on the number of ground motions considered in the analysis. Recently, alternatives to the sufficiency-based assessments have been proposed including the use of information theory [18–20] or methods and principles from the field of causal inference [14]. The latter has recently been shown to produce results that are invariant to the number of considered ground motions.

Because PBEE assessments are typically performed using simulated EDPs, the aforementioned studies utilized responses generated from NRHAs to evaluate the adequacy of IMs. However, as noted earlier, there is a growing emphasis on using strong motion building response data to conduct rapid PBEE-type assessments in the post-earthquake environment. A major challenge in this context is that the response measurements in a given building or portfolio are never "complete". Meaning, within in a given building, only a subset of the floor levels are instrumented. Also, in a given portfolio, only a subset of buildings are instrumented. For this reason, response reconstruction models are used to generate EDPs in the uninstrumented locations (i.e., a subset of the stories in a given building or a subset of the buildings in a given inventory) [1, 21–26]. These response reconstruction models often utilize one or more IMs as predictor variables. Therefore, understanding the relative suitability of different IMs based on different criteria can inform the development of the EDP reconstruction models, especially when they are incorporated into a PBEE-type evaluation.

In this study, we assess the effectiveness of several different IMs using data from building strong motion response measurements. In addition to the sufficiency- and efficiency-based criteria, we evaluate the IMs using the principles and methods from the field of causal inference. The six IMs considered in the evaluation include PGA, PGV, cumulative absolute velocity (CAV), arias intensity (I_a), Sa_{T1} and Sa_{avg}. The response measurements are taken from the database developed by [27], which contains historical data from more than 200 buildings located in California that have been subjected to $\mathbf{M} > 4$ earthquakes during a period of more than 35 years. The recorded floor acceleration histories are used to generate peak story drift ratios (PSDRs) and peak floor accelerations (PFAs), which are the considered EDPs. The acceleration history measured at the ground level is taken as the input motion for each building and used to compute IM values. Several different causal inference methods are considered in the evaluation to determine whether the findings are invariant to the adopted model.

The next section provides an overview of the considered data set including the geometric and structural characteristics of the buildings that make up the inventory and the IM and EDP distributions. A brief but broad overview of the relevant aspects of causal inference is then presented followed by a detailed description of the considered methods. The sufficiency- and efficiency-based assessment criteria are then summarized. The subsequent section presents the results of the IM evaluation using the three approaches (sufficiency, efficiency and causal inference). The paper concludes with a summary of the key findings, limitations of the study, and suggestions for future related work.

2 Summary of Building, Event and Strong Motion Response Data

The building strong motion responses are obtained from the relational database developed in [27]. After removing data points with relevant missing information, 150 buildings with recordings from 35 earthquakes are considered, giving a total of 479 response measurements. The pertinent building properties include the number of stories (above ground), lateral force resisting system (LFRS) type, period estimated by ASCE 7-16 [28] ($T_{1,ASCE}$), and the response measurements, including their locations within the structure. The site properties include the location, source-to-site distance (epicentral), time averaged shear wave velocity to a depth of 30m (V_{S30}), and site class. The event magnitude associated with each response measurement is also considered. Figure 1 shows histograms of the V_{S30}, source-to-site distance, PGA, PGV, and the

LFRS type. For the LFRS type histogram, many of the buildings have multiple systems, so each system is counted individually for this plot. In Figure 2, a histogram of the number of stories is overlayed by plots showing the bin averages of (a) the ASCE period and (b) Sa_{T_1} and Sa_{Avg}. In (c), the coefficient of variation (COV) for each bin is plotted. These plots help to understand the data distribution and serve as a preliminary check. For example, in plot (a), we see that as the number of stories increases, the average period within a bin increases. Conversely, the spectral IMs in (b) decrease with the increase in number of stories (i.e. increase in period). From a physical standpoint, this is expected. The COV plots in (c) help to understand how the number of data points within a bin may effect the results we obtain from the analysis performed later. There are multiple factors considered in the analysis, and given that there are more data points in the low to mid-rise buildings, there is more variation in the building, site, and event characteristics, leading to the higher observed COV. This information will aid in interpreting the results and drawing conclusions.

The ASCE 7-16 estimated period is used to determine Sa_{T_1} and Sa_{avg}. Figure 3 shows the geometric mean spectra (considering ground motions in the two orthogonal directions) for all buildings in the set. Sa_{avg} is calculated as the geometric mean for the range $0.2T_1$ to T_1, as most of the responses are linear-elastic. It is important to note that the lack of moderate to highly nonlinear responses is one of the limitations of the current study. Specifically, the findings in terms of the most suitable IM(s) would only be applicable to problems involving structures that respond in the near-elastic range. Nevertheless, we believe the IM evaluation using measured responses provides new information that has overall relevance to PBEE-type assessments.

The EDPs (maxPSDR and maxPFA) are extracted from the time-series data. The story drift ratio is calculated as the difference in the displacement values in two adjacent stories divided by the story height. This value is calculated at each floor, using the maximum over the entire response history as the PSDR for that floor. The maximum over all floors is then taken as the maxPSDR. The instruments record the accelerations at their respective floors. The maximum at each floor over the entire response history is computed (PFA), and the maximum of these values over all floors is the maxPFA for a given building.

3 Fundamentals of Causal Inference

3.1 Potential Outcomes

This section provides a brief overview of the relevant fundamental aspects of causal inference. In a broad sense, causal inference aims to utilize observational data (i.e., data not generated by a controlled experiment) to identify and/or quantify the causal effect of one variable (described as the "treatment") on another (described as the "outcome" variable) in the presence of "covariate"

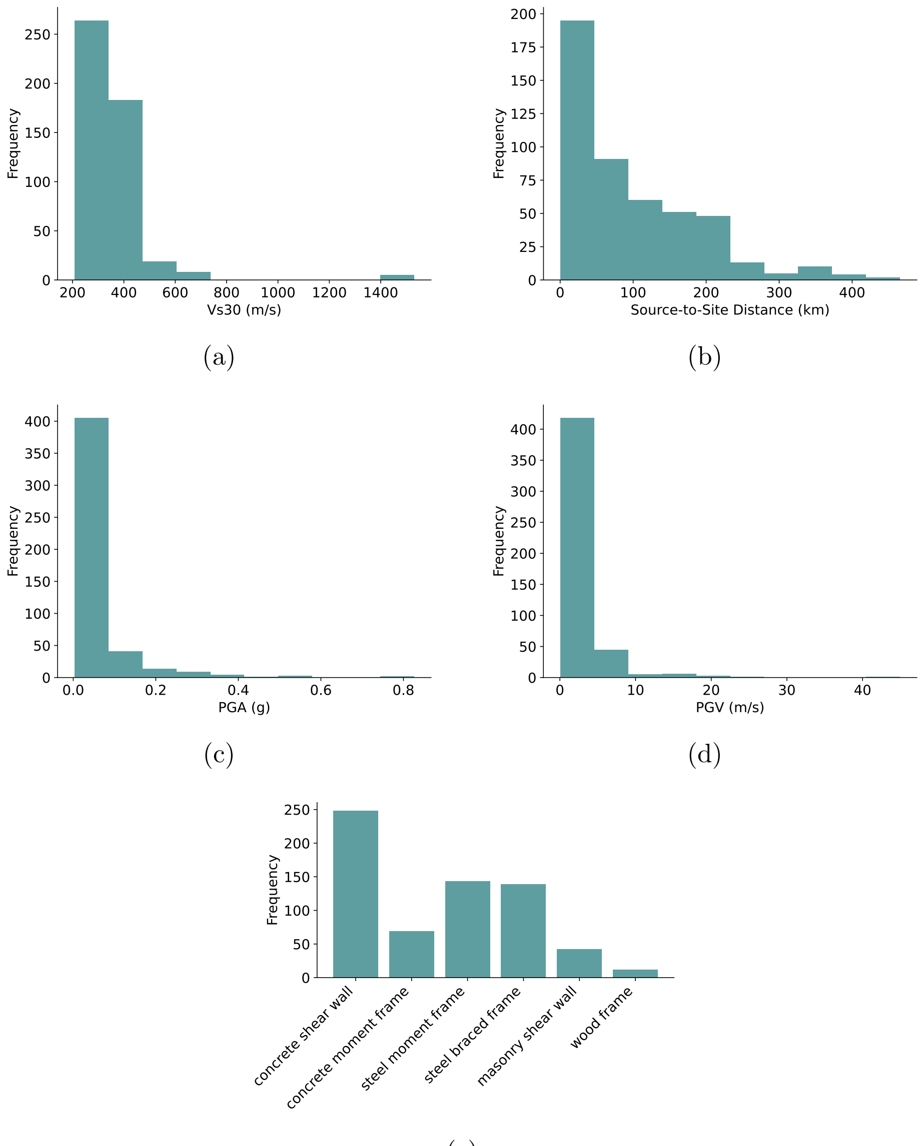

Fig. 1 Histograms showing the empirical distribution of (a) V_{S30}, (b) source-to site distance, (c) PGA, (d) PGV, and (e) the lateral force resisting system types

or "confounding" variables. This can be contrasted with traditional statistics which is centered around establishing associative relationships among a set of variables. Potential outcomes, a cornerstone of causal inference, provide a conceptual mathematical framework for representing causal relationships. Originating from the works of Neyman and Rubin, potential outcomes are

