11th International Operational Modal Analysis Conference
20 - 23 May 2025 | Rennes, France

A Deep Learning Model for Cross-Building Seismic Response Sequence Reconstruction

Hongyu HUANG[1], *Yuxin PAN*[2], *and Teng LI*[3]

[1] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong China, hhuangca@connect.ust.hk
[2] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong China, ceypan@ust.hk
[3] School of Control Science and Engineering, Shandong University, Shandong, China, li.teng@sdu.edu.cn

ABSTRACT

Seismic structural health monitoring (SHM) system typically covers a limited number of buildings in a region to measure their dynamic responses, while the responses of the other non-instrumented buildings remain unknown during seismic events. It is widely used to predict these responses by nonlinear time history analysis (NLTHA). In recent years, deep learning (DL) models have emerged as efficient surrogate models for NLTHA due to their remarkable accuracy and efficiency. However, models like long-short-term memory (LSTM) are limited in predicting seismic response sequences for individual buildings, hindering comprehensive regional assessments. To address this, we propose a novel DL model based on LSTM to reconstruct seismic response sequences across multiple buildings within the same cluster. The inputs of the proposed model consist of ground motion acceleration, seismic responses of instrumented indicator buildings, and building attributes of non-instrumented target buildings, with outputs providing the seismic response of these target buildings. The feasibility of the proposed model will be demonstrated in a hypothetical region with 31 high-rise reinforced concrete shear walls (1 indicator building and 30 target buildings). The building attributes include the number of stories and floor area. Each building will be simplified as a multi-degree of freedom (MDOF) shear model with a pinching hysteretic model to characterize the inter-story behavior. NLTHA will be conducted under a total of 65 ground motion records to obtain response sequences to generate training data. The results show that the proposed DL model achieves 93.5% accuracy when predicting roof acceleration response sequences of 30 non-instrumented buildings under 14 testing seismic waves. One prediction can be completed in 0.05 ms, capable of real-time application. Overall, the proposed DL model has the potential to enhance regional seismic risk assessment with a limited SHM system.

Keywords: Regional seismic risk, Response reconstruction, Nonlinear time history analysis, Deep learning, Building attribute

1. INTRODUCTION

In a regional area, a seismic structural health monitoring (SHM) system can alarm residents of dangerous zones early and guide engineers to conduct damage assessments and post-earthquake repairs by acquiring real-time seismic responses of buildings. However, the SHM system typically only covers limited buildings in a region; buildings without such a system are also essential to accurately assess seismic risk in that region. As a result, many scholars have probed into reconstructing seismic responses of buildings without monitoring. The non-instrumented buildings are often defined as target buildings, while buildings installed with sensors can be identified as indicator buildings because they are representative ones of the same building cluster [1].

Several studies have reconstructed seismic responses of target buildings at a regional scale, but they focused on single metric responses rather than the whole sequences in the time domain. Sun et al. (2017) [2] first proposed to obtain peak floor accelerations (PFAs) and peak story drift ratios (PSDRs) of a single location within target buildings using interpolation based on the structural and spatial correlations between them and indicator buildings. Then, Sun et al. (2019) [3] extended it to a full-profile reconstruction, and Sun et al. (2022) [4] demonstrated the viability of reconstructing responses of diverse building portfolios under future earthquakes. However, they were all limited to elastic responses. Xu et al. (2023) [5] facilitated regional-scale nonlinear seismic response prediction by converting it to a matrix completion task, providing potential for a larger region but only involving a single structure. Some research also considered the nonlinearity of buildings but only reconstructed seismic damage states, which was a less complicated classification task than response reconstruction [6][7].

Meanwhile, many scholars have investigated the prediction performance of the seismic response sequences by conducting nonlinear time-history analysis (NLTHA). NLTHA is a relatively reliable method to obtain comprehensive seismic responses of structures without monitoring. However, dealing with numerous buildings or complex structures requires tremendous computing time and resources. Establishing a deep learning (DL) model as a surrogate for NLTHA has recently become a trend due to its remarkable accuracy and much less computational demand [8][9], and more importantly, it can achieve real-time prediction. However, these studies mainly focused on predicting the responses within a single building. The application of DL models for reconstructing regional response sequences is still limited.

