Leveraging machine learning techniques to support a holistic performancebased seismic design

Mohsen Zaker Esteghamati

Department of Civil and Environmental Engineering, Utah State University, Logan, UT, USA

Abstract

The increasing vulnerability of communities to natural hazards motivates novel design and assessment methods to ensure that the built environment performs optimally during its lifetime. The current design methodologies do not account for life-cycle impacts across multiple performance domains such as economy and environment. Therefore, low-effort and designer-centric computational methods are needed to support a multi-objective performance-based design, from the conceptual stage to design development. This study presents a framework for a holistic performance-based seismic design of buildings. The proposed framework leverages machine learning techniques to extract the implicit, and highly complex, relationship between design parameters, geometric configuration, and performance measures. At early design, data-driven surrogate models (trained on performance inventories) are used to identify candidate structural systems and their approximate design parameters. At the detailed design stage, a deep learning-based engine generates seismic risk estimates based on simpler nonlinear static analysis on the candidate systems or their equivalent low-order dynamic models. A case study illustrates the framework's application for performance-based seismic design of multistory commercial buildings in Charleston, SC.

Keywords: Machine learning, performance-based earthquake engineering, seismic design, deep learning

1. Introduction

Performance-based earthquake engineering (PBEE) [1,2] is a modular framework to quantify the impact of earthquakes on the built environment through global performance measures. The framework consists of four independent modules of seismic hazard, structural response, damage, and loss analyses. Each module is represented as a conditional probability, where analysis data are used to characterize the probability distribution of the module's random variable(s). This modularity facilitates a rigorous and interdisciplinary approach to account for different factors that shape seismic performance, enabling a pathway toward a more *holistic* seismic design. A holistic design (Figure 1) directly accounts for performance across different domains over the entire design stages, from early design concepts to design development and post-construction operation.

Despite the promising features of PBEE, this framework is still not widely used after nearly two decades since its inception, except for special buildings that justify the additional time and effort needed. Most PBEE efforts are limited to seismic evaluation of novel structural systems such as buildings supplemented with dissipative devices (e.g., structural fuses [3,4], hysteretic devices [5,6]), fiber reinforced polymer- or shape memory alloy-based components [7–10], special braces (e.g., gap inclined [11], crescent-shaped [12], knee bracing[13]) or strongback assemblies [14,15].



Figure 1. A holistic design approach: the relationship between design and performance is accounted at every stage.

Perhaps the most notable effort in using PBEE for design was directed at tall buildings as part of the PEER/ATC project in 2006 [16].

A critical barrier in using PBEE is the simulation challenges and modeling complexities of developing sophisticated nonlinear finite element models, selecting appropriate ground motions, and performing time-consuming time history analysis. Such complexities substantially increase when applying PBEE to generate design alternatives (i.e., at the early design stage), or considering different performance objectives simultaneously (particularly if a different type of analysis is needed for each objective). Therefore, different studies aimed to simplify performing PBEE by introducing approximate analysis methods or developing efficient computational platforms.

Machine learning (ML)-based methods have been recently introduced to alleviate PBEE computational challenges, such as predicting structural response and developing models based on experimental, health monitoring, or field reconnaissance data [17,18]. ML methods can aid with both achieving computationally inexpensive models and design automation. Nevertheless, most studies on structural response estimation focused on predicting nonlinear responses (or fragility functions) based on ground motion-related and detailed design parameters [17]. Section 2 reviews the current literature on ML applications for seismic design and performance evaluation.

This study proposes an ML-assisted modular framework to implement PBEE for a holistic seismic design, encompassing both the early stage and design development. To this end, the framework provides risk-informed guidance at the early design on system selection and initial ranges for design and configuration parameters, and ensures that the desired performance (e.g., total repair cost) is achieved at the design development stage. The proposed framework relies on the hypothesis that there is a complex relationship between building performance, configuration, and design. This relationship allows shifting the time and effort needed to perform such assessments from designers into ML models, hence sidestepping performing PBEE assessments over a large design space. In addition, the ML models facilitate rigorous sensitivity assessment

and optimization to identify the most critical design decisions. Lastly, the framework considers earthquakes' economic and environmental impacts and their interaction. The organization of this chapter is as follows: Section 2 reviews the literature on using ML for seismic design and assessment, Section 3 provides an overview of the framework and each module, and Section 4 presents a case study that illustrates different modules' applications.

2. Background

The application of ML in PBEE has mainly focused on its predictive capabilities to serve as *surrogates* for computationally-expensive finite element simulations. The majority of previous studies have addressed component-level predictions such as RC columns [19,20], concrete-filled steel columns [21,22], masonry walls[23], or RC walls [24,25]. Several authors aimed to predict seismic responses of archetypal buildings based on design- and hazard-related parameters, as shown in Table 1. Furthermore, instead of developing ML-based mapping functions, a few studies focused on creating novel ML architectures to directly estimate time-history responses based on ground motion (GM) sequences. For example, Ahmed *et al.* showed that stacked long short-term memory (LSTM) accurately predicts damage states for ductile and non-ductile frames [26]. On the other hand, Luo and Paal developed an ML-based solver to estimate structure response by constructing the stiffness matrix of RC frames using experimental data of RC components [27].

In a quest to generalize ML models, efforts have been made to predict the seismic performance of building inventories with varying configurations and designs. For example, Nguyen *et al.* applied 8 ML models to a database of 468 steel moment-resisting frames and showed that random forest could predict building damage states with an accuracy level of 98% [28]. In addition, they concluded that spectral acceleration at 1 s is the most important predictor. Guan *et al.* examined data-driven and hybrid ML models to predict the drift demand of 621 steel frames, and showed that random forests could accurately predict the seismic demands for both low- and high-rise buildings, although hybrid ML models provide higher accuracy than purely data-driven ones. Moreover, they found that the ratio of floor height to the building height and spectral acceleration and displacement at the fundamental period are the most influential for hybrid ML models, whereas spectral acceleration at the first mode dominated the purely data-driven models [29].