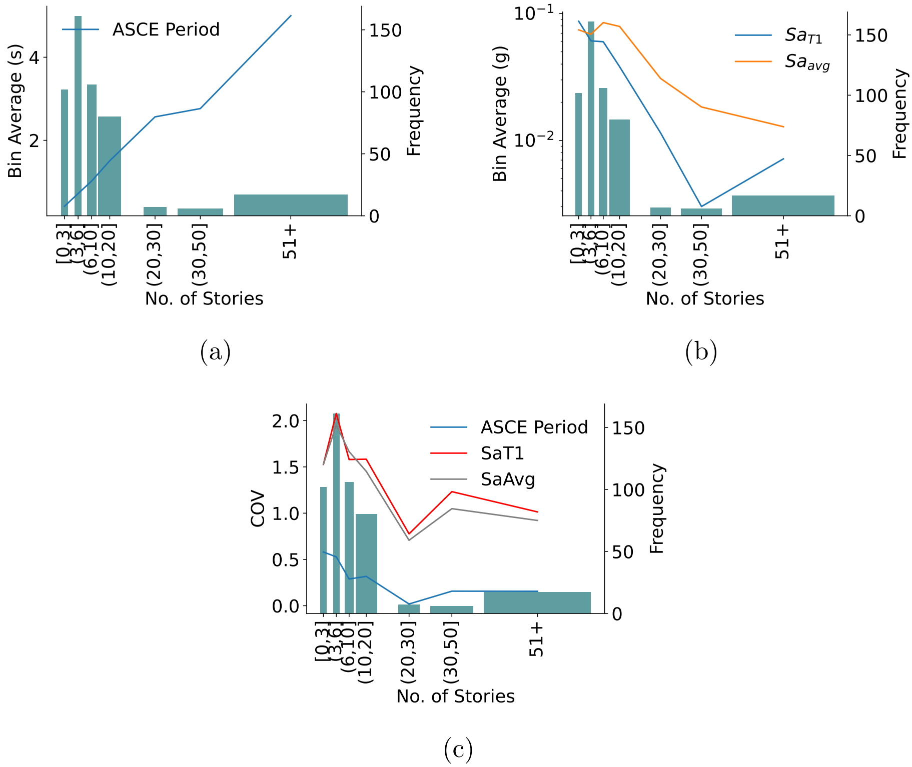

Fig. 2 Bin averages and histograms for (a) the ASCE Period (b) Sa_{T_1} and Sa_{avg} and (c) the corresponding coefficients of variation (COV)

hypothetical outcomes that would have been observed under different treatment conditions [29, 30]. This framework lays the foundation for estimating causal effects and drawing causal conclusions.

In any causal problem, each individual or unit possesses two potential outcomes: the outcome that would be observed if the unit receives the treatment (denoted as Y_1), and the outcome under the condition that the treatment is not assigned to the same unit (denoted as Y_0). However, the fundamental challenge of causal inference stems from the fact that only one of these potential outcomes can be realized in an observational data set, depending on whether a given unit actually receives the treatment or not. The potential outcomes framework allows us to define the causal effect of a treatment as the difference between Y_1 and Y_0 or $Y_1 - Y_0$. This causal effect captures the change in the outcome that can be attributed solely to the treatment itself while holding all other factors constant. To estimate causal effects within this framework, researchers rely on data from various study designs such as observational studies or randomized controlled trials. Regardless of the context, the main challenge lies in addressing the unobserved potential outcomes. Overcoming this challenge requires making identification assumptions that relate

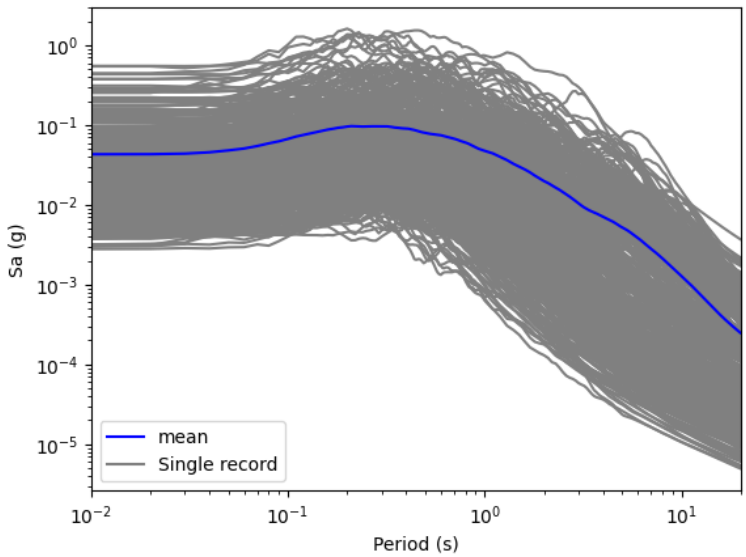

Fig. 3 Geometric mean spectra (considering ground motions in the two orthogonal directions

to the relationship between the potential outcomes and the treatment assignment mechanism. These assumptions play a critical role in estimating causal effects.

One widely recognized identification assumption is the stable unit treatment value assumption (SUTVA) [31]. SUTVA assumes that the potential outcome for a given unit is not influenced by the treatment assignment or potential outcomes of other units. This assumption ensures that the estimated causal effect can be solely attributed to the treatment itself. Another important criterion within the potential outcomes framework is related to the assignment mechanism or the process by which units are assigned to different treatment conditions. In randomized controlled trials, the assignment mechanism is typically random, ensuring that treatment assignment is independent of unit characteristics or covariate variables. In observational studies, however, the assignment mechanism may be affected by confounding factors, which can introduce bias into the estimated causal effects. This leads to the ignorability assumption, which says that the treatment assignment should be independent of potential confounding factors (variables affecting both the treatment and the outcome). Lastly, the consistency assumption says that the potential outcome of unit i corresponding to the treatment condition for the same unit is the outcome that is observed. So, in other words, if the treatment condition for unit i, $T_i = t$, then outcome, $Y_i^t = Y_i^T = Y_i$ [32].

The potential outcome framework provides a principled approach to understanding causal relationships. By considering the hypothetical potential outcomes and making appropriate identification assumptions, effect estimates can be quantified and more defensible conclusions can be drawn about the causal relationships between variables.

3.2 Directed Acyclic Graphs (DAGs) and Structural Causal Models (SCMs)

Causal graphical models have been proposed as a compliment to the potential outcomes framework [32]. Causal graphical models provide a visual representation of causal relationships and facilitate the identification of causal effects based on structural causal models. Directed Acyclic Graphs (DAGs) and Structural Causal Models (SCMs) are used to represent and understand causal relationships between variables. They provide a formal framework for modeling causal structures and making causal claims based on observed data.

In a DAG, variables are represented as nodes, and causal relationships are depicted by directed edges connecting the nodes. Importantly, DAGs impose a structure where causal effects flow from parent nodes to child nodes, creating a clear directionality of influence. The acyclic property ensures that there are no feedback loops or circular dependencies among variables. DAGs are useful for encoding prior knowledge about the causal relationships between variables. They help define the assumptions underlying the causal models and identify the confounding variables that need to be addressed when estimating causal effects. Figure 4 shows an example of a DAG for a hypothetical causal problem with outcome Y, treatment T, a set of covariates [X], and the corresponding unmeasured influences, U.

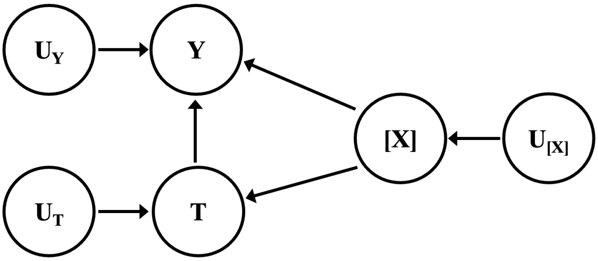

Fig. 4 DAG for a hypothetical causal problem with treatment T, outcome Y and covariates [X]

SCMs are mathematical representations of the causal relationships described by a DAG. SCMs specify the interactions among variables in the system and how their values are determined. Each variable is associated with a structural equation that describes its dependence on its parent variables. The structural equations encode the causal mechanisms or processes that generate the values of the variables. In an SCM, there are endogenous variables and exogenous variables. Endogenous variables are those whose values are determined by other variables in the system, while exogenous variables are treated as inputs or sources of external influence, which are not affected by other variables in the system. The structural equations in an SCM provide a mathematical representation of the causal relationships implied by the DAG. By specifying the functional forms and parameters of these equations, causal effects can be estimated using statistical methods under different scenarios. The SCM associated with the DAG in Figure 4 is described in Equation 1.

$$X_i = f_{X_i}(U_{X_i}) \tag{1a}$$
$$T = f_T(X_1, ..., X_n, U_T) \tag{1b}$$
$$Y = f_Y(X_1, ..., X_n, T, U_Y) \tag{1c}$$

In summary, the combination of DAGs and SCMs enables causal conclusions to be drawn from observational data. DAGs provide a clear graphical representation of the causal assumptions, while SCMs formalize these assumptions into mathematical models. These models can then be used to estimate causal effects, perform counterfactual analyses, and assess the sensitivity of results to different assumptions.