In this regard, this research proposes a novel DL model fusing static (building attributes) and dynamic features (seismic waves and seismic responses of indicator building) to realize real-time reconstruction of seismic response sequences of target buildings. The proposed model is expected to reconstruct complete seismic response sequences instead of several metrics. Moreover, seismic responses of target buildings include linearity and nonlinearity.

The main contents of this study are summarized as follows. Section 2 introduces the methodology and architecture of the DL model. Section 3 prepares data of a hypothetical region for a case study, including the selection of buildings, earthquakes, and the modeling method suitable for buildings at a regional scale. Section 4 presents the results of this case and illustrates the model performance.

2. METHODOLOGY

This section proposes a DL model that can reconstruct the responses of non-instrumented buildings at a regional scale. Since the model will fuse different data types, this section will also clarify the data preprocessing.

2.1. Model architecture

In general, the seismic responses of a building are affected by ground motions (GMs) input and its building attributes (e.g., story height, floor area, period, etc.). However, this research incorporates seismic responses of indicator buildings as a reference since more attention will be paid to reconstruct responses of buildings in the same region instead of an individual building prediction. Accordingly, the inputs of the proposed DL model consist of target building attributes, GMs, and seismic responses of indicator buildings.

Building attributes are classified as static features since building natural properties remain constant over time. In contrast, GMs and responses of indicator buildings are dynamic time series data. The proposed model accounts for both static and dynamic features simultaneously. Given the different characteristics of static and dynamic features, for example, the significantly greater length of time series sequence compared to static building attributes in a single sample, the proposed DL model is divided into two parts to process static and dynamic inputs respectively and subsequently merge them to produce the final output, as illustrated in the right part of Fig. 1.

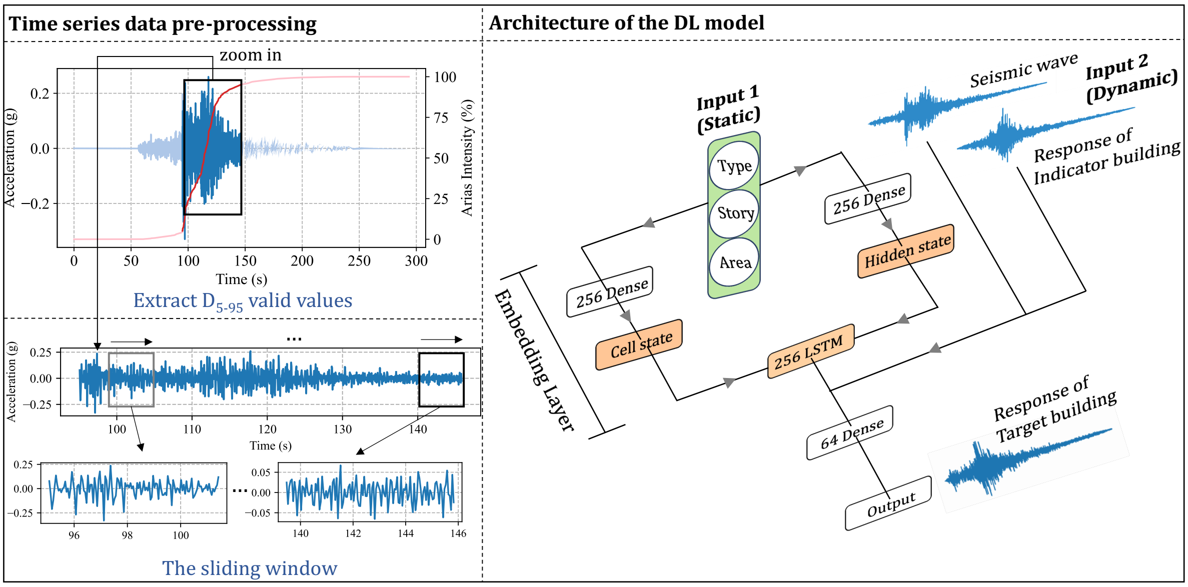


Figure 1. The architecture of the proposed DL model and time series data preprocessing.

The basic idea of dealing with Input 1 building attributes derives from the model "temporal fusion transformer (TFT)" [10]. An embedding block formed by two fully connected (FC) layers follows the input layer only of Input 1 because it can arbitrarily change the length of Input 1 to any value equivalent to the neuron number in FC layers by transforming Input 1 to another space dimension. Long-short-term memory (LSTM) was chosen to deal with time series data Input 2 after trial and error.