Study	Year	Structure type	ML method description		
			Predictors	Target	Selected algorithm(s)
Hwang et al.[30]	2021	RC frame (2) ^a	Plastic hinge properties, damping, period, $S_a(T_1)$	MIDR, Collapse	Boosting (6) ^b
Demirtzis <i>et al.</i> [31]	2022	RC buildings (30)	Height, eccentricity, ratio of walls base shear, and 14 GM parameters	MIDR	LightGBM (15)
Kazemi et al. [32]	2023	Steel frames (8)	Weight, period, Sa(T1), soil type, record number	MIDR	XGboost,Breg, HistGBR, ERTReg (11)
Dabiri et al. [33] Kiani et al	2022	fragility database (214)	Footprint, height, period	Fragility parameters	Decision trees (5)

Table 1. Literature review of recent studies usi	ing ML to predict the	seismic performance of	buildings
--	-----------------------	------------------------	-----------

^a number in parenthesis shows the number of studied buildings

^b number in parenthesis shows the number of studied algorithms

Recently, ML-based approaches have been applied to regional-scale and community-level seismic assessments of buildings. Kourehpaz and Molina Mutt applied ML algorithms on a database of 36 RC frames, and used the trained model for portfolio assessment of RC wall-building across the Metro Seattle region [34]. Instead of using an ML model trained on small datasets, Lu *et al.* [35] developed an open-source scientific framework that can perform city-scale nonlinear time history analysis and applied this framework to 1.8 million buildings in San Francisco Bay Area using SimCenter [36] workflow. Nevertheless, the reviewed literature suggests that despite the significant existing literature on using ML at different scales, an ML-assisted framework is still needed to achieve performance-consistent design throughout the entire design process.

3. ML-assisted holistic seismic design

This section presents the motivation and overview of the developed framework. Next, the application of the methodology for each design stage is discussed. Lastly, suitable databases to supplement the framework are introduced.

3.1. Objectives & Motivation

A conventional design often follows the designer's intuition and judgment from a precedent-based perspective [37]. These cognitive biases [38] can lead to missing more resilient and sustainable structural systems. In addition, the conventional seismic design aims to satisfy the safety requirements (such as required strength and stability) of building code under severe earthquakes. However, this approach cannot explicitly consider other performance objectives (e.g., minimizing repair costs or downtime) over the building life cycle, or even guarantee that the life safety objectives will be met. Computational tools integrate simulation and quantitative measures to supplement the design procedure and lessen the role of intuition and precedence.

PBEE can be a practical computational approach to quantify the seismic-related consequences of different design decisions. PBEE also offers a systematic approach to treat different sources of uncertainties (due to hazard and structural modeling) using global performance measures that are easily understandable by the entire design team. Furthermore, due to PBEE's modular nature, the assessment results can be seamlessly combined with other quantitative risk measures to provide a comprehensive evaluation.

Several challenges need to be addressed to increase the application of PBEE in the design. First, PBEE requires ample data and particular skills (e.g., hazard modeling and simulation) that are not common in large design firms, and non-existent in small ones. These sequential analyses span different disciplines ranging from engineering seismology to structural engineering. Second, PBEE assessment is time-consuming and computationally expensive, and the needed effort and time significantly increase as the number of candidate design alternatives grows. For example, for each new design alternative, new structural models should be developed, and based on the assessment type, a new ground motion selection needs to be performed if the decision notably changes the structure's stiffness and mass.



Figure 2. Motivation and objectives of the proposed framework

The main objective of this study is to combine PBEE and data-driven modeling, where datadriven models overcome the computational challenges and offer a fast means to extend and scale PBEE over the design space. As shown in Figure 2, this data-driven performance-based seismic design framework provides the following advantages over traditional heuristics:

- 1. Compilation of a broad design space: A successful design should consider many design alternatives (i.e., design space exploration) so that no high-performing candidate system is excluded. Here, statistical sampling-based approaches are used to populate the design space efficiently. Sampling-based approaches also allow space reduction to reduce the sheer effort needed for the high cardinality of real-world design problems.
- 2. Consideration of holistic life-cycle performance: A holistic design should consider multiple performance objectives such as environmental, economic, and hazard-related performance. Most previous efforts are aimed at the combined economic and environmental impacts of the design decisions, primarily for shape optimization of structural components. In addition, these studies overlook the interaction of different performance measures. The framework uses multiple performance measures, where the interaction is considered through refined assessments that generate data needed for ML algorithms.
- **3.** Characterization of seismic performance range: Commonly, PBEE is not applied for designing ordinary buildings (such as multistory concrete or steel frames), or is reserved

for later stages of design or assessment. This framework extends PBEE application from early stages (to characterize seismic performance *range* over the approximate description of different candidate systems) to detailed design and optimization.

- 4. Treatment of uncertainties: Both PBEE and data-driven introduces different sources of uncertainties into the design. PBEE assessments are characterized by uncertainties related to hazard characterization, structural response analysis, and loss estimation, whereas data-driven models introduce uncertainties into the input variables (e.g., inherent randomness, insufficient training data), algorithm hyperparameters, and mapping algorithms. The proposed framework leverages the uncertainty quantification methods native to both methodologies to treat the most governing sources of uncertainties comprehensively.
- **5. Optimization and sensitivity analysis of design decisions:** The data-driven surrogate models facilitates performing a larger number of PBEE analysis at low computational costs, hence allowing to identify the most critical design decisions through sensitivity assessments (such as variance-based sensitivity), or determining the optimum range of these parameters to achieve the intended performance range.

3.2. Framework Overview

The proposed framework consists of two modules: early design and detailed design. At early design, the framework uses imprecise information provided by the designer in a pipeline of different simplified models to aid with selecting suitable structural systems, and providing preliminary insights on possible sizing. At detailed design, the framework uses a deep learning (DL) model to provide precise loss estimates based on more detailed design descriptors for the selected systems, where the DL can readily map seismic fragility to changes in structural properties.

3.2.1. Early design module

Early design defines the design problem and explores the solution space to identify candidate systems. The proposed framework provides risk-informed insights on the best possible structural systems and preliminary estimates on system design and configuration (e.g., weight, footprints). Figure 3 shows the schematics of the early design module. This module relies on surrogate models to estimate the performance range based on crude design and topology information. The word "surrogate" here refers to different lower-order models, such as knowledge-based or data-driven models, that are computationally inexpensive and conform to the limited data availability of earlier design stages. For example, approximating performance based on assessments of similar buildings from literature (i.e., knowledge-based), or using decision trees that can predict performance based on a set of assumed building characteristics (i.e., data-driven).