4 Methods used to Evaluate the Efficacy of Ground Motion Intensity Measures

In addition to the well-established efficiency and sufficiency criteria, three different causal inference methods are used to evaluate the relative performance of the considered IMs. The subsequent subsections provide an overview of the causal inference approach to the IM evaluation problem as well a detailed description of the three methods. A brief summary of the efficiency and sufficiency-based assessment is also presented.

4.1 Causal Inference-Based Intensity Measure Evaluation

4.1.1 Overview

According to the sufficiency criterion, an effective IM is one that produces EDPs that are independent of the upstream parameters used in ground motions models (e.g., event magnitude, source-to-site distance) after conditioning on said IM. As discussed in the Introduction section, the sufficiency-based assessment of this conditional independence criterion can be evaluated using one of two approaches: (i) regressing the conditional EDP (i.e., EDPs associated with a single IM level) against the upstream parameter of interest or (ii) a two-stage process that first regresses the EDP against the IM and then regresses the residuals from the first model against the upstream parameter of interest. In either approach, if the upstream parameter of interest is found to be statistically insignificant (based on the associated p-value), the IM is deemed sufficient.

As described in [14], the conditional independence criterion can also be evaluated through the lens of causal inference. The DAG shown in Figure 5 provides a conceptual representation of the IM evaluation problem from a causal inference perspective. In this formulation, the EDP is the outcome variable and the IM is the treatment. The considered upstream parameters include the event magnitude (M), epicentral source-to-site distance (R), and site class (as defined by ASCE 7-16). Note that site class (SC in Figure 5) is used (as a categorical variable) in lieu of Vs_{30} because of the large uncertainty associated with the latter. Also, many of the Vs_{30} values provided by the Center for Strong Motion Data [33] are not directly measured but inferred from topological features. In the Figure 5 DAG, M, R and the site class (denoted by the vector $X_1 = \{M, R, SC\}$) are confounders because they affect both the IM and the EDP. It then follows that the causal effect of a given upstream parameter (e.g., M) on the EDP can be interpreted as having two parts. The direct effect represented in Figure 5 by the arrow from the upstream parameter to the EDP and the "mediated" effect which flows through the IM. From a conditional independence standpoint, a desirable IM is one that maximizes the "mediated" effect and minimizes the direct effect. However, since the direct effect is independent of the IM, the IM that maximizes the effect on the EDP while controlling for the upstream parameters is the most desirable (from a causal inference perspective).

In a typical NRHA-based IM evaluation study (like the prior studies described in the Introduction section), each structure is considered individually and the uncertainty is derived from using multiple ground motions. In the current study, the set of measured EDPs are from a portfolio of buildings. In other words, the IMs are being evaluated across multiple structure types. As such, the number of stories (N_s), LFRS-type and ASCE 7-16 estimated period (T_1) (denoted by the vector $X_2 = \{N_s, LFRS, T_1\}$) are also confounders because they affect both the IM and the EDP. Therefore, the causal inference-based IM evaluation considers both the upstream parameters and building/structural properties shown in Figure 5 as confounders (denoted by the vector $X = \{X_1, X_2\}$). Note that T_1 only affects the Sa_{T1} and Sa_{avg} IMs, which is why the arrow between the two variables is shown using a dashed line.

The IM evaluation problem shown in Figure 5 is solved using the following three causal inference methods: Inverse Propensity Weighting (IPW), Covariate Balancing Propensity Score (CBPS) and Entropy Balancing for Continuous Treatments (EBCT). Recall that for any causal problem, a primary challenge with estimating the effect of the treatment on the outcome variable is the presence of one or more confounders. One approach to addressing this challenge (illustrated in Figure 6) is to "balance" the effect of the confounding variables (or covariates) before estimating the treatment effects using statistical regression. The differences among the IPW, CBPS and EBCT methods lie primarily in the covariate balancing strategy that is used before estimating the treatment effect. Considering all three methods will allow us to assess the invariance of the causal inference-based IM evaluation results to the covariate balancing strategy. Ideally, the relative performance of the considered IMs in estimating the EDPs should should not vary across the different covariate balancing approaches. As shown in Figure 6, two types of regression are considered after the covariate balancing is performed. Using simple linear regression provides a single average treatment effect (ATE) value. However, to consider any nonlinearity in the post-balancing relationship between the treatment and outcome, local linear regression is also performed to obtain an exposure outcome relationship.

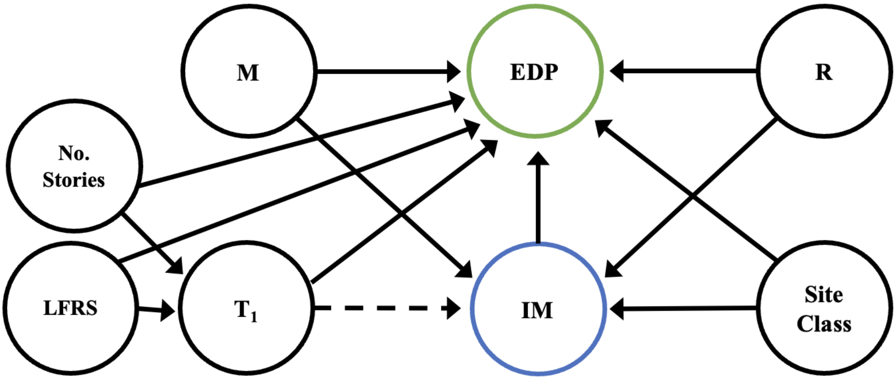

Fig. 5 DAG showing a conceptual representation of the IM evaluation problem. The unmeasured influences are not shown.

4.1.2 Inverse Propensity Weighting (IPW)

To account for non-random assignment of the treatment in observational studies, inverse propensity weighting (IPW) [34] is a widely used technique in causal inference. It assigns weights to each observation based on their propensity scores, which represent the conditional probability of receiving the treatment given the observed covariates.

Consider a binary treatment setting with a dataset comprised of N observations denoted by (Y_i, T_i, X_i) for $i = 1, 2, ..., N$. Here, Y_i is the outcome of interest, T_i is the binary treatment indicator (0 for control, 1 for treatment), and X_i represents the vector of observed covariates for the observation i. The propensity score, denoted as $\pi(X_i)$, is the conditional probability of receiving the treatment given the covariates X_i, i.e., $\pi(X_i) = P(T_i = 1|X_i)$. The goal is to estimate the average treatment effect (ATE), which is the average difference in outcomes between the treated and control groups. To balance the covariates across the treatment groups, we assign weights to each observation based on the inverse of the propensity scores. The weight for the i^{th} observation is given by $w(X_i) = 1/\pi(X_i)$ for the treated group $(T_i = 1)$ and $w(X_i) = 1/(1-\pi(X_i))$ for the control group $(T_i = 0)$. The $w(X_i)$ operates by upweighting the underrepresented group and downweighting the overrepresented group, effectively balancing the covariate distributions between the treatment groups. The IPW estimate of the ATE is then calculated as:

