

Article

Training Data Selection for Machine Learning-Enhanced Monte Carlo Simulations in Structural Dynamics

Denny Thaler *, Leonard Elezaj, Franz Bamer * and Bernd Markert 

Institute of General Mechanics, RWTH Aachen University, 52062 Aachen, Germany;
leonard.elezaj@rwth-aachen.de (L.E.); markert@iam.rwth-aachen.de (B.M.)

* Correspondence: thaler@iam.rwth-aachen.de (D.T.); bamer@iam.rwth-aachen.de (F.B.)

Abstract: The evaluation of structural response constitutes a fundamental task in the design of ground-excited structures. In this context, the Monte Carlo simulation is a powerful tool to estimate the response statistics of nonlinear systems, which cannot be represented analytically. Unfortunately, the number of samples which is required for estimations with high confidence increases disproportionately to obtain a reliable estimation of low-probability events. As a consequence, the Monte Carlo simulation becomes a non-realizable task from a computational perspective. We show that the application of machine learning algorithms significantly lowers the computational burden of the Monte Carlo method. We use artificial neural networks to predict structural response behavior using supervised learning. However, one shortcoming of supervised learning is the inability of a sufficiently accurate prediction when extrapolating to data the neural network has not seen yet. In this paper, neural networks predict the response of structures subjected to non-stationary ground excitations. In doing so, we propose a novel selection process for the training data to provide the required samples to reliably predict rare events. We, finally, prove that the new strategy results in a significant improvement of the prediction of the response statistics in the tail end of the distribution.

Keywords: machine learning; training data selection; Monte Carlo simulation; artificial neural networks; supervised learning; earthquake generation; Kanai–Tajimi filter; probability of failure; nonlinear structural dynamics



Citation: Thaler, D.; Elezaj, L.; Bamer, F.; Markert, B. Training Data Selection for Machine Learning-Enhanced Monte Carlo Simulations in Structural Dynamics. *Appl. Sci.* **2022**, *12*, 581. <https://doi.org/10.3390/app12020581>

Academic Editors: Konstantinos Morfidis and Konstantinos Kostinakis

Received: 20 October 2021
Accepted: 9 December 2021
Published: 7 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As engineers, we are responsible for the reliable design of structures and infrastructures. This task turns out to be challenging, especially when structures are supposed to withstand natural hazards, such as earthquakes. A powerful tool to evaluate the response statistics of structures provides the statistical investigation of a set of experiments, commonly known as the Monte Carlo method.

Structural failure should occur very rarely in order to save infrastructures and, consequently, human life so that the number of samples of the crude Monte Carlo method must be chosen high. Generally, experimental setups on full-scale structures are the most reliable way to collect data of the response behavior [1–3]. However, on the one hand, measured earthquake acceleration data are limited and, on the other hand, experimental setups of such a high number of samples are not realizable. One strategy to overcome this problem is the realization of hybrid simulations [4]. Although this approach significantly decreases the costs, it is still inefficient to collect enough data for reliable Monte Carlo predictions of low failure probabilities.

In order to obtain a practically realizable framework, purely numerical models are used to simulate the statistical response behavior of structures. Especially when inducing structural failure, numerical simulations are much easier realizable in an engineer's every-day life environment. Nevertheless, for complex structures modeled by high-dimensional finite element models, the computational cost of the crude Monte Carlo simulation becomes excessive.

There are two major options to deal with this issue:

- Methods were developed to reduce the number of the required samples employing advanced Monte Carlo strategies to estimate the probability of failure, such as importance and asymptotic sampling [5,6] or subset simulation techniques [7].
- Methods were developed to reduce the effort of each simulation, in this regard, model order reduction techniques have shown significant improvement of the computational burden [8–10]. Furthermore, the application of machine learning has received attention in the field of structural mechanics and showed promising results [11–14].

Several studies have used neural networks for the design in earthquake engineering [15] and for the investigation of hazard events [16]. Researchers used machine learning approaches for structural response prediction, health monitoring, post-earthquake assessment, and interpretation of experimental data [17–25]. Special focus has been laid on the design of the structures to estimate the response or classify the damage of reinforced concrete frames and masonry shear walls [26,27]. Furthermore, in-plane failure modes have been investigated using machine learning algorithms [28]. With particular emphasis on the estimation of the tail end probability, i.e., the occurrence of very extreme events, extended data sets were used to predict the probability of failure using various neural network architectures [29,30]. To this end, the earthquake samples for the training sets of the neural network were generated using random factorization. This leads to sample sets with higher variance and, therefore, a larger portion of extreme events. However, these strategies need the consideration of additional extended data sets for the training procedure to obtain a suitable neural network prediction in case of very rare events—a strategy that requires a careful choice of additional intensity factors when generating earthquake training sets.