As shown in Figure 3, the primary workflow in the proposed data-driven framework uses supervised-learning ML algorithms to build surrogate models from PBEE assessment data. This data can be generated by alternate pathways and is supported through a knowledge-based module. As a result, the framework provides a fast means to explore design space, where the designer only needs to change the statistical surrogate model input accordingly to get instant assessment results. At the same time, the designer does not need to perform detailed performance-based assessments, but rather utilize a performance data inventory. The latter could be accomplished by exploiting



Figure 3. Early design module

available open databases, prior assessments, or performing simplified PBEE assessments. The first module requires performance databases to support the surrogate models. Currently, few open databases provide adequate data granularity over different building taxonomies and performance measures. Section 3.3 discusses several available databases for seismic and environmental performance.

The framework provides designers with a sequential approach to leverage three different surrogate modeling techniques for performing convergence-divergence cycles, as shown in Figure 4. In a typical early design procedure, the designer performs a series of convergence (i.e., removing unfavorable design alternatives) and divergence (i.e., introducing new design alternatives) cycles to arrive at the final alternatives. This sequential approach uses surrogate models with lower fidelity at the earlier stage to remove redundant alternatives from the large and highly variable (both intra- and inter-system variabilities) initial design space. Next, data-driven surrogate models are implemented to estimate hazard performance ranges and compare alternatives in the second convergence cycle. Lastly, low-order dynamic models, such as simplified single-degree-of-freedom (SDOF) systems, are developed for the few selected candidates, and parametric studies are performed to explore those design alternatives as a divergence cycle.



Figure 4. Sequential workflow to use surrogate models for early design

3.2.2. Detailed Design Module

After selecting the final candidate structural system(s) and initializing their parameter range, the second module provides a more detailed seismic design. Here, the designer needs to provide a more detailed description of the structural system through simplified nonlinear static (i.e., pushover) analysis. A deep learning-based engine then translates the resulting force-deformation relationship (i.e., pushover curves) into probabilistic seismic demand models (PSDMs), which can be readily mapped to seismic loss through pre-defined damage states and vulnerability functions.

As shown in Figure 5, the deep learning model (i.e., *P2M*) comprises an encoder to encode pushover curves and a decoder to construct the PSDMS through the encoded values. The loss function couples the changes in estimated intercept and slope values, avoiding separate training and overfitting [39]. PSDM's error characterizes fragility variance and impact loss prediction. Therefore, *P2M* explicitly estimates PSDM error using a separate network to track and update error estimation as a moving target. Different network types can be implemented for each network; however, preliminary results indicate that long short-term memory (LSTM) could be slightly more efficient in capturing PSDM slope and intercept [39].

The *P2M* mapping also provides an efficient means to incorporate structural modeling uncertainties into the assessment through the "stochastic pushover" [39] method. This method changes the structural model properties, such as plastic hinges properties, and repeats pushover analysis for each set of new parameters. Next, *P2M* estimates the variation in PSDMs due to structural models change, creating a direct mapping between resultant fragility curves and structural models through pushover analysis. Therefore, the designer can perform a new pushover analysis for any given design decision and instantly calculates the changes in the structure's fragility. The designer can also perform a few pushover analyses to estimate upper and lower-bond pushover curves and characterize the seismic performance range in terms of fragility or loss variances. In addition, since this approach directly measured the impact of different design



Figure 5. Detailed design module

decisions on the fragility curves (and consequently loss estimates) without performing a nonlinear dynamic analysis, it is suitable for the dynamic and fast-paced nature of the design.

3.3. Performance Data Inventories

Preparing data, including acquisition and preprocessing, is the first and perhaps the most critical step of any ML-based framework. The proposed methodology requires extensive open data on the seismic and environmental performance of different building systems. Future hazard performance databases can be presented as relational databases (RDBs) to promote easier data sharing and collaboration. RDBs organize interrelated data tables using shared fields (i.e., keys), facilitating more efficient data storage and integration. These databases can be hosted on Design-Safe CI, which allows users to extract data by defining a search criterion through web-based SQL queries. The criterion is defined by using database keys to combine tables (i.e., performing a *join*), and filtering the results to meet user conditions. Adopting RDBs and open data practices scale the individual efforts in PBEE to community-driven scientific workflows necessary for addressing seismic resiliency challenges.

3.5.1. Vulnerability data

The global synergies on vulnerability modeling resulted in several databases that can be used to quickly characterize seismic performance in terms of loss or probability of experiencing a given damage state. Such data (particularly fragility curves) can also be used as the input to calculate higher-level performance measures. Among these databases, the Global Earthquake Model (GEM) provides fragility curves, damage-to-loss models, or capacity curves for different structures [40].

Curated published literature on PBEE for different structural systems can also be used as a resource to provide such data. Therefore, Esteghamati *et al.* compiled an open relational database, denoted as "INventory of Seismic Structural Evaluation, Performance functions, and Taxonomies (INSSEPT)" [41,42], to aggregate 222 PBEE case studies from 39 papers. The current version of INNSEPT provides simulation-based data on building topology, design, location, and seismic performance in terms of fragility and PSDM parameters for 144 moment-resisting frames, 30 braced frames, and 24 wall-building systems.

3.5.2. Environmental performance data

Life cycle assessments (LCAs) provide a systematic approach to evaluating environmental impacts. Different LCA approaches exist with different levels of complexity, such as process-based and economic input-out methods. Unlike PBEE assessments, several guidelines establish standardized approaches to performing LCA. However, standardized LCA studies require the definition of scope and system boundary, which vary between different studies and prevent their reuse for other projects. A few studies performed comparative LCAs between different structural systems, where their published impact data can be used as an approximate value, useful for the early design stage. For example, Esteghamati *et al.* compared six commercial buildings with varying foundation, structural, and envelope systems [42], provided impact data, and LCA and energy models on a public repository[43].

3.5.3. Seismic recovery data

The current PBEE assessments mainly focus on seismic risk in terms of monetary metrics such as repair cost. However, the recovery of buildings after an earthquake event is an essential predictor of community resilience. Post-earthquake recovery models rely on empirical data, such as duration-based parameters that define recovery activities. Omoya *et al.* developed a relational database to compile the recovery efforts of 3695 buildings after the 2014 Napa earthquake [44]. The relational database provides empirical data on building general topology, site, observed damage, and duration-based recovery measures (such as timestamps for initiation and completion of permitting or repair. Such data can then be integrated into recovery models through analytical approaches [45,46]

4. Case study

This section discusses the preliminary application of the proposed framework for the seismic design of mid-rise commercial buildings for a site in Charleston, SC.