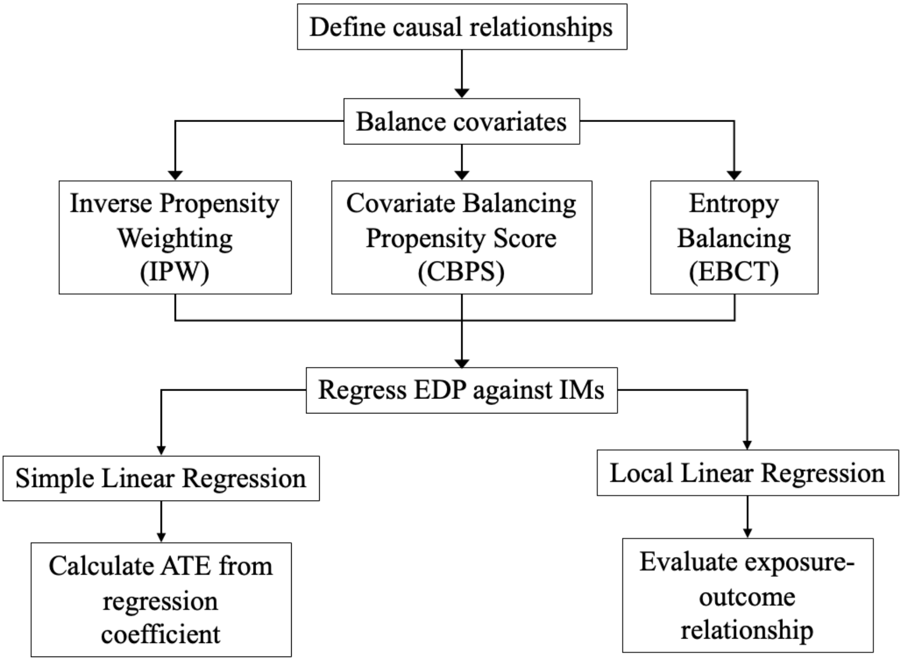

Fig. 6 Schematic illustration of the covariate balancing approach used to solve the IM evaluation problem using causal inference

$$ATE_{IPW} = (1/N)\sum[T_iY_iw(X_i)] - (1/N)\sum[(1-T_i)Y_iw(X_i)] \tag{2}$$

where the first term on the right-hand side represents the weighted sum of outcomes in the treated group, and the second term represents the weighted sum of outcomes in the control group. By incorporating the weights based on the inverse propensity scores, the IPW estimator effectively creates a pseudo-population where the treatment assignment is independent of the observed covariates. This allows for unbiased estimation of the treatment effect by removing the confounding bias induced by the non-random treatment assignment.

For continuous treatment variables (as is the case in the IM evaluation problem), the IPW methodology can be extended by replacing the binary treatment indicator with the continuous treatment values. The propensity scores are then estimated using the appropriate models, such as logistic regression for binary treatments or generalized propensity score models for continuous treatments. The weights are calculated as the inverse of the estimated propensity scores, similar to the binary treatment case. The IPW estimator for the average treatment effect with continuous treatments follows a similar formulation as the binary treatment case:

$$ATE_{IPW} = (1/N) \sum (T_i - \mu(T)) w_i Y_i \tag{3}$$

where T_i now represents the continuous treatment values, and $w(X_i)$ is the corresponding weight based on the inverse propensity score. μ_T is the conditional mean of the treatment given the covariates.

Estimating the propensity scores and choosing the appropriate model for estimating treatment effects are the two main steps in the IPW methodology. Various techniques, such as logistic regression, machine learning algorithms, or propensity score matching, can be employed to estimate the propensity scores.

IPW is a useful technique in causal inference that addresses the issue of non-random treatment assignment in observational studies. By assigning weights based on the inverse of propensity scores, IPW balances the covariate distributions between treatment groups and allows for unbiased estimation of treatment effects. The methodology can be applied to both binary and continuous treatment settings, with appropriate modifications in the calculation of weights and estimation procedures.

4.1.3 Covariate Balancing Propensity Score (CBPS)

Covariate Balancing Propensity Score (CBPS), originally developed by [35], is a method used in causal inference to estimate treatment effects while achieving balance in the covariates between the treatment and control groups. The method was extended by [36] to consider continuous treatments. It combines the principles of propensity score estimation and weighting techniques to address confounding bias and enhance the validity of causal analyses. Similar to IPW, the CBPS method begins by estimating the propensity scores, $\pi_\beta(X_i) = P(T_i = 1|X_i)$ using logistic regression:

$$\pi_\beta(X_i) = \frac{exp(X_i^T \beta)}{1 + exp(X_i^T \beta)} \tag{4}$$

where β is an unknown parameter vector of size equal to the number of covariates, which can be estimated by maximizing the log-likelihood in Equation 5. Once the propensity scores are estimated, the CBPS method assigns weights to each unit based on their propensity score and covariate values. The definition of weights based on the covariates is where CBPS differs from IPW. In the CBPS framework, the weight assigned to unit i, W_i, is given by Equation 6.

$$\hat{\beta}_{MLE} = \arg\max_\beta \sum_{i=1}^{N} T_i log(\pi_\beta(X_i)) + (1 - T_i) log(\pi_\beta(X_i)) \tag{5}$$

$$W_\beta(T_i, X_i) = \frac{T_i - \pi_\beta(X_i)}{\pi_\beta(X_i)(1 - \pi_\beta(X_i))} \tag{6}$$

This ensures that units with covariate values that are unbalanced between the treatment and control groups receive higher weights, facilitating covariate balance. When extended to continuous treatments, transformation of the treatment and covariate vectors is performed so that the means and variances are 0 and 1, respectively. The transformed vectors are denoted as $\tilde{T}_i$ and $\tilde{X}_i$. The transformation of the treatment variable is shown in Equation 7. The same equation is used to transform the covariate vector.

$$\tilde{T}_i = \frac{T_i - \bar{T}}{\sqrt{s_T}} \tag{7a}$$

$$\bar{T} = \sum_{i=1}^{N} \frac{T_i}{N} \tag{7b}$$

$$s_T = \sum_{i=1}^{N} \frac{(T_i - \bar{T}^2)}{N-1} \tag{7c}$$

The weights for the continuous treatment case are defined by:

$$w_i = \frac{f(\tilde{T}_i)}{f(\tilde{T}_i|\tilde{X}_i)} \tag{8}$$

To satisfy the covariate balancing conditions that the weights (i) remove the correlation between $\tilde{X}_i$ and $\tilde{T}_i$ and (ii) maintain their marginal means, the following constraints are applied.

$$\sum_{i=1}^{N} w_i g(\tilde{X}_i, \tilde{T}_i) = 0 \tag{9a}$$

$$\sum_{i=1}^{N} w_i - N = 0 \tag{9b}$$

where $g(\tilde{X}_i, \tilde{T}_i) = (\tilde{X}_i, \tilde{T}_i, \tilde{X}_i\tilde{T}_i)^\top$. Equation 8 can be rewritten in terms of w_i as $f(\tilde{T}_i, \tilde{X}_i) = \frac{1}{w_i} f(\tilde{T}_i) f(\tilde{X}_i)$. To maximize the likelihood of the full sample, we maximize the following function with the corresponding contraints:

$$\prod_{i=1}^{N} f(\tilde{T}_i, \tilde{X}_i) = \prod_{i=1}^{N} \frac{1}{w_i} f(\tilde{T}_i) f(\tilde{X}_i) \tag{10}$$

$$\text{s.t.} \sum_{i=1}^{N} w_i g(\tilde{X}_i, \tilde{T}_i) = 0$$

$$\sum_{i=1}^{N} w_i = N$$

$$\sum_{i=1}^{N} w_i \tilde{X}_i \tilde{T}_i = 0$$

$$w_i > 0 \text{ for all i}$$

The likelihood is equivalent to $\sum_{i=1}^{N} log(f(\tilde{T}_i)) + log(f(\tilde{X}_i)) - log(w_i)$. Since we are only interested in w_i, this equation simplifies to

$$\arg\min \sum_{i=1}^{N} log(w_i) \tag{11}$$

Using the Lagrange method to evaluate the optimization problem and solve for w_i, we obtain

$$w_i = \frac{1}{1 - \gamma^\top g(\tilde{X}_i, \tilde{T}_i)} \tag{12}$$

where γ is the Lagrange multiplier associated with the first constraint. With the assigned weights, the CBPS method computes the weighted average treatment effect. The w_i's are incorporated in the estimation of the ATE by giving more weight to units whose covariate values are unbalanced between the treatment and control groups.

By achieving covariate balance, the CBPS method aims to reduce confounding bias and provide more reliable estimates of the treatment effect. It allows for the adjustment of multiple covariates simultaneously and can handle both binary and continuous treatments.

CBPS combines propensity score estimation and weighting techniques to achieve covariate balance. By assigning weights based on the estimated propensity scores and covariate values, CBPS aims to reduce confounding bias and provide more accurate treatment effect estimates.

4.1.4 Entropy Balancing for Continuous Treatments (EBCT)

Entropy Balancing for Continuous Treatments, or EBCT, was developed by [37] as a continuous treatment extension to the original entropy balancing method for binary treatments [38]. The primary function of EBCT is to minimize the correlation between the moments of the pre-treatment covariates and the treatment by weighting each unit. The weighted set can then be used to estimate the ATE with minimal influence from the confounders. EBCT works by first partitioning the observations into bins based on the treatment variable. Then, within each bin, the covariate distributions are balanced using a weighting scheme that minimizes the dispersion of the weights (the entropy). This ensures that the weights are evenly distributed across the treatment and control groups. Once these weights are obtained, they can be used in techniques such as weighted linear regression to estimate the ATE.