In order to deal with this issue, this paper proposes a training data selection approach that enables a reliable prediction within the entire domain without extending the training data set. The proposed training data selection process is demonstrated on three different structures using two-, three-, and six-dimensional selection processes. This strategy improves the prediction of the response statistics in the case of rare events, i.e., in the tail region of the distribution and turns out to result in more stable response distributions.

2. Materials and Methods

2.1. Estimation of Failure Probability

To determine whether structural failure occurs, one must define a limit state based on engineering decisions. Using a limit state function $g(\mathbf{x})$, failure occurs if this function is below or equal to zero, i.e., $g(\mathbf{x}) \leq 0$. The probability of failure can be evaluated in terms of a multi-dimensional integral of the probability density function $f_X(\mathbf{x})$ [5]:

$$P_f = \int \cdots \int_{g(\mathbf{x}) \leq 0} f_X(\mathbf{x}) \, dx_1 \dots dx_n . \tag{1}$$

Introducing an indicator function $I_g(x_1, \dots, x_n)$, which is 1, if $g(x_1, \dots, x_n) \leq 0$, and 0 otherwise, Equation (1) can be reformulated as:

$$p_f = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} I_g(x_1, \dots, x_n) f_{X_1, \dots, X_n}(x_1, \dots, x_n) \, dx_1 \dots dx_n . \tag{2}$$

Generating a set of random samples and evaluating the corresponding indicator functions $I_g(\mathbf{x}^{(k)})$, the Monte Carlo method allows one to define the probability of failure in terms of an expected value:

$$p_f = \frac{1}{m} \sum_{k=1}^m I_g(\mathbf{x}^{(k)}) . \tag{3}$$

This approach works well for response estimations around the mean of the distribution, however, the general drawback is the low confidence of this estimator around the tail end of the distribution, i.e., the estimation of low failure probabilities:

$$\sigma_{pf}^2 = \frac{p_f}{m} - \frac{p_f^2}{m} \approx \frac{p_f}{m} \rightarrow \sigma_{pf} = \sqrt{\frac{p_f}{m}} . \tag{4}$$

Therefore, a disproportionately high number of samples must be evaluated in order to estimate small probabilities of failure with high confidence.

2.2. Artificial Non-Stationary Seismic Excitation Using the Kanai–Tajimi Filter

During the design process, earthquake excitations are employed to perform nonlinear dynamic time history analyses. In this context, the Kanai–Tajimi [31,32] filter is used to model the movement of the ground in terms of a single degree of freedom oscillator, which is subjected to a random white noise excitation. In doing so, site dependencies are included by using the natural frequency and the damping ratio of the ground. However, the conventional Kanai–Tajimi filter models earthquakes as stationary random processes and does not account for temporal changes in the properties of the excitation history. To this end, non-stationary Kanai–Tajimi models have been proposed to create more realistic excitations [33,34]. In this way, the time-dependent frequency content $\omega_g(t)$ and the time-dependent intensity parameter $e(t)$ are extracted from a measured benchmark record.

We adopt this procedure to provide many realistic and site-dependent artificial ground motions based on one recorded accelerogram chosen from the region around the building site. Exemplary, we use the Northridge earthquake accelerogram, recorded in California in 1994. In particular, the chosen time history is the 360° component from a Station of USGS (Sepulveda VA Hospital, CA: VA Hospital Bldg 40) on the ground level [35].

A time window that moves from t_0 to t_{end} of the accelerogram is defined to capture the time-varying features, as shown in Figure 1.

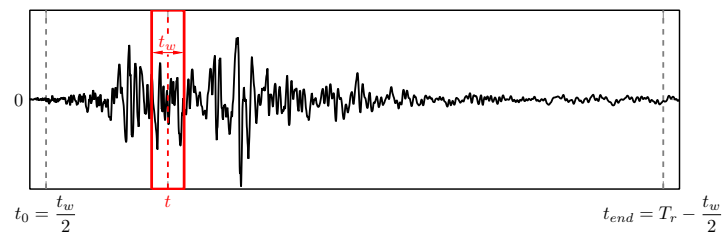


Figure 1. Parameter extraction from a record of the Northridge earthquake in 1994 [35] using a moving time window.