4.1. Design problem definition

A structural engineer is assumed to be consulted on identifying high-performing design alternatives for a multistory commercial building project in Charleston, SC. This site is subjected to high wind and earthquake hazards, although only earthquake hazard is considered. For the sake of this illustration, several assumptions are made. First, building footprints are commonly determined based on zoning and ordinance requirements and thus have a fixed value. However, it is assumed that the designer can find the best general topology that meets certain limitations (e.g., mid-rise, building dimension in each direction between 42 ft to 180 ft). Second, although the framework can consider multiple performance objectives, this illustration only focuses on seismic-



Figure 6. Compiling concrete frame inventory: (a) the workflow to generate performance data,(b) a typical plan of a symmetric concrete frame building, (c) general configuration of the nonlinear models, (d) generated synthetic GMs for the site of this study

related repair costs and embodied carbon during the building's service life of 50 years. Third, while the framework aims to guide structural system selection, the illustration scope is limited to only multistory concrete frames due to the unavailability of adequate data for the considered site.

4.2. Building inventory description

Currently, a comprehensive performance inventory does not exist for the considered site. Therefore, an inventory of 720 RC moment frames was developed through an automated workflow as a proof-of-concept to provide data for the framework. Figure 6.a shows the schematics of the workflow to generate this inventory. The geometric configuration of a multistory symmetric concrete frame can be defined through four parameters of dimensions in two perpendicular directions, the number of stories, and bay length (Figure 6.b). A Latin Hypercube Sampling was then used to generate 60 samples of these topology parameters based on their practical ranges. For each set of sampled topology parameters, 12 different sets of section sizes are generated for frame beams and columns through a pseudo-directional sampling. Additional information on frame and section sampling can be found in [47]. The paired topology and designed sections are then used to create finite element models suitable for performance-based seismic assessments.

4.2.1. Finite element modeling

A two-dimensional finite element model was developed in OpenSees [48] for each sampled pair of topology and design parameters. As shown in Figure 6.c, the finite element model uses a concentrated plasticity approach, representing frame members as elastic elements with two nonlinear plastic hinges at both ends. The hinge properties were determined based on Ibarra-Medina-Krawinkler (IMK) model [49]. The parameters of the IMK model's backbone curve (such

as pre-and post-capping deformation) and the cycle deterioration were obtained from a set of regression equations proposed by Haselton for symmetric RC sections [50]. The stiffness of plastic hinges and elastic elements were adjusted to remove the unrealistic damping forces due to plastic hinge formulation [51]. The leaning column concept accounted for the P-delta effects of gravity frames [52].

4.2.2. GM selection & time-history analysis

Due to Charleston's special seismicity, conventional GM selection methods based on generic GM suites are inefficient. Therefore, a careful geologically-realistic seismic hazard analysis [53] was used to generate synthetic GMs. Four hazard levels were selected corresponding to 2%, 5%, 7%, and 10% exceedance probabilities in 50 years. Earthquake magnitude (M) and distance (R) pairs were generated based on an inverse transform sampling from the hazard deaggregation at each considered hazard level, where the number of M-R pairs in each bin was proportional to the bin's contribution to the hazard [54]. A synthetic GM record is generated for each M-R pair using a stochastic method described in [55]. Two GM sets were developed to perform the nonlinear time-history analysis: (a) 20 records scaled to maximum considered earthquake levels, (b) 80 unscaled records (20 records for each considered hazard level). The scaled record set was used to check the structural models based on code requirements for drift and collapse, whereas the second set was used to perform a cloud analysis considering different response levels [56,57]. In cloud analysis, the spectral acceleration at the first mode (i.e., $Sa(T_I)$) was used as the GM intensity measure (IM). Floor peak acceleration and maximum interstory drift responses were recorded as the engineering demand parameters (EDPs).

4.2.3. Seismic loss assessment

An assembly-based approach [58] was used to derive life-cycle repair costs based on considered EDPs. This approach aggregates building component losses to assembly categories, reducing the required data and computation, which is suitable for portfolio assessments.

In this approach, collapse and non-collapse losses are separated. The collapse losses were taken as building replacement and demolishing costs multiplied by collapse probability, where collapse was defined as the building reaching a maximum inter-story drift of 10% [59]. Logistic distribution was used to derive collapse probability. The average collapse probability for the compiled building inventory was 0.52%, with a standard deviation of 0.19%. Out of 720 buildings, only 1.5% did not collapse under any given record, whereas 2.4% had a collapse probability over 1% with a maximum value of 2.4%. For non-collapse losses, the repair cost for three assemblies (structural, non-structural drift-sensitive, and non-structural acceleration sensitive) and their corresponding damage states were derived from the HAZUS manual [60]. A lognormal distribution was fitted to EDP-IM pairs from cloud analysis to calculate damage distributions. Since HAZUS repair costs are in terms of building construction costs, the RSMeans database was used to calculate the monetary values. Lastly, the expected total loss values were integrated over the IM intensities to estimate the expected annual loss (EAL) values for each assembly.

Figure 7 compares the variation of different EAL values (normalized by building replacement cost) with respect to the structure's fundamental period, floor area, and average beam section size. While structures with a higher fundamental period show a larger structural loss, they have a smaller



Figure 7. Comparison of loss values variation with respects to building period, floor area, and average beam area

non-structural acceleration-sensitive loss. Similarly, a positive correlation between floor area and structural losses (and a negative correlation to non-structural acceleration-sensitive losses) is observed. However, increasing average beam section sizes reduces structural loss and slightly increases non-structural acceleration-sensitive losses for the considered inventory. Nevertheless, the impact of all these parameters on collapse loss is small. The ML models must capture these underlying relationships to predict the total loss due to different topology and design parameters.

4.2.4. Environmental impact assessment

Whole-building life cycle assessments (WLCA) were performed on the developed building inventory following a process-based approach described in ISO14044. The WLCA boundary included the structural frames, floors, envelopes, and partitions [61]. The frame topology and design information were obtained from the automated workflow output, whereas partition material was calculated based on FEMA P-58 normative tool[47]. A brick-veneered concrete masonry block was assumed for the envelope system with a window-to-wall ratio of 0.31. The environmental impact assessment was carried out for embodied global warming potential (GWP) during a building service life of 50 years.