Generally, LSTM revolves around cells and hidden states, whose roles are similar to building attributes among all inputs. Hence, considering embedded values as initial states of LSTM is a key principle in data fusion, which also bridges Inputs 1 and 2. Finally, the embedded values of Input 1 are taken as cell states and hidden states of LSTM for Input 2 in the proposed model.

2.2. Time series data preprocessing

In this research, time series data needs to be preprocessed due to its inclusion of invalid values and varying length, including GMs and seismic responses of buildings. As shown in the left part of Fig. 1, time series data preprocessing includes two steps: extracting valid slices and cropping them consistently.

In a raw seismic wave, the starting and ending parts always fluctuate around zero, which is valueless to the model due to the similarity between values. As a result, partial contents in these two parts need to be trimmed. Significant duration ($D_{5\text{-}95}$) is a popular index evaluating the effectiveness of a seismic wave, and it refers to the duration reaching 5%-95% of the total accumulative arias intensity (AI) in a seismic wave [11]. It was adopted to extract meaningful portions in the first step.

In theory, immediately taking each extracted GM in the first step as a sample is feasible, but it presents two challenges: uneven sequence inputs and insufficient samples. To address these issues, a sliding window approach is adopted in the second step. Meanwhile, ground motion with any frequency can be a candidate for input. The window size and stride were set as 128 and 10, respectively.

3. CASE STUDY: DATA PREPARATION

This section will detail the implementation of the proposed methodology. Due to the scarcity of published real databases involving building information, earthquakes, and corresponding seismic responses simultaneously, a well-studied region in Vancouver, in the southwest part of Canada, is chosen [12]. For illustration, a small-scale hypothetical area with multiple buildings is assumed. The seismic responses of all buildings in this region are obtained through NLTHA, and NLTHA results are assumed as ground truth in this research.

The generation of regional building configuration and the NLTHA modeling method suitable for these buildings will be introduced, along with the selection of GMs. Data splitting for model training and testing will also be clarified.

3.1. Regional buildings modeling

Given that regional buildings are often numerous but similar in structure type, a concise and uniform modeling method is required. Lu's modeling method designed for regional buildings based on HAZUS is adopted here [13][14]. Some details are modified, and a brief introduction to this method (Fig. 2) is as follows. Firstly, since comprehensive building information is impossible to obtain, each regional building is simplified as a multi-degree-of-freedom (MDOF) shear model instead of a detailed model. Then, the shear type is assumed to be the dominant deformation mode. Besides, inter-story behavior is regarded as hysteretic, which can simulate the nonlinearity of each story. The pinching model is adopted to characterize the hysteretic behaviors of structures.

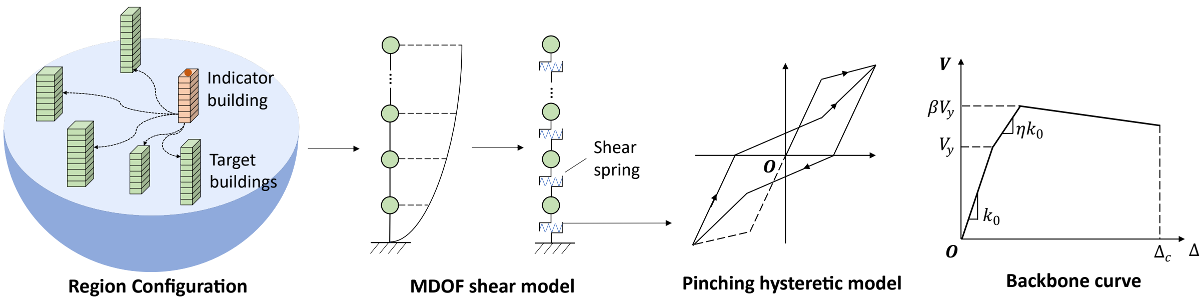


Figure 2. The building modeling method and configuration of the virtual area.

The pinching model adopts a trilinear backbone curve. Unlike Lu's method, as he defined the last curve as a horizontal straight line, the third stage of the curve is modified to simulate strength deterioration. The inter-story shear strength of the final point is defined as 80% of the peak one [15][16]. In this research, NLTHA was conducted in OpenSeesPy, an official Python library [17]. The seismic wave along only one direction is considered. The zero-length 2D defines elements whose material is uniaxial hysteretic. The classical Rayleigh damping is adopted here.