In this study, we chose a time window of size $t_w = 1.5$ s, which results in a smooth extraction of the parameters, while covering the frequency oscillations. Decreasing the time windows increases the deviation from window to window. By contrast, large time windows flatten the curves of the extracted parameters. The intensity of the acceleration at time t within the time-window $t - \frac{t_w}{2} < t < t + \frac{t_w}{2}$ is calculated by:

$$\hat{e}(t) = \int_{t - \frac{t_w}{2}}^{t + \frac{t_w}{2}} \left(\ddot{x}_g^{(r)} \right)^2 dt . \tag{5}$$

By counting the number of zero crossings in the interval $[t - \frac{t_w}{2}, t + \frac{t_w}{2}]$, the time dependent frequency at time t is evaluated:

$$\hat{\omega}_g(t) = \frac{\hat{n}(t)}{t_w} 2\pi . \tag{6}$$

We fit polynomial functions to the extracted properties, as shown in Figure 2. Therefore, next to the size of the scanning time, the order of the polynomial function must be chosen carefully to subsequently generate the nonstationary excitation. Next to these time-varying parameters, other ground properties must be chosen, such as damping of the ground. This parameter may be found through soil tests or from trial and error methods. We choose a damping ratio of $\zeta_g = 0.3$ based on numerical fitting methods [36,37] and visual inspection.

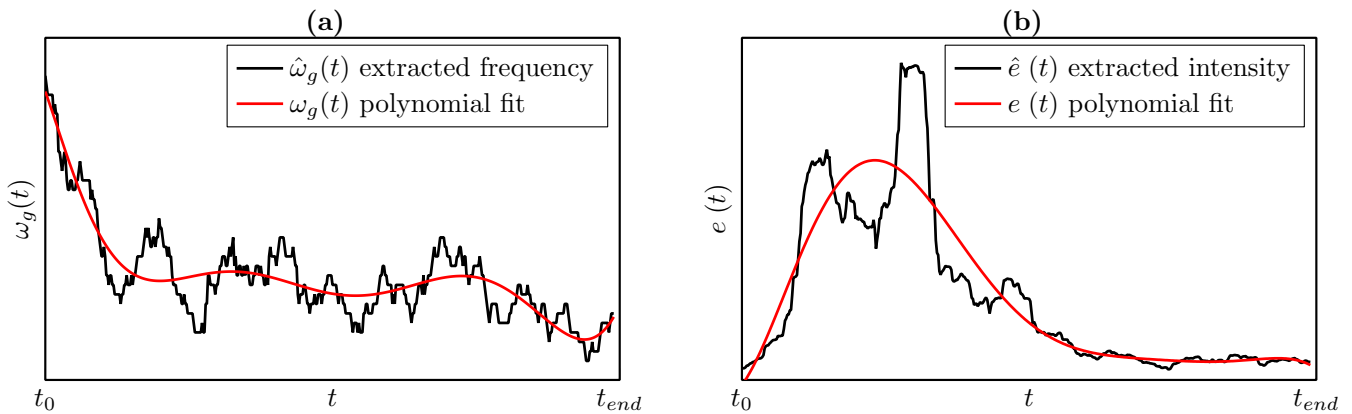


Figure 2. Extracted time-dependent properties of the recorded Northridge accelerogram, shown in Figure 1: (a) intensity $e(t)$ extracted by the moving time window from the Northridge event; (b) frequency content $\omega_g(t)$ from counted zero-crossing extracted by the moving time window of the Northridge event. The polynomial fits of the extracted parameters are highlighted in red.

The nonstationary Kanai–Tajimi filter is represented as the solution to the following nonlinear single degree of freedom system [33]:

$$\dot{x}_f^{(k)} + 2\zeta_g\omega_g(t)\dot{x}_f^{(k)} + \omega_g^2(t)x_f^{(k)} = n(t). \tag{7}$$

To obtain the filter response, x_f and \dot{x}_f , the single degree of freedom system is solved by the explicit central difference integration scheme [38,39]. The filter acceleration is written as:

$$\ddot{x}_g^{(k)} = -2\zeta_g\omega_g(t)\dot{x}_f^{(k)} - \omega_g^2(t)x_f^{(k)}. \tag{8}$$

The extracted envelope function $e(t)$ magnifies this response based on the extracted intensity parameter over the excitation time history. Finally, one obtains the synthetic accelerogram that includes the essential physical properties of the chosen benchmark excitation:

$$\ddot{x}_g^{(k)} = \ddot{x}_g^{(k)} e(t). \tag{9}$$

Using a value of $S_0 = 1.8$ for the power spectral density, we generated 10^4 samples for the numerical demonstration of this study, which we will discuss in Section 3 more detailed. One real and one synthetic accelerogram are presented in Figure 3a,b, respectively. As can be observed from these figures, the artificial ground excitation preserves the essential properties of the benchmark accelerogram while revealing the required level of randomness for our Monte Carlo simulation approaches.