Figure 8. Comparison of embodied GWP relationship to height and floor area

Embodied GWP due to earthquake-related repairs was also accounted for through the cost ratio approach. In this approach, it is assumed that the ratio of repair-related GWP to initial embodied GWP is directly related to the ratio of repair cost to the building construction cost. As shown in Figure 8, embodied GWP is linearly related to building gross area (and consequently weight) with an R-squared value of 0.98. This high R-squared value is mainly due to the small values of seismic loss, resulting in embodied GWP being a function of initial construction GWP (and hence proportional to building footprint). It should be noted that not all topology-related parameters have a high correlation to the embodied GWP [47].

4.3. ML-assisted early design

4.3.1. Data-driven surrogate models

The performance inventory was used to generate data-driven surrogate models through an ML pipeline depicted in Figure 9.a. This ML pipeline splits the database into training and testing sets, selects candidate features (using statistical or recursive feature elimination (RFE) methods), selects an algorithm, tunes the ML model hyperparameters, and finally, evaluates them based on measures such as R-squared or root mean square error. Five algorithms were examined, and the model



Figure 9. ML-assisten early design module results: (a) ML pipeline to train data-driven surrogate models, (b) SVM prediction of total loss on training set, (c) Sensitivity assessments of SVM to model predictors: height (*H*), Y-dimnesion (L_y), floor area (*FA*), lateral weight (W_L), average beam area ($\overline{B_A}$), average beam area in the first floor($\overline{B_{A,1}}$), average beam reinforcement ratio ()



Figure 10. Comparison of embodied GWP relationship to height and floor area

accuracy was compared. The results show that support vector machines (SVM) provide the highest accuracy, predicting total seismic loss with an R-squared value of 0.92 (Figure 9.b) using only crude topological and design parameters. Nevertheless, the loss prediction values were sensitive to how the building cost function is formulated. The highest accuracy is achieved when the loss is defined in terms of actual dollars spent (cost function as a multiplier of building footprint), whereas defining loss as the percentage of replacement cost (cost function of unity) leads to the lowest accuracy [47].

ML models error can be decomposed into two competing factors: bias and variance. Bias quantifies the average difference between actual response values and model prediction, whereas variance measures the variability in ML model prediction for a given data. There is a tradeoff between ML model bias and variance: simpler models might not be adequate for prediction (i.e., high bias) yet exhibit low variance, whereas flexible models might result in excellent prediction (i.e., low bias), yet the models' accuracy might not be consistent (i.e., high variance). Figure 10 compares the bias and variance of studied ML algorithms to predict the structural and collapse losses. First, the *k*-nearest neighbor shows high bias and variance for both loss types, resulting in a large overall error. Multiple linear regression generally shows lower variance for both loss types; however, it is highly biased for collapse loss prediction. For structural losses, it can be observed that MLR shows the lowest variance, whereas SVM shows the lowest bias. All models except for extreme gradient boosting and SVM show high bias for collapse loss.

To predict total loss, the selected SVM model uses building height (*H*), dimension perpendicular to the analyzed frame (L_y), floor area (*FA*), lateral weight (W_L), average beam area over the entire building ($\overline{B_A}$), average beam area over the first floor ($\overline{B_{A,1}}$), and average beam longitudinal rebar percentage ($\overline{B_S}$). As shown in Figure 9.c, a variance-based sensitivity assessment showed that the SVM model's prediction variance is dominated by the higher-order interaction of all these predictors. Among different predictors, floor area, height, and average beam area were the most influential predictors for total loss.

4.3.2. Sequential design space exploration

As described in Section 3.2.1, a sequential workflow was created by linking knowledge-based, data-driven, and low-order mechanistic surrogate models. The knowledge-based model is achieved by regressing loss values from literature (query through INSSEPT, among other databases) for the reported building gross area. The developed SVM model was used as the data-driven surrogate model. Lastly, an equivalent SDOF was developed to derive low-order mechanistic models, where spring characteristics were calibrated based on the result of the pushover analysis of similar concrete frames in the literature.

The assumption made to develop the equivalent SDOF assumptions could impact loss prediction. Therefore, sensitivity assessments were performed to identify the critical decisions. Figure 11 compares different loss values obtained from equivalent SDOF and detailed assessment for three considered frames (denoted as frames 1,2, and 3). Overall, the difference between simplified and detailed dynamic models vary across different loss type. For example, while there is a significant difference between structural losses of SDOF and MDOF models for the second frame, the collapse losses are very close. Second, as can be observed by comparing Figures 11.a-c to 10.d-f, the inclusion of cyclic deterioration in the SDOF model did not notably change the loss prediction (collapse losses are more sensitive to this value). Lastly, the estimated loss values are highly sensitive to the factor relating SDOF displacement to MDOF drift. Currently, empirical values are suggested for this mapping [62]. However, the comparison between Figure 11.d-f and 11.g-i



Figure 11. Comparison of different loss types estimated from equivalent SDOF and detailed analysis. Figures a-c are calculated using a γ =0 and empirical factors to related SDOF displacement to MDOF drift, Figures d-f are caucluated using γ =50 and empirical factors, Figures g-i are calculated using γ =50 and calibrated factors.



Figure 12. KBS: knowledge-based, DDS-F: data-driven (full), DDS-R: data-driven (reduced), PBS: physics-based

suggests that even when this value is obtained directly from sensitivity assessment, the final loss values can be drastically different from the detailed assessment. In addition, empirical factors could result in a woefully inaccurate loss estimate for a given frame (e.g., frame 2). Lastly, the difference between losses obtained by different factors varies across frames and loss types, pointing to a more complex relationship.

The workflow was applied to estimate the loss for a 4-story building with five 27-foot bays in each direction. Here, the only available information is building footprint, location, and construction cost. Figure 12 shows the resultant loss ranges predicted by the workflow. The results of the surrogate models are compared to the 12 design variations assessed based on a detailed PBEE assessment. The knowledge-based surrogate models provide a larger range, mainly because the current performance inventories are governed by concrete frames designed for California that must resist larger seismic forces. The developed SVM used ranges of input variables for lateral weight and beam design information (from code minimum and maximum range), resulting in a loss range that encompasses the median and variance predicted by the detailed assessment. Lastly, the pushover results of a 4-story building from the literature were used to calibrate the SDOF model. Since the literature building was located in California, a scaled pushover curve was also used to represent the lower bound for the site in Charleston. As shown in Figure 12, the sequence approximately converges to the median loss results from a detailed assessment using incomplete data. The average prediction from data-driven models is 10.5% larger than the average loss predicted by detailed design. The average prediction from the low-order dynamic model is about 11.9% smaller than the detailed model.