The binary entropy balancing case will be first introduced then expanded to deal with continuous treatments. The binary treatment variable is denoted as T such that each data point is assigned to either the treatment (1) or control (0) group. For a set of N observations ($N = N_0 + N_1$), each unit, i, has a treatment status T_i and covariates X_i. Entropy balancing seeks to find the weight, w_i, for each unit that balances the covariate distribution in the treatment and control groups. The formulation uses the Kullback entropy metric ($h(w)$ in Equation 13).

$$h(w_i) = w_i \ln\left(\frac{w_i}{\pi_i}\right) \tag{13}$$

where π_i, the selected base weight, is set equal to $1/N_1$ for all i when not specified. The goal is to minimize this entropy subject to certain constraints via the loss function $H(w)$ in Equation 14.

$$\min_{w} H(w) = \sum_{i|T_i=0} h(w_i) \tag{14}$$

$$\begin{aligned} \text{s.t.} \sum_{i|T_i=0} w_i c_{ri}(X_i) &= m_r,\ r \in 1, ..., R \\ \sum_{i|T_i=0} w_i &= 1 \\ w_i \geq 0 \text{ for all } i &\text{ such that } T_i = 0 \end{aligned}$$

where $c_{ri}(X_i)$ is the covariate moment functions of the control group, which can be defined as $c_{ri}(X_i) = X_{ij}^r$ or a de-meaned version of the moment functions given by $c_{ri}(X_{ij}) = (X_{ij} - \mu_j)^r$, where μ_j is the mean and j is a corresponding covariate for observation i. m_r represents the r^{th} moment of the treatment group for covariate X_j. For example, if r = 1, the constraint is such that only the means of the covariates are balanced. If r = 2, then the means and the variances are balanced, and so on. In summary, the constraints impose that the sum of the weighted covariate moments in the control group must equal the moments of the covariates in the treatment group.

The extension of EBCT to consider continuous treatment variables is based on [37]. The treatment T is now a continuous variable, with $T_i > 0$ for treated units. The covariate moment function for the binary case, $c_{ri}(X_i)$, is rewritten as $\tilde{X}_i = f(X_i) - \frac{1}{N_1}\sum_{i|T_i>0} f(X_i)$ where $f(\cdot)$ is a function of the covariate and any interaction terms to be balanced, and the de-meaning is done with respect to the treatment group. The variables p and $\tilde{T}$ are used to denote the order of the treatment term to be balanced and a de-meaned representation of the treatment (similar to $\tilde{X}$ for the covariates), respectively. The column vector $g(p, \tilde{X}_i, \tilde{T}_i) = [\tilde{X}_i{}', \tilde{T}_i, ..., \tilde{T}_i{}^p, \tilde{X}_i{}' \cdot \tilde{T}_i, ..., \tilde{X}_i{}' \cdot \tilde{T}_i{}^p]$ considers all necessary covariate and treatment moments and interactions. Equation 14 can then be re-written for continuous treatments as

$$\min_{w} H(w) = \sum_{i|T_i>0} h(w_i) \tag{15}$$

$$\text{s.t.} \sum_{i|T_i>0} w_i g(p, \tilde{X}_i, \tilde{T}_i) = 0$$
$$\sum_{i|T_i>0} w_i = 1$$
$$w_i > 0 \text{ for all } i \text{ such that } T_i > 0$$

To obtain the weights, the constrained optimization problem can be reformulated by applying Lagrange multipliers (λ, γ) to the constraints. Recall from Equation 13 that $h(w) = w_i \ln\left(\frac{w_i}{\pi_i}\right)$, so the new unconstrained problem becomes

$$\min_{w,\lambda,\gamma} \mathscr{L}, (w, \lambda, \gamma) = w_i \ln\left(\frac{w_i}{\pi_i}\right) - \lambda\Big\{\sum_{i=1}^{N_1} w_i - 1\Big\} - \gamma'\Big\{\sum_{i=1}^{N_1} w_i g(p,,, \tilde{X}_i,,, \tilde{T}_i)\Big\} \tag{16}$$

Differentiating with respect to λ and w_i produces the equation for calculating the balancing weights

$$w_i = \frac{q_i \exp\big\{\gamma' g(p,,, \tilde{X}_i,,, \tilde{T}_i))\big\}}{\sum_{i=1}^{N_1} q_i \exp\big\{\gamma' g(p,,, \tilde{X}_i,,, \tilde{T}_i)\big\}} \tag{17}$$

4.2 Efficiency and Sufficiency

In general (statistical) terms, efficiency and sufficiency are two approaches to quantifying the quality of an estimator. The efficiency criterion gauges how effectively the available data is utilized to estimate the outcome of interest. Sufficiency assesses whether the estimator alone is sufficient with respect to one or more covariates in explaining the response. Sufficiency is evaluated by conditioning on the estimator and evaluating the dependence of the response on the covariate(s) of interest. In the context of the current study, the IM is the estimator and the upstream parameters are the covariates.

The primary objective in efficiency-based evaluation is to identify the IM that offers the most precise estimates of the EDPs given certain assumptions or conditions. Specifically, an estimator is deemed efficient when it achieves the smallest variance among the class of estimators being compared. In other words, it strives to minimize the dispersion between estimated values and the true value of the parameter. Consider the following regression:

$$ln\,\hat{Y} = ln\,a + b \cdot ln\,X + \epsilon \tag{18}$$

where $\hat{Y}$ is the response variable of interest, X is a predictor variable, a and b are regression parameters and ϵ is the error. An efficient estimator is one that minimizes the variance in the error term ϵ.

The sufficiency criterion is a measure of how well the estimator alone predicts the outcome. The two stage regression approach is adopted in the current study. The first regression is between the intensity measure and the response, the same as for efficiency in Equation 18. The second regression takes the residuals from the first stage and regresses them against each covariate, shown in Equation 19. The p-values obtained from this second regression indicate the sufficiency of the IM. If the p-value is greater than 0.05 (a commonly used threshold for statistical significance), then the covariate is rendered independent of the first stage regression residuals conditioned on the treatment, deeming the IM sufficient.

$$\epsilon = c \cdot Z + \eta \tag{19}$$

where Z is the covariate or upstream parameter, c is a regression parameter and η is a second error term. The baseline unbalanced data set is evaluated for both efficiency and sufficiency for comparison with the causal inference approaches that use covariate balancing.

5 Ground Motion Intensity Measure Evaluation and Results

This section evaluates the relative effectiveness of five ground motion intensity measures in estimating two types of engineering demand parameters. The considered IMs include PGA, PGV, CAV, I_a, Sa_{T1} and Sa_{avg}. Each sensor in a building that is subjected to a given event provides response history data at a specific location (i.e., floor level). The response histories are recorded at small intervals over the duration of the event. In most buildings, sensors are not placed on every floor, so linear interpolation was performed to calculate the response histories on those floors that are not instrumented [24]. This information is then used to calculate the peak response demands. The peak story drift ratio (PSDR) and peak floor acceleration (PFA) are used to denote the peak response corresponding to each story and floor respectively. The maximum PSDR and PFA over all stories and floors, respectively, denoted as maxPSDR and maxPFA, are the considered EDPs. To calculate the peak response demands, the diaphragm is assumed to be rigid and torsional effects are neglected. The causal inference-based evaluation via the three alternative covariate balancing methods is implemented in addition to the efficiency and sufficiency criteria.

A procedure similar to that outlined in [37] is followed. The original sample consists of N = 479 samples. Bootstrapping is used to resample the data with replacement over $N_b = 500$ repetitions. The samples are considered individually for the covariate balancing schemes. Weights are estimated for each sample and a regression is performed. The estimated exposure-outcome curves in Figure 7 and Figure 8 are determined by using the fitted regression to predict at evenly spaced values of the considered IM between the fifth and 95-th percentile, as we do not expect to draw confident inferences outside of this range.

5.1 Covariate Balancing

This subsection presents the results of the covariate balancing sub-step of the causal inference-based evaluation illustrated in Figure 6. Recall that the goal of covariate balancing is to remove the correlation between the confounding covariates and the treatment variable (EDP) to minimize bias in the effect estimation. Balancing also reduces the effective sample size (ESS) which are summarized in Table 2. ESS measures the amount of precision in the data. Specifically, higher correlations in the data results in lower ESS due to redundancy. The considered covariates include M, R, SC, and T_1. Since the number of stories (N_s) and lateral force resisting system (LFRS) only affect the IM through the first mode period, they are not included as covariates. More formally, it can be shown that considering T_1 in the covariate balancing also addresses the confounding effects of LFRS and N_s. For this data set, it is expected that the ESS will be significantly reduced given the high correlations that M has with the exposure variables.