3.2. Building inventory information

Since 94% of high-rise buildings in Vancouver are concrete shear walls, this type of construction (corresponding to the 'C2H' type in HAZUS) is selected as the structure type in the hypothetical region to test model performance [18]. In total, 31 buildings are set for the building cluster, including 1 indicator building and 30 target buildings. Attributes of target buildings in this region are randomly generated based on statistical data. This research selects dozens of specific C2H buildings from a real database called 'building-footprints-2009' published by the City of Vancouver [19]. Since the adopted modeling method neglects the effects of flexural behavior, the upper bound for the number of stories is set to 12 in this study, while the lower bound for the number of stories is set as 8 according to HAZUS. This limitation may reduce the simulation accuracy for taller buildings [20]. The indicator building is assigned 12 stories with a floor area of 665 m^2, reflecting the average value of the target buildings.

3.3. GMs selection

The design spectrum here was assumed to be the one in the case of Vancouver with site class C. The ground motion dataset consists of three categories of earthquakes by source type, including crustal, subcrustal, and subduction earthquake sources. In total, 65 seismic waves were selected from Selection and Scaling of Ground Motions (S2GM) [21]. To further examine the capacity of the proposed DL model for nonlinear response, GMs were manually scaled up with a factor of 2 to induce buildings' nonlinear behaviors.

51 of 65 seismic waves participated in the model training, and the remaining 14 waves became testing samples. The same data distribution was guaranteed as much as possible to avoid a dataset shift between training and testing datasets [22]. In total, the training dataset consists of 1530 samples, while the testing one includes 420 samples.

4. CASE STUDY: RESULTS

The model results were formed into complete waves through concatenating predicted slices. Since adjacent slices have overlapping parts resulting from the shorter stride of the sliding window than its length, a special way of concatenation was adopted. First, slices from the same wave are aligned according to the original timesteps due to their different starting timesteps. Then, values from different slices but at the same timings are averaged as the final predicted values of the corresponding timing in the complete wave. After repeated trials, this process was proven to enhance prediction accuracy stably.

Given that the acceleration of GMs is known, the roof acceleration was chosen to represent the seismic response and maintain consistency. The coefficient of determination (R^2, Eq. (1)) and mean absolute error (MAE, Eq. (2)) between the model predictions and NLTHA results of testing samples were used to evaluate the model performance.

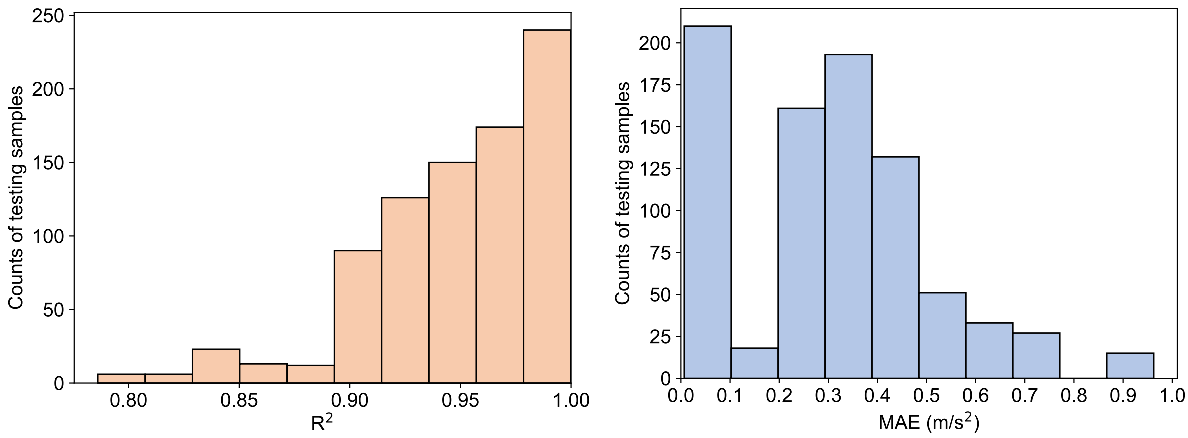
$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{1}$$

$$MAE(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

where y, $\hat{y}$, and $\bar{y}$ are truth, prediction, and the average value of truth respectively; n represents the number of samples.