4.4. ML-assisted detailed design

The results from the early design stage characterize the RC frame properties such as lateral weight, footprint, and average beam area. An experienced structural engineer can use this information to generate some preliminary designs. Next, the designer can perform pushover analysis on the generated alternatives and uses the *P2M* network to evaluate the seismic performance. As discussed in Section 3.2, *P2M* also provides an efficient means to change frame properties and map the changes to the final seismic performance measures. It should be noted that future work to create mapping functions between early design estimates and member sizing could sidestep this step, resulting in an end-to-end simulation framework.

The discussed network was trained on the compiled frame inventory to develop *P2M* for the current site. The results showed average prediction accuracy of 84% for a test-to-training ratio of unity (half of the database was used for testing) [39]. A sensitivity assessment on different neural network types suggested that LSTM networks can provide the highest accuracy with an average R2 of 87%, only 4.4% lower accuracy than the detailed analysis results. The average difference between P2M and the detailed analysis estimate for the slope and intercept of the PSDM model were 2.7% and 5.7%, respectively [39].

5. Conclusion

5.1. Summary

This study presented a conceptual framework to integrate ML methods into the PBEE framework for a holistic seismic design. The presented framework comprises two modules for early and detailed design. The early design module uses supervised ML algorithms to directly estimate seismic loss using geometric configuration and average design parameters estimable by an experienced designer. Combing these models with simplified mathematical relationships based on the knowledge base and low-order mechanistic models provided a means to quickly predict a performance range for a given frame without any detailed design information. After parametrizing the possible range for design parameters, finite element models could be developed for average (or minimum and maximum) values to perform a pushover analysis. The resultant pushover curve (and its variations for parametrized frame properties) can be fed into a deep-learning engine to estimate seismic demand models, and consequently, fragility and loss values, preventing the need for complicated time-history analysis. The same approach can be extended to other performance measures reliant on quantitative analysis (such as environmental impacts), leading to a holistic seismic design.

Several major findings of this study are as follows:

- ML methods can scale PBEE assessments over a building inventory and provide accurate predictions subjected to different levels of available information on topology and design.
- ML models can be trained to estimate earthquake-related repair costs using only crude topological and design parameters for a given site and structural system. ML model accuracy was sensitive to the building replacement cost values.
- Among studied algorithms, support vector machines and gradient boosting methods provide the highest accuracy. The prediction of support vector machines was dominated by higher-order interactions of predictors, where building height, floor area, and average beam section area were the most influential features in predicting concrete frame repair cost.
- Although simpler regression-based models could be adequate for structural loss predictions, they show high bias for collapse loss estimation.
- An equivalent SDOF system can provide reasonable accuracy to predict seismic loss of symmetric concrete frames. The accuracy of SDOF systems varies across different loss types, and is sensitive to the factors relating its displacement to the building inter-story drift.

• The "shape" of pushover curves successfully encodes salient dynamic properties of the frame systems, which can be decoded into probabilistic seismic demand models through a deep learning architecture. This model can propagate component-level uncertainties into fragility functions, allowing for a fast estimate of seismic loss for different design decisions.

5.2. Limitations

Despite the current framework advantages to extend PBEE to different design stages, several limitations need further research to improve frameworks automation and easy scaling as follows:

- The methodology required an automated seismic design module to provide codeconforming member sizing. In the current case study, a pseudo-directional sampling-based approach was used to generate member sizes, which does not meet the requirements of a careful seismic design and prevent an end-to-end framework.
- The framework is applied to a site with moderate seismicity. To ensure generalizability, testbeds for sites with high seismicity should be investigated. Nevertheless, the current literature indicates that ML models can estimate drift with high accuracy for sites with high seismicity, suggesting that applying the framework to those sites is possible.
- Although pushover analyses are relatively less computationally expensive, they require developing a finite element model. Future work is needed to create a mapping function between the properties of an equivalent SDOF and detailed FEM to sidestep this issue.
- The framework was applied to an inventory of only concrete frames. However, a more extensive and diverse inventory comprising different structural systems, is needed to evaluate the framework generalizability, particularly for system selection at early design.

References

- [1] Bozorgnia Y, Bertero VV. Earthquake Engineering: From Engineering Seismology to Performance-Based Engineering. CRC Press; 2004.
- [2] Moehle J, Deierlein G. A framework methodology for performance-based earthquake engineering 2004.
- [3] Zaker Esteghamati M, Farzampour A. Probabilistic seismic performance and loss evaluation of a multistory steel building equipped with butterfly-shaped fuses. J Constr Steel Res 2020;172:106187. https://doi.org/10.1016/j.jcsr.2020.106187.
- [4] Chen P, Pan J, Hu F, Wang Z. Numerical investigation on seismic resilient steel beam-tocolumn connections with replaceable buckling-restrained fuses. J Constr Steel Res 2022;199:107598. https://doi.org/10.1016/j.jcsr.2022.107598.
- [5] Zhu R, Guo T, Tesfamariam S. Seismic performance assessment of steel moment-resisting frames with self-centering viscous-hysteretic devices. J Constr Steel Res 2021;187:106987. https://doi.org/10.1016/j.jcsr.2021.106987.
- [6] Wang J, Men J, Zhang Q, Fan D, Zhang Z, Huang C-H. Seismic performance evaluation of a novel shape-optimized composite metallic yielding damper. Eng Struct 2022;268:114714. https://doi.org/10.1016/j.engstruct.2022.114714.
- [7] Muntasir Billah A, Rahman J, Zhang Q. Shape memory alloys (SMAs) for resilient bridges: A state-of-the-art review. Structures 2022;37:514–27. https://doi.org/10.1016/j.istruc.2022.01.034.
- [8] Hadad AA, Shahrooz BM, Fortney PJ. Innovative resilient steel braced frame with Belleville disk and shape memory alloy assemblies. Eng Struct 2021;237:112166. https://doi.org/10.1016/j.engstruct.2021.112166.
- [9] Das S, Tesfamariam S. Multi-objective design optimization of multi-outrigger tall-timber building: Using SMA-based damper and Lagrangian model. J Build Eng 2022;51:104358. https://doi.org/10.1016/j.jobe.2022.104358.
- [10] Tarfan S, Banazadeh M, Zaker Esteghamati M. Probabilistic seismic assessment of nonductile RC buildings retrofitted using pre-tensioned aramid fiber reinforced polymer belts. Compos Struct 2019;208:865–78. https://doi.org/10.1016/j.compstruct.2018.10.048.
- [11] Shahsahebi A, Waezi Z, Hashemi MJ. Seismic performance assessment of multistory RC buildings with soft-story collapse mechanism equipped with gapped inclined bracing (GIB). Structures 2020;28:2448–66. https://doi.org/10.1016/j.istruc.2020.10.068.
- [12] Kammouh O, Silvestri S, Palermo M, Cimellaro GP. Performance-based seismic design of multistory frame structures equipped with crescent-shaped brace. Struct Control Health Monit 2018;25:e2079. https://doi.org/10.1002/stc.2079.
- [13] Nobahar E, Farahi M, Mofid M. Quantification of seismic performance factors of the buildings consisting of disposable knee bracing frames. J Constr Steel Res 2016;124:132–41. https://doi.org/10.1016/j.jcsr.2016.05.007.
- [14] Faramarzi MS, Taghikhany T. A comparative performance-based seismic assessment of strongback steel braced frames. J Build Eng 2021;44:102983. https://doi.org/10.1016/j.jobe.2021.102983.
- [15] Simpson BG, Mahin SA. Experimental and numerical investigation of strongback braced frame system to mitigate weak story behavior. J Struct Eng 2018;144:04017211.
- [16] PEER. Guidelines for PerformanceBased Seismic Design of Tall Buildings. Pacific Earthquake Engineering Research Center; 2017.