The results from the covariate balancing are shown in Table 1 which summarizes the Pearson correlations between the confounding covariates and treatment variable (IM) before and after balancing is performed. The post-balancing correlations, which are computed after applying the weights (w_i) to each observation, serve as an indication of the relative balancing ability of each method. The values reported in the table are the Pearson correlations between the covariates and treatment variables. Prior to balancing, the earthquake magnitude has the highest absolute correlation - which is positive, as expected - with the IMs. Also, as expected, T_1 has relatively high pre-balancing correlations with Sa_{T1} and Sa_{avg}. R has low to moderate correlations with PGV, Sa_{T1} and Sa_{avg} and low positive correlations with PGV, CAV and I_a before balancing. Recall that SC is a categorical variable, which, for the purpose of the analysis, is assigned values of 0 and 1 for site class C and D, respectively. As such, the positive (but low) pre-balancing correlations between SC and the various IMs indicate that the site class D IM levels are generally higher relative to site class C. Based on the observed post-balancing correlations, it is clear that EBCT provides the best balance, effectively eliminating the correlations for all variables while maintaining a larger ESS (see Table 2) than the other two methods. Nevertheless, the results using all three balancing methods are reported to evaluate the consistency of the relative performance of the different IMs. It should also be noted that EBCT was performed with $r = 1$ and $p = 1$, i.e. only the first moment (the mean) of the treatment and covariates are balanced. This decision was based on convergence issues that arose when attempting to balance higher order moments. IM-covariate pairs that converged did so with the consequence of a low ESS. A larger dataset might alleviate these issues, but for the current study, higher order moments were omitted.

Table 1 Pearson correlations between the IMs and covariates before and after balancing

Covariate	IM	Unweighted	IPW	CBPS	EBCT
	PGA	0.279	-0.143	0.013	0.000
	PGV	0.661	0.196	0.001	0.000
M	CAV	0.742	0.367	0.198	0.000
	I_a	0.742	0.330	0.166	0.000
	Sa_{T1}	0.514	0.001	0.041	0.000
	Sa_{avg}	0.533	-0.055	0.110	0.000
	PGA	-0.381	-0.354	-0.111	0.000
	PGV	0.046	0.146	0.003	0.000
R	CAV	0.091	0.084	-0.002	0.000
	I_a	0.092	0.051	-0.020	0.000
	Sa_{T1}	-0.007	-0.263	-0.161	0.000
	Sa_{avg}	-0.041	-0.284	-0.115	0.000
	PGA	0.011	-0.039	0.015	0.000
	PGV	0.056	0.078	-0.005	0.000
Site Class	CAV	0.079	0.047	0.071	0.000
	I_a	0.070	0.021	0.046	0.000
	Sa_{T1}	0.196	0.069	0.056	0.000
	Sa_{avg}	0.135	0.039	0.049	0.000
	Sa_{T1}	-0.470	-0.521	-0.201	0.000
T_1	Sa_{avg}	-0.309	-0.178	-0.117	0.000

Table 2 Effective sample size corresponding to the different covariate balancing methods

IM	IPW	CBPS	EBCT
PGA	132.9	69.9	220.1
PGV	99.8	52.8	171.7
CAV	126.3	53.9	129.3
I_a	90.5	36.2	130.5
Sa_{T1}	46.8	122.3	154.0
Sa_{avg}	154.2	131.7	198.3

5.2 Causal Effect Estimation Results

After the balancing is performed and the weights (w_i) for all samples are determined, a weighted linear regression is performed for each method. Both simple and local linear regressions are performed. The local linear regression provides more flexibility in the functional relationship between the exposure and outcome, so the form does not need to be assumed. It serves as a way to evaluate how the effect may vary at different levels for a given IM. The simple linear regression allows for a more direct and intuitive comparison of the population ATE, which is taken as the regression coefficient. When the relationship is nonlinear, the effect cannot be simply summarized into a single value for the population. However, the exposure-outcome curve is useful for highlighting important aspects of the relationship between the treatment (IM) and outcome (EDP).

The results of the simple linear regression are summarized in Table 3 and Table 4. The reported values are the regression coefficient (β) (causal effect) which is interpreted as a β% increase in the outcome (EDP) caused by a 1% increase in the treatment (IM). Recall that, from a conditional independence standpoint, the most desirable IM is the one with the largest β value (shown in bold font). From Table 3, it is observed that, across all three covariate balancing methods, CAV has the highest causal effect for PSDR and is therefore the best performing IM (from a causal inference perspective). It is notable that CAV is one of the least studied IMs specifically as it relates to its relative performance in estimating EDPs. However, this finding is consistent with that of the study by [39] which found that CAV is well-correlated with damage, even moreso than other IMs such as PGA. The worst performing IM for PSDR, which is the one with the lowest causal effect, is PGA. Again, this is consistent with several prior studies which showed that PGA does not perform well for displacement-based EDPs [12, 13].

The results from Table 4 show that CAV is also the best performing IM for PFA across all balancing methods. It is a bit unusual that PGA is outperformed by most of the other IMs (with the exception of Sa_{T1}) when estimating PFA. Most prior simulation-based studies have shown that, at least from an efficiency standpoint (discussed later in this section), PGA performs relatively well for acceleration-based EDPs. On the other hand, the fact PGA generally performs much better for PFA than for PSDR (when comparing the causal effects in Table 3 and Table 4)) is consistent with the principles of structural dynamics. Further, when attempting to benchmark the current results against findings from the prior literature, it is important to highlight that (i) all prior studies utilize simulation-based data whereas this study uses response measurements and (ii) the current study is based on aggregated data from multiple structure types and configurations whereas all prior studies utilize data from one structure type at a time. Although the covariate balancing is implemented to consider these differences, the analysis is limited by the amount of available data. The strata for some variables have few data points compared to others (i.e. high-rise vs low-rise buildings). The results are also dependent on the performance of the balancing methods and the resulting effective sample sizes. Results where the ESS is less than 10% of the original sample size are excluded due to the low precision.

When evaluating the results across the different methods, it is important to acknowledge their respective performance. For example, it can be observed that the causal effect estimation for PGA on PSDR is significantly larger when balanced using EBCT compared to IPW and CBPS. This may initially seem concerning without considering that not only was EBCT better at eliminating the correlations in the covariates, but for this IM-EDP pair, EBCT maintained a larger ESS. The ESS for CBPS, for example, is only just barely above the 10% cutoff criterion.

The results of the weighted local linear regressions are shown in Figure 7 and Figure 8. For brevity, only the EBCT-weighted plots are shown and inferences will be drawn from this weighting scheme as it performed best at eliminating the correlations between the covariates and IMs and maintaining a reasonable ESS. The estimated exposure-outcome curve is plotted with normally approximated 95% confidence bands using the boostrapped standard errors. The unweighted and EBCT-weighted simple regression curves are also plotted to illustrate their differences. In general, we see that although the linear model assumption was not unreasonable for some IM-EDP pairs, others display a potentially more complex relationship. Specifically, for CAV and I_a at low intensities, the response is nearly flat (in the cases of both PSDR and PFA), indicating there is almost no causal effect of the IM on the EDP at these intensity levels. This is unsurprising as these are cumulative metrics, so data points in this range likely belong to buildings that experienced little shaking and consequently, had very small responses with negligible differences. The mid-level intensities show a steep upward slope, indicating a significant increase in the causal effect, but it again decreases at the very high end of intensities, exhibiting a negative slope. The latter observation is unusual, but may be explained by the limited number of responses in this range. This conclusion is further supported by the wide confidence bands in this range. With the current size and linear-elastic nature of the set, drawing inferences in this range is not recommended. Another observation to consider is with respect to PGV. When regressed against PFA, it exhibits a notable increase in the causal effect around 1 g-s, indicating it may be a better quality predictor for this EDP at higher intensity levels. Under the assumption that the EBCT-weighted LOESS better reflects the true relationships, it shows how the simple linear regression may underestimate the effectiveness of this IM. Although limited by the size and breadth of the available responses, the weighted LOESS provides potential insight into the underlying relationships between IMs and EDPs that the simple regression does not capture.

Table 3 Estimated causal effect (β) for PSDR from post-balancing linear regression

IM	IPW	CBPS	EBCT
PGA	0.315	0.367	0.675
PGV	0.929	0.852	0.850
CAV	**1.29**	**1.62**	**1.694**
I_a	1.018	-	1.385
Sa_{T1}	-	0.577	1.112
Sa_{avg}	0.297	0.862	1.105

Table 4 Estimated causal effect (β) for PFA from post-balancing linear regression

IM	IPW	CBPS	EBCT
PGA	0.682	0.666	0.658
PGV	0.742	1.03	1.031
CAV	**1.165**	**1.576**	**1.54**
I_a	0.972	-	1.361
Sa_{T1}	-	0.625	0.603
Sa_{avg}	0.791	0.844	1.01