This procedure was repeated 10^4 times to generate the sample set used in the numerical experiments of this paper. The range of the acceleration response spectra of all generated ground excitation is shown in Figure 4. Furthermore, the 25 percentile, mean, and 75 percentile of the spectra are highlighted in this visualization. The acceleration response spectrum of the benchmark earthquake is also plotted and is embedded within the envelope, defined by the upper and lower bound of all generated spectra.

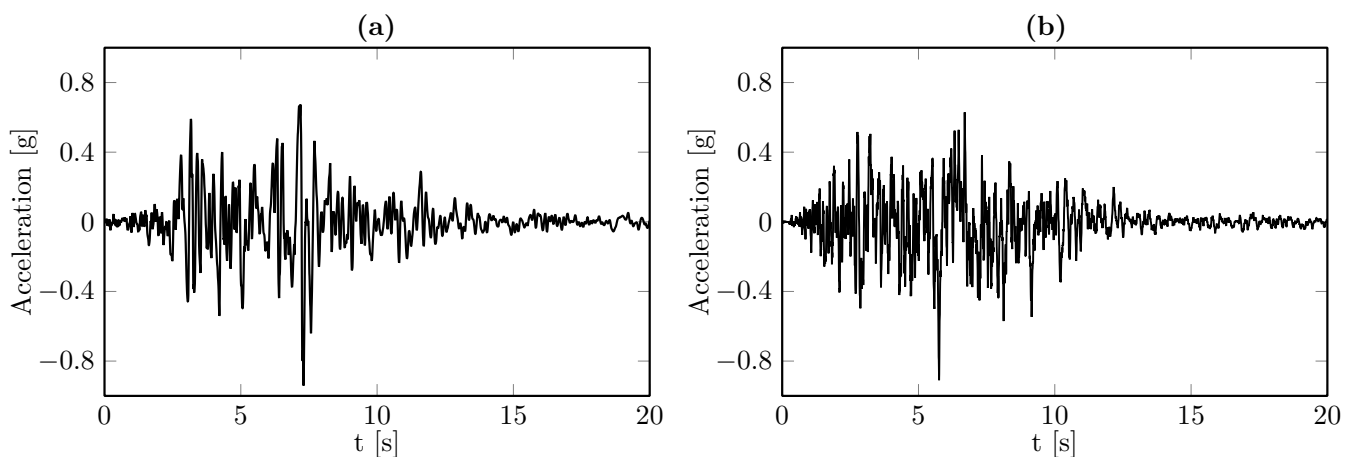


Figure 3. (a) Accelerogram of the Northridge earthquake [35] in 1994 and (b) time history of one sample of the synthetic generated accelerograms using a nonstationary Kanai–Tajimi filter.

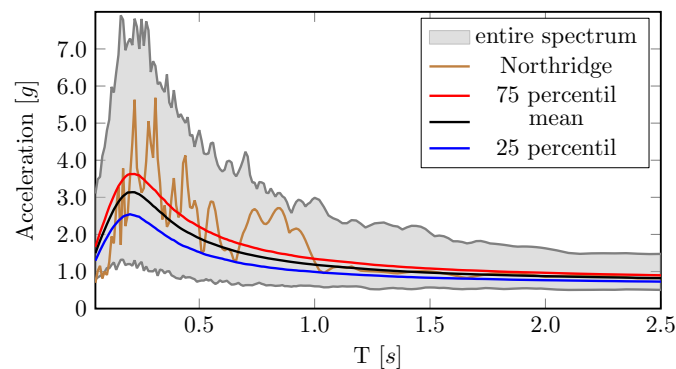


Figure 4. Acceleration response spectra of all generated earthquakes using the presented procedure. The mean, 25 percentile, and 75 percentile spectrum are compared with the spectrum of the benchmark earthquake (Northridge 1995).

2.3. Feedforward Neural Networks

Machine learning has become an indispensable tool in many applications. The incorporation of artificial neural networks in complex problems of mechanics has been shown useful in terms of structural response prediction. In this section, we provide the basic idea and key mathematical expressions, summarizing the fundamentals of feedforward neural networks, which can be found in detail in literature [40,41]. The term feedforward neural networks implies already the data flow of the neural network, as the information is transported forward from layer to layer. These architectures comprise at least two layers: an input and an output layer. In this case, they are called single-layer perceptrons. For this type, the output is calculated by the weighted sum of all inputs. However, this architecture is limited to the classification of separable patterns [40]. By adding so-called hidden layers between the input and output layer, one overcomes these limitations. These architectures are called multi-layer or deep feedforward neural networks or multi-layer perceptrons. The simplest architecture has one hidden layer, as depicted in Figure 5. First, one chooses appropriate information parameters for the input layer and the output parameters as the desired targets. In our case, the input parameters are chosen intensity measures, while the output parameter is the failure criterion. The data is passed forward from layer to layer via matrix multiplications, vector additions, and activations, where, in case of supervised learning, the coefficients of these matrices and vectors are the neural network parameters to be found through an optimization process.