Overall, the proposed model achieved outstanding accuracy, with 93.5% R^2 and 0.364 m/s^2 MAE values on 14 integrated testing waves. Fig. 3 shows the R^2 and MAE distributions of all testing samples. It highlights the number of testing samples within each accuracy interval. It could be observed that most samples fell within more than 90% R^2 and less than 0.5 m/s^2 MAE intervals, indicating a stable prediction performance of the model. For illustration, Fig. 4 compares predicted and ground truth acceleration responses for target building No. 5 under future earthquake No. 7. The model can capture the trend and amplitude of seismic response sequences for the entire duration with very little differences in terms of peak values.

In addition, the time consumption for predicting each slice with a length of 128 is around 0.05 ms. This demonstrates that the prediction of the trained model can be achieved in real time, showing significant advantages over traditional NLTHA but also maintaining excellent accuracy.



(a) R^2 distribution; (b) MAE distribution;

Figure 3. The R^2 and MAE distributions of all testing samples.

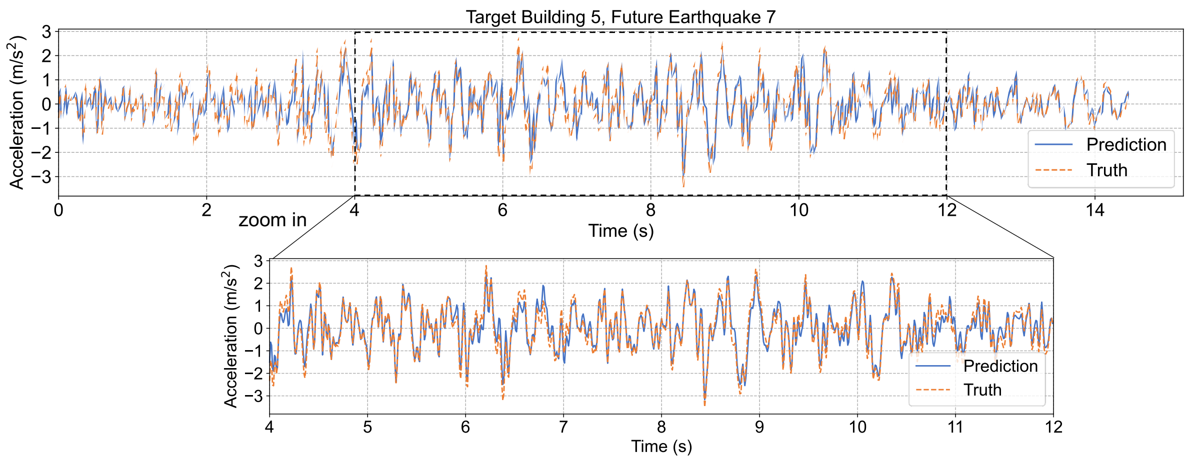


Figure 4. Comparison between predicted and true seismic response sequences.

5. CONCLUSIONS

This study proposed a novel DL model as an alternative way for NLTHA to reconstruct complete nonlinear seismic responses of non-instrumented buildings in a region. The proposed model designed a fusion of static (building attributes) and dynamic (GMs and seismic responses of buildings with sensors) inputs to surrogate the conventional NLTHA with superior and real-time performance. To validate the feasibility of the model, a case study was implemented in a hypothetical region with 31 reinforced concrete shear wall ('C2H') buildings in Vancouver, subjected to 65 seismic waves. The model performance of reconstructing roof accelerations was tested for 30 non-instrumented buildings under 14 seismic waves, while 51 waves were utilized for the model training.

The proposed model can rapidly reconstruct nonlinear seismic response sequences of non-instrumented buildings in a region with high accuracy. The DL model obtained 93.5% R^2 and 0.364 m/s^2 MAE value when predicting roof accelerations of 420 testing samples, including diverse building attributes and earthquakes, with the majority of results exceeding 90%. An excellent prediction of curve trends was observed. The error was mainly derived from the missed caption of peak values.

It is worth mentioning that this study has not considered the variability of seismic wave inputs at different locations. To have a more comprehensive analysis and assessment, such work is undergoing in our research group.

ACKNOWLEDGEMENTS

The funding support from the National Natural Science Foundation of China (Grant No. 52308280) for this research is greatly acknowledged.

REFERENCES

[1] A. Ghasemi, M. T. Stephens, and K. J. Elwood. Regional seismic response assessment using indicator buildings. *Bulletin of Earthquake Engineering*, 22(5), 2405-2431, 2024.