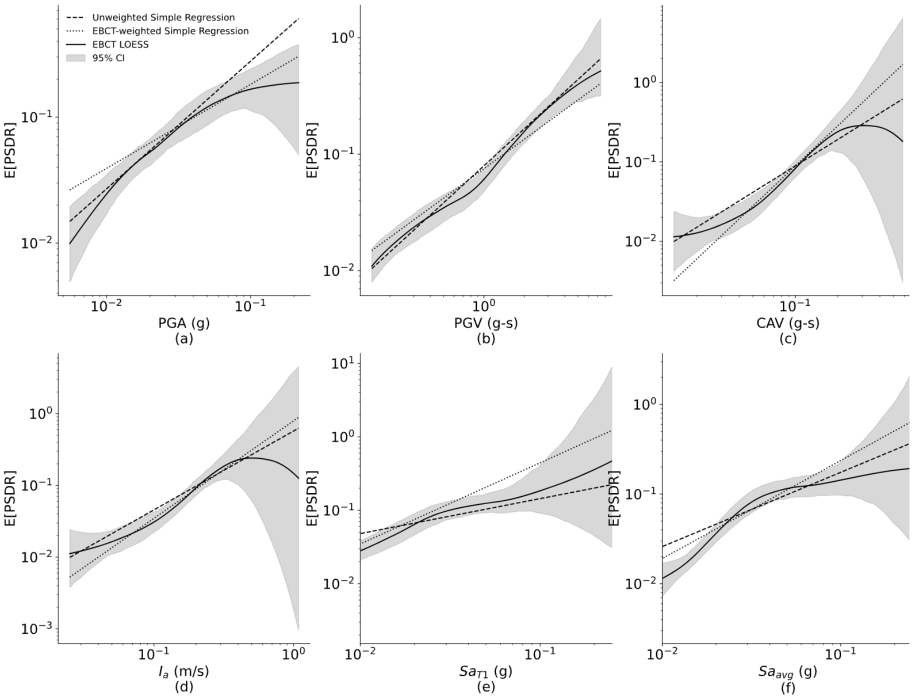

Fig. 7 Exposure-outcome Curves for PSDR vs a) PGA, b) PGV, c) CAV, d) Ia, e) Sa_{T_1} and f) Sa_{Avg} using Naive simple linear regression, EBCT-weighted simple linear regression, and EBCT-weighted LOESS

5.3 Efficiency and Sufficiency

This section presents the results of the efficiency and sufficiency-based evaluations. Table 5 summarizes the standard errors from the probabilistic seismic demand models (based on Equation 18) which serve as a measure of efficiency. For PSDR, PGV produces the lowest dispersion and is therefore the best performing IM for that EDP. Although, it is noteworthy that CAV and I_a have dispersion values that are comparable with (5% higher than) PGV, which is generally consistent with the findings from the causal analysis. It is interesting that PGA outperforms Sa_{T1} and Sa_{avg} in estimating PSDR. This runs counter to the findings from prior simulation-based studies. One possible explanation is that the inventory considered for the current study comprises primarily low-rise buildings (as summarized in Figure 2). In fact, more than half of the buildings in the inventory have five stories or less. This, coupled with the fact that most of the responses are in the linear elastic realm may explain the superior performance of PGA relative to Sa_{T1} and Sa_{avg}. For PFA, PGA produces the lowest dispersion and is therefore the most efficient. This is consistent with the findings from prior simulation-based studies which have shown that PGA performs well for acceleration-based EDPs.

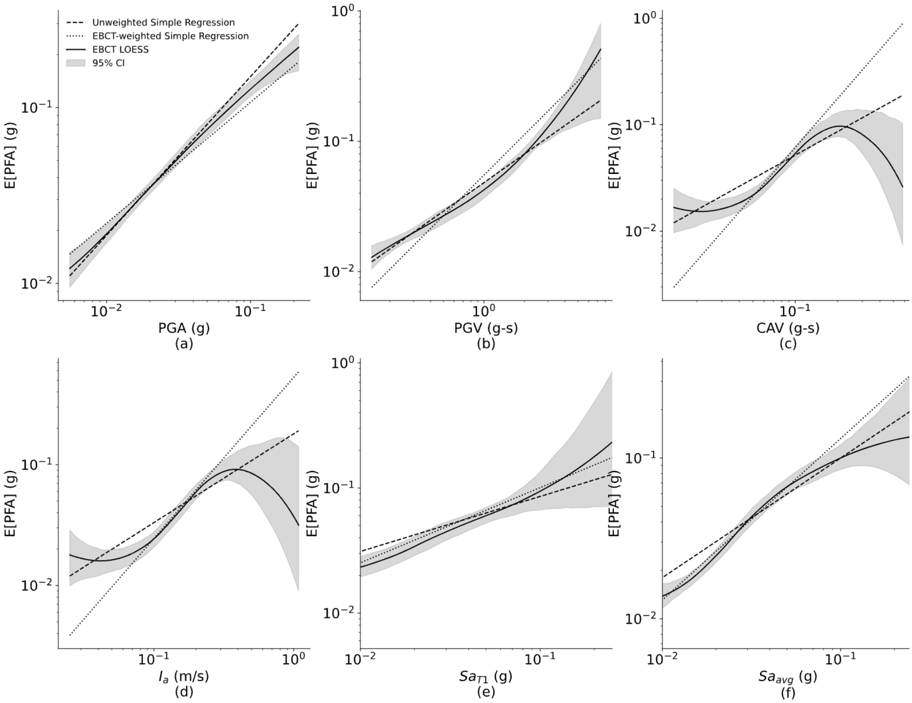

Fig. 8 Exposure-outcome Curves for PFA vs a) PGA, b) PGV, c) CAV, d) Ia, e) Sa_{T_1} and f) Sa_{Avg} using Naive simple linear regression, EBCT-weighted simple linear regression, and EBCT-weighted LOESS

The p-values from the sufficiency-based evaluations are summarized in Table 6. For each upstream parameter (i.e., M and R), the p-values for the sufficient IMs (i.e., p-values above the 5% significance threshold) are highlighted in bold. The fact that CAV and I_a are sufficient with respect to both M and R when estimating PSDR while PGA isn't, is consistent with the findings from the causal analysis. For PFA, only Sa_{T1} and Sa_{avg} are sufficient with respect to M and none of the IMs are sufficient with respect to R. The latter finding is a bit surprising and inconsistent with the results from prior studies [16, 40, 41]. However, as described in [14], this may be explained by the sensitivity of the sufficiency-based findings to the size of the data set.

Table 5 Efficiency of IMs as measured by the standard error from the PSDM (Equation 18)

IM	PSDR	PFA
PGA	1.30	**0.43**
PGV	**1.07**	0.58
CAV	1.12	0.66
I_a	1.13	0.67
Sa_{T1}	1.55	0.85
Sa_{avg}	1.41	0.63

Table 6 The p-values from the sufficiency-based IMs evaluations

Covariate	IM	PSDR	PFA
M	PGA	0	0
	PGV	**0.35**	0
	CAV	**0.56**	0
	I_a	**0.56**	0
	Sa_{T1}	0	**0.24**
	Sa_{avg}	0	**0.38**
R	PGA	0	0
	PGV	**0.44**	0
	CAV	**0.44**	0
	I_a	**0.42**	0
	Sa_{T1}	**0.27**	0
	Sa_{avg}	0	0.01

6 Conclusion

This study assessed the effectiveness of ground intensity measures (IM's) using strong motion building response data. Sufficiency, efficiency, and causal analyses are used to evaluate the performance of different IMs for potential use in a structural response reconstruction context. In the causal analyses, covariate balancing methods are implemented to control for known confounding variables that contribute to building response, and the weights from these methods were used to perform regressions that define the causal relationship between the IMs and engineering demand parameters (EDPs).

The IMs considered include peak ground acceleration (PGA), peak ground velocity (PGV), cumulative absolute velocity (CAV), Arias Intensity (Ia), first mode period spectal acceleration (Sa_{T_1}), and the geometric mean spectral acceleration over a range of periods (Sa_{avg}). For the causal analysis, each IM was balanced against common cause covariates of the IM and EDP (or confounders), then regressed using weights calculated from the covariate balancing. The simple linear regression provided an intuitive understanding of the population level average treatment effect (ATE), which represents how a change in the intensity level of the IM, on average, changes the response. A local linear regression was also performed to allow for a more complex functional relationship that captures the variation in the effect across the intensity levels. This revealed potential heterogeneous effects in the data, where the effect varies among covariate subgroups. The effects of some of the IMs on the EDPs, including when accounting for confounding, were shown to be mostly linearly positive, but the results also highlighted some complexity in the relationships.

The results from the causal analysis indicated that CAV generally outperforms the other IMs. A prior study by [39] also found that CAV generally outperforms other IMs, especially in terms of damage detection. Also consistent with prior studies, PGA was found to be the worst performing IM when estimating PSDR. There were some surprising observations, such as PGA being outperformed by most of the other IMs in the causal analysis when estimating PFA. However, this was explained by a number of factors including the structural heterogeneity within the data. In terms of efficiency, the findings were generally consistent with the causal evaluation. The sufficiency-based evaluation also produced some unexpected results. For example, the spectral IMs (Sa_{T1} and Sa_{avg}) were generally outperformed by PGA and none of the IMs were sufficient with respect to source-to-site distance. The inconsistencies (with respect to prior studies) in the sufficiency analysis may be explained by the relatively small size of the data set as it has been shown in past studies to be sensitive to sample size [14].