- [17] Sun H, Burton HV, Huang H. Machine learning applications for building structural design and performance assessment: State-of-the-art review. J Build Eng 2021;33:101816. https://doi.org/10.1016/j.jobe.2020.101816.
- [18] Xie Y, Ebad Sichani M, Padgett JE, DesRoches R. The promise of implementing machine learning in earthquake engineering: A state-of-the-art review. Earthq Spectra 2020;36:1769– 801. https://doi.org/10.1177/8755293020919419.
- [19] Luo H, Paal SG. Data-driven seismic response prediction of structural components. Earthq Spectra 2022;38:1382–416. https://doi.org/10.1177/87552930211053345.
- [20] Feng D-C, Liu Z-T, Wang X-D, Jiang Z-M, Liang S-X. Failure mode classification and bearing capacity prediction for reinforced concrete columns based on ensemble machine learning algorithm. Adv Eng Inform 2020;45:101126. https://doi.org/10.1016/j.aei.2020.101126.
- [21] Cakiroglu C, Islam K, Bekdaş G, Isikdag U, Mangalathu S. Explainable machine learning models for predicting the axial compression capacity of concrete filled steel tubular columns. Constr Build Mater 2022;356:129227. https://doi.org/10.1016/j.conbuildmat.2022.129227.
- [22] Vu Q-V, Truong V-H, Thai H-T. Machine learning-based prediction of CFST columns using gradient tree boosting algorithm. Compos Struct 2021;259:113505. https://doi.org/10.1016/j.compstruct.2020.113505.
- [23] Siam A, Ezzeldin M, El-Dakhakhni W. Machine learning algorithms for structural performance classifications and predictions: Application to reinforced masonry shear walls. Structures 2019;22:252–65. https://doi.org/10.1016/j.istruc.2019.06.017.
- [24] Feng D-C, Wang W-J, Mangalathu S, Taciroglu E. Interpretable XGBoost-SHAP Machine-Learning Model for Shear Strength Prediction of Squat RC Walls. J Struct Eng 2021;147:04021173. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003115.
- [25] Zhang H, Cheng X, Li Y, Du X. Prediction of failure modes, strength, and deformation capacity of RC shear walls through machine learning. J Build Eng 2022;50:104145. https://doi.org/10.1016/j.jobe.2022.104145.
- [26] Ahmed B, Mangalathu S, Jeon J-S. Seismic damage state predictions of reinforced concrete structures using stacked long short-term memory neural networks. J Build Eng 2022;46:103737. https://doi.org/10.1016/j.jobe.2021.103737.
- [27] Luo H, Paal SG. Artificial intelligence-enhanced seismic response prediction of reinforced concrete frames. Adv Eng Inform 2022;52:101568. https://doi.org/10.1016/j.aei.2022.101568.
- [28] Nguyen HD, LaFave JM, Lee Y-J, Shin M. Rapid seismic damage-state assessment of steel moment frames using machine learning. Eng Struct 2022;252:113737. https://doi.org/10.1016/j.engstruct.2021.113737.
- [29] Guan X, Burton H, Shokrabadi M, Yi Z. Seismic Drift Demand Estimation for Steel Moment Frame Buildings: From Mechanics-Based to Data-Driven Models. J Struct Eng 2021;147:04021058. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003004.
- [30] Hwang S-H, Mangalathu S, Shin J, Jeon J-S. Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames. J Build Eng 2021;34:101905. https://doi.org/10.1016/j.jobe.2020.101905.
- [31] Demertzis K, Kostinakis K, Morfidis K, Iliadis L. An interpretable machine learning method for the prediction of R/C buildings' seismic response. J Build Eng 2023;63:105493. https://doi.org/10.1016/j.jobe.2022.105493.