[2] H. Sun, H. Burton, Y. Zhang, and J. Wallace. Interbuilding interpolation of peak seismic response using spatially correlated demand parameters. *Earthquake Engineering & Structural Dynamics*, 47(5), 1148-1168, 2017.

[3] H. Sun, H. Burton, and J. Wallace. Reconstructing seismic response demands across multiple tall buildings using kernel-based machine learning methods. *Structural Control and Health Monitoring*, 26(7), e2359, 2019.

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

[5] Z. Xu, J. Chen, J. Shen, and M. Xiang. Regional-scale nonlinear structural seismic response prediction by neural network. *Engineering Failure Analysis*, 154, 107707, 2023.

[6] P. Kourehpaz, and C. M. Hutt. Machine learning for enhanced regional seismic risk assessments. *Journal of Structural Engineering*, 148(9), 04022126, 2022.

[7] S. Bhatta, and J. Dang. Machine learning-based classification for rapid seismic damage assessment of buildings at a regional scale. *Journal of Earthquake Engineering*, 28(7), 1861-1891, 2023.

[8] Y. Liao, R. Lin, R. Zhang, and G. Wu. Attention-based LSTM (AttLSTM) neural network for seismic response modeling of bridges. *Computers & Structures*, 275, 106915, 2023.

[9] C. Ning, Y. Xie, and L. Sun. LSTM, WaveNet, and 2D CNN for nonlinear time history prediction of seismic responses. *Engineering Structures*, 286, 116083, 2023.

[10] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748-1764, 2021.

[11] T. Li, Y. Pan, K. Tong, C. E. Ventura, and C. W. Silva. Attention-based sequence-to-sequence learning for online structural response forecasting under seismic excitation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52(4), 2184-2200, 2021.

[12] C. Ventura and L. Finn. A study of seismic risk in southwestern British Columbia. In *9th US National and 10th Canadian Conference on Earthquake Engineering*, Toronto, Canada, 2010.

[13] X. Lu, B. Han, M. Hori, C. Xiong, and Z. Xu. A coarse-grained parallel approach for seismic damage simulations of urban areas based on refined models and GPU/CPU cooperative computing. *Advances in Engineering Software*, 70, 90-103, 2014.

[14] X. Lu, F. McKenna, Q. Cheng, Z. Xu, X. Zeng, and S. A. Mahin. An open-source framework for regional earthquake loss estimation using the city-scale nonlinear time history analysis. *Earthquake Spectra*, 36(2), 806-831, 2020.

[15] H. Elgamel, M. K. Ismail, A. Ashour, and W. El-Dakhakhni. Backbone model for reinforced concrete block shear wall components and systems using controlled multigene genetic programming. *Engineering Structures*, 274, 115173, 2023.

[16] G. Ma, Y. Wang, and H. J. Hwang. Genetic programming-based backbone curve model of reinforced concrete walls. *Engineering Structures*, 283, 115824, 2023.

[17] M. Zhu, F. McKenna, and M. H. Scott. OpenSeesPy: Python library for the OpenSees finite element framework. *SoftwareX*, 7, 6-11, 2018.

[18] C. E. Ventura, W. L. Finn, T. Onur, A. Blanquera and M. Rezai. Regional seismic risk in British Columbia—classification of buildings and development of damage probability functions. *Canadian Journal of Civil Engineering*, 32(2), 372-387, 2005.

[19] City of Vancouver. Building - footprints - 2009. https://opendata.vancouver.ca/explore/dataset/building-footprints-2009/.

[20] C. Xiong, X. Lu, H. Guan, and Z. Xu. A nonlinear computational model for regional seismic simulation of tall buildings. *Bulletin of Earthquake Engineering*, 14, 1047-1069, 2016.

[21] R. Tremblay, G. M. Atkinson, N. Bouaanani, P. Daneshvar, P. Léger, and S. Koboevic. Selection and scaling of ground motion time histories for seismic analysis using NBCC 2015. In *11th Canadian Conference on Earthquake Engineering*, Victoria, Canada, 2015.

[22] J. G. Moreno-Torres, T. Raeder, R. Alaiz-Rodríguez, N. V. Chawla, and F. Herrera. A unifying view on dataset shift in classification. *Pattern recognition*, 45(1), 521-530, 2012.