There are several limitations of this study which serve as motivation for future related research. The EDP data set comprises mostly elastic responses, which means that the findings cannot be generalized to nonlinear demand levels. The imbalance in the distribution of structural characteristics also produced some challenges. For instance, the majority of buildings can be classified as low to mid-rise with very few high rises. While it is conceptually possible to account for heterogeneous effects (e.g., variations in IM performance across different building heights) in the causal analysis, this was not feasible in the current study because of the relatively small size of the data set. The size of the set also limited the EBCT balancing scheme, as higher order moments were unable to be included due to non-convergence and effective sample sizes below the chosen cutoff threshold.

Future studies can address some of the aforementioned limitations by expanding the suite of strong motion data to consider a broader range of demand levels (i.e., elastic and inelastic). Alternatively, a hybrid data set can be constructed by combining recorded and simulated responses. In addition to addressing issues related to sample size, a hybrid data set can be used to ensure greater balance in the structural characteristics of the considered building inventory. This would enable an assessment of heterogeneous effects (in the causal evaluation) using subgroup analysis and potentially allow for balancing higher order moments of the IMs and covariates to eliminate correlations that may exist.

Supplementary information.

Acknowledgments. The first-author of this paper is partially supported by the GEM Fellowship Program (https://www.gemfellowship.org/).

References

[1] Roohi, M., Hernandez, E.M., Rosowsky, D.: Nonlinear seismic response reconstruction and performance assessment of instrumented wood-frame buildings—validation using neeswood capstone full-scale tests. Structural Control and Health Monitoring **26**(9), 2373 (2019)

[2] Roohi, M., Erazo, K., Rosowsky, D., Hernandez, E.M.: An extended model-based observer for state estimation in nonlinear hysteretic structural systems. Mechanical Systems and Signal Processing **146**, 107015 (2021)

[3] Erazo, K., Hernandez, E.M.: Uncertainty quantification of state estimation in nonlinear structural systems with application to seismic response in buildings. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering **2**(3), 5015001 (2016)

[4] Ching, J., Beck, J.L., Porter, K.A., Shaikhutdinov, R.: Bayesian state estimation method for nonlinear systems and its application to recorded seismic response. Journal of Engineering Mechanics **132**(4), 396–410 (2006)

[5] Çelebi, M., Sanli, A., Sinclair, M., Gallant, S., Radulescu, D.: Seismic monitoring instrumentation needs of a building owner and the solution: A cooperative effort. In: 2003 ASCE/SEI Structures Congress and Exposition: Engineering Smarter, pp. 359–360 (2003)

[6] Miranda, E.: Use of probability-based measures for automated damage assessment. The structural design of tall and special buildings **15**(1), 35–50 (2006)

[7] Porter, K., Mitrani-Reiser, J., Beck, J.L.: Near-real-time loss estimation for instrumented buildings. The Structural Design of Tall and Special Buildings **15**(1), 3–20 (2006)

[8] Roohi, M., Hernandez, E.M.: Performance-based post-earthquake decision making for instrumented buildings. Journal of Civil Structural Health Monitoring **10**, 775–792 (2020)

[9] Shome, N.: Probabilistic Seismic Demand Analysis of Nonlinear Structures. Stanford University, Stanford (1999)

[10] Mehanny, S.S.F.: Modeling and Assessment of Seismic Performance of Composite Frames with Reinforced Concrete Columns and Steel Beams. Stanford University, Stanford (2000)

[11] Cordova, P.P., Deierlein, G.G., Mehanny, S.S., Cornell, C.A.: Development of a two-parameter seismic intensity measure and probabilistic assessment procedure. In: The Second US-Japan Workshop on Performance-based Earthquake Engineering Methodology for Reinforced Concrete Building Structures, vol. 20, p. 0 (2000)

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

[13] Shokrabadi, M., Burton, H.V.: Ground motion intensity measures for rocking building systems. Earthquake spectra **33**(4), 1533–1554 (2017)

[14] Burton, H.V., Baker, J.W.: Evaluating the effectiveness of ground motion intensity measures through the lens of causal inference. Earthquake Engineering & Structural Dynamics (2023)

[15] Dávalos, H., Miranda, E.: Filtered incremental velocity: A novel approach in intensity measures for seismic collapse estimation. Earthquake Engineering & Structural Dynamics **48**(12), 1384–1405 (2019)

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

[17] Tothong, P., Luco, N.: Probabilistic seismic demand analysis using advanced ground motion intensity measures. Earthquake Engineering & Structural Dynamics **36**(13), 1837–1860 (2007)

[18] Du, A., Padgett, J.E.: Entropy-based intensity measure selection for site-specific probabilistic seismic risk assessment. Earthquake Engineering & Structural Dynamics **50**(2), 560–579 (2021)

[19] Dhulipala, S.L., Rodriguez-Marek, A., Ranganathan, S., Flint, M.M.: A site-consistent method to quantify sufficiency of alternative ims in relation to psda. Earthquake Engineering & Structural Dynamics **47**(2), 377–396 (2018)

[20] Jalayer, F., Beck, J., Zareian, F.: Analyzing the sufficiency of alternative scalar and vector intensity measures of ground shaking based on information theory. Journal of Engineering Mechanics **138**(3), 307–316 (2012)

[21] Abdelmalek-Lee, E., Burton, H.: A dual kriging-xgboost model for reconstructing building seismic responses using strong motion data. Bull Earthquake Eng (2023)

[22] Sun, H., Burton, H., Zhang, Y., Wallace, J.: Interbuilding interpolation of peak seismic response using spatially correlated demand parameters. Earthquake Engineering & Structural Dynamics **47**(5), 1148–1168 (2018)

[23] Sun, H., Burton, H., Wallace, J.: Reconstructing seismic response demands across multiple tall buildings using kernel-based machine learning methods. Structural Control and Health Monitoring **26**(7), 2359 (2019)

[24] Sun, H., Burton, H.V., Stewart, J.P., Wallace, J.W.: Development of a generalized cross-building structural response reconstruction model using strong motion data. Journal of Structural Engineering **148**(6), 04022053 (2022)

[25] Ni, P., Sun, L., Yang, J., Li, Y.: Multi-end physics-informed deep learning for seismic response estimation. Sensors **22**(10), 3697 (2022)

[26] Roohi, M., Hernandez, E.M., Rosowsky, D.: Nonlinear seismic response reconstruction in minimally instrumented buildingsâ€" validation using neeswood capstone full-scale tests. Structural Health Monitoring 2019 (2019)

[27] Abdelmalek-Lee, E., Jain, T., Madero, S., Sun, H., Burton, H., Wallace, J.: Relational database for building strong motion recordings used for seismic impact assessments. Earthquake Spectra (2023)

[28] ASCE: Minimum design loads and associated criteria for buildings and other structures. American Institute of Steel Construction, Chicago-Illinois American Society of Civil Engineers (2017)

[29] Splawa-Neyman, J., Dabrowska, D.M., Speed, T.P.: On the application of probability theory to agricultural experiments. essay on principles. section 9. Statistical Science **5**(4), 465–472 (1923)

[30] Rubin, D.B.: Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association **100**(469), 422–331 (2005)

[31] Hinkley, D.V.: Randomization analysis of experimental data: The fisher randomization test comment. Journal of the American Statistical Association **75**(371), 582–584 (1980). Accessed 2023-06-20

[32] Pearl, J.: Causality. Cambridge University Press, Cambridge (2019)

[33] CESMD: Center for Engineering Strong Motion Data. https://www.strongmotioncenter.org/

[34] Rubin, D.B.: Using multivariate matched sampling and regression adjustment to control bias in observational studies. Journal of the American Statistical Association **74**(366), 318–328 (1979). Accessed 2023-06-20

[35] Imai, K., Ratkovic, M.: Covariate Balancing Propensity Score. Journal of the Royal Statistical Society Series B: Statistical Methodology **76**(1), 243–263 (2013) https://academic.oup.com/jrsssb/article-pdf/76/1/243/49514325/jrsssb_76_1_243.pdf. https://doi.org/10.1111/rssb.12027

[36] Fong, C., Hazlett, C., Imai, K.: Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. The Annals of Applied Statistics **12**, 156–177 (2018). https://doi.org/10.1214/17-AOAS1101

[37] Tübbicke, S.: Entropy balancing for continuous treatments. Journal of Econometric Methods **11**(1), 71–89 (2022). https://doi.org/10.1515/jem-2021-0002

[38] Hainmueller, J.: Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. Political Analysis **20**(1), 25–46 (2012). https://doi.org/10.1093/pan/mpr025

[39] Muin, S., Mosalam, K.M.: Cumulative absolute velocity as a local damage indicator of instrumented structures. Earthquake Spectra **33**(2), 641–664 (2017) https://doi.org/10.1193/090416EQS142M. https://doi.org/10.1193/090416EQS142M

[40] Padgett, J.E., Nielson, B.G., DesRoches, R.: Selection of optimal intensity measures in probabilistic seismic demand models of highway bridge portfolios. Earthquake engineering structural dynamics **37**(5), 711–725 (2008)

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