- [32] Kazemi F, Asgarkhani N, Jankowski R. Predicting seismic response of SMRFs founded on different soil types using machine learning techniques. Eng Struct 2023;274:114953. https://doi.org/10.1016/j.engstruct.2022.114953.
- [33] Dabiri H, Faramarzi A, Dall'Asta A, Tondi E, Micozzi F. A machine learning-based analysis for predicting fragility curve parameters of buildings. J Build Eng 2022;62:105367. https://doi.org/10.1016/j.jobe.2022.105367.
- [34] Kourehpaz P, Molina Hutt C. Machine Learning for Enhanced Regional Seismic Risk Assessments. J Struct Eng 2022;148:04022126. https://doi.org/10.1061/(ASCE)ST.1943-541X.0003421.
- [35] Lu X, McKenna F, Cheng Q, Xu Z, Zeng X, Mahin SA. An open-source framework for regional earthquake loss estimation using the city-scale nonlinear time history analysis. Earthq Spectra 2020;36:806–31. https://doi.org/10.1177/8755293019891724.
- [36] Deierlein GG, McKenna F, Zsarnóczay A, Kijewski-Correa T, Kareem A, Elhaddad W, et al. A Cloud-Enabled Application Framework for Simulating Regional-Scale Impacts of Natural Hazards on the Built Environment. Front Built Environ 2020;6.
- [37] Rezaee R, Brown J, Augenbroe G, Kim J. Assessment of uncertainty and confidence in building design exploration. AI EDAM 2015;29:429–41. https://doi.org/10.1017/S0890060415000426.
- [38] Arnott D. Cognitive biases and decision support systems development: a design science approach. Inf Syst J 2006;16:55–78. https://doi.org/10.1111/j.1365-2575.2006.00208.x.
- [39] Soleimani-Babakamali MH, Zaker Esteghamati M. Estimating seismic demand models of a building inventory from nonlinear static analysis using deep learning methods. Eng Struct 2022;266:114576. https://doi.org/10.1016/j.engstruct.2022.114576.
- [40] Yepes-Estrada C, Silva V, Rossetto T, D'Ayala D, Ioannou I, Meslem A, et al. The Global Earthquake Model Physical Vulnerability Database. Earthq Spectra 2016;32:2567–85. https://doi.org/10.1193/011816EQS015DP.
- [41] Flint MM, Zaker Esteghamati M, Lee J, Musetich M, Sharifi Mood M. Inventory of Seismic Structural Evaluations, Performance Functions and Taxonomies for Buildings (INSSEPT) 2019.
- [42] Zaker Esteghamati M, Lee J, Musetich M, Flint MM. INSSEPT: An open-source relational database of seismic performance estimation to aid with early design of buildings. Earthq Spectra 2020;36:2177–97. https://doi.org/10.1177/8755293020919857.
- [43] Zaker Esteghamati M, Sharifnia H, Ton D, Asiatco P, Reichard G, Flint MM. Life Cycle Assessment Data & Energy Models of Archetype Commercial Buildings Located in the Central and Eastern United States 2021.
- [44] Omoya M, Ero I, Zaker Esteghamati M, Burton HV, Brandenberg S, Sun H, et al. A relational database to support post-earthquake building damage and recovery assessment. Earthq Spectra 2022;38:1549–69. https://doi.org/10.1177/87552930211061167.
- [45] Kang H, Burton HV, Miao H. Replicating the Recovery following the 2014 South Napa Earthquake using Stochastic Process Models. Earthq Spectra 2018;34:1247–66. https://doi.org/10.1193/012917EQS020M.
- [46] Cook DT, Liel AB, Haselton CB, Koliou M. A framework for operationalizing the assessment of post-earthquake functional recovery of buildings. Earthq Spectra 2022;38:1972–2007. https://doi.org/10.1177/87552930221081538.

- [47] Zaker Esteghamati M, Flint MM. Developing data-driven surrogate models for holistic performance-based assessment of mid-rise RC frame buildings at early design. Eng Struct 2021;245:112971. https://doi.org/10.1016/j.engstruct.2021.112971.
- [48] McKenna F. OpenSees: A Framework for Earthquake Engineering Simulation. Comput Sci Eng 2011;13:58–66. https://doi.org/10.1109/MCSE.2011.66.
- [49] Ibarra LF, Medina RA, Krawinkler H. Hysteretic models that incorporate strength and stiffness deterioration. Earthq Eng Struct Dyn 2005;34:1489–511. https://doi.org/10.1002/eqe.495.
- [50] Haselton CB, Pacific Earthquake Engineering Research Center. Beam-column element model calibrated for predicting flexural response leading to global collapse of RC frame buildings. Pacific Earthquake Engineering Research Center; 2008.
- [51] Zareian F, Medina RA. A practical method for proper modeling of structural damping in inelastic plane structural systems. Comput Struct 2010;88:45–53. https://doi.org/10.1016/j.compstruc.2009.08.001.
- [52] Zaker Esteghamati M, Banazadeh M, Huang Q. The effect of design drift limit on the seismic performance of RC dual high-rise buildings. Struct Des Tall Spec Build 2018;27:e1464.
- [53] Chapman MC, Talwani P. Seismic Hazard Mapping for Bridge and Highway Design in South Carolina. 2006.
- [54] Zaker Esteghamati M, Bahrampouri M, Rodriguez-Marek A. The Impact of Hazard-Consistent Ground Motion Scenarios Selection on Structural Seismic Risk Estimation 2021:347–56. https://doi.org/10.1061/9780784483695.034.
- [55] Boore DM. Simulation of Ground Motion Using the Stochastic Method. Pure Appl Geophys 2003;160:635–76. https://doi.org/10.1007/PL00012553.
- [56] Zaker Esteghamati M, Huang Q. An efficient stratified-based ground motion selection for cloud analysis, 2019. https://doi.org/10.22725/ICASP13.399.
- [57] Zaker Esteghamati M. A Holistic Review of GM/IM Selection Methods from a Structural
Performance-Based Perspective.Sustainability2022;14:12994.https://doi.org/10.3390/su142012994.
- [58] Ramirez CM, Liel AB, Mitrani-Reiser J, Haselton CB, Spear AD, Steiner J, et al. Expected earthquake damage and repair costs in reinforced concrete frame buildings. Earthq Eng Struct Dyn 2012;41:1455–75. https://doi.org/10.1002/eqe.2216.
- [59] Lignos D, Krawinkler H. Sidesway Collapse of Deteriorating Structural Systems Under Seismic Excitations. 2012.
- [60] FEMA. Multi-hazard Loss Estimation Methodology Earthquake Model Technical Manual (Hazus–MH 2.1). 2011.
- [61] Zaker Esteghamati M, Sharifnia H, Ton D, Asiatico P, Reichard G, Flint MM. Sustainable early design exploration of mid-rise office buildings with different subsystems using comparative life cycle assessment. J Build Eng 2022;48:104004. https://doi.org/10.1016/j.jobe.2022.104004.
- [62] Krawinkler H, Zareian F, Medina RA, Ibarra LF. Decision support for conceptual performance-based design. Earthq Eng Struct Dyn 2006;35:115–33. https://doi.org/10.1002/eqe.536.