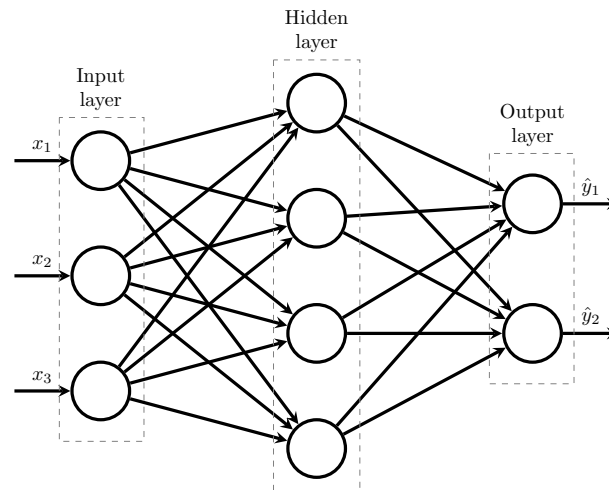


Figure 5. Graph of a feedforward neural network with one hidden layer.

The procedure of one forward pass is expressed mathematically as follows. The first layer ($l = 1$), is provided with normalized input data, such as, min-max feature scaling or standard score normalization, cf. [41]. The following layers ($l > 1$) are fed by the output of the previous layer, $\hat{y}^{(l)}$. The outputs are weighted by a matrix $\mathbf{W}^{(l)}$. Furthermore, a bias vector is added $\mathbf{b}^{(l)}$, i.e., $\mathbf{z}^{(l)} = \mathbf{W}^{(l)}\mathbf{x}^{(l)} + \mathbf{b}^{(l)}$. Additionally, each neuron within each layer is provided by an activation function $f(\mathbf{z}^{(l)})$. The most common activation functions are the sigmoid function $f(z) = \frac{e^z}{1+e^z}$, the tangent hyperbolicus function $f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$, and the rectified linear activation unit $f(z) = \max(0, z)$. The output of every hidden layer is written as $\hat{\mathbf{y}}^{(l)} = f(\mathbf{z}^{(l)})$.

In this manner, the information is transported through all the layers until the output layer $\hat{y}^{(l=L)} = \hat{y}$. During parameter optimization, the output layer is provided by the targets \mathbf{y} of the sample. The error is calculated by comparing this target with the output of the last layer. The weights and biases can be optimized by back-propagation of the error [42], e.g., the mean squared error $\epsilon = \frac{1}{N_y}(\mathbf{y} - \hat{\mathbf{y}})^2$:

$$\Delta \mathbf{W}^{(l)} = -\alpha \frac{\partial \epsilon}{\partial \mathbf{W}^{(l)}} \quad \text{with} \quad \frac{\partial \epsilon}{\partial \mathbf{W}^{(l)}} = \frac{\partial \mathbf{z}^{(l)}}{\partial \mathbf{W}^{(l)}} \frac{\partial \hat{\mathbf{y}}^{(l)}}{\partial \mathbf{z}^{(l)}} \frac{\partial \epsilon}{\partial \hat{\mathbf{y}}^{(l)}}. \tag{10}$$

Hereby, the gradient $\frac{\partial \epsilon}{\partial \hat{\mathbf{y}}^{(l)}}$ of the last layer ($l = L$) leads to $\frac{2}{N_y}(\hat{\mathbf{y}} - \mathbf{y})$. For all previous layers applies $\frac{\partial \epsilon}{\partial \hat{\mathbf{y}}^{(l)}} = \mathbf{W}^{(l+1)} \frac{\partial \hat{\mathbf{y}}^{(l+1)}}{\partial \hat{\mathbf{y}}^{(l)}}$.

The forward computation, the back-propagation of the error, and the respective update of the weights \mathbf{W} and biases \mathbf{b} are performed in each training iteration. This procedure is repeated until the error falls below a desired magnitude. Implementation-wise it has proven useful to update the neural network parameters after a certain training subset or so-called batches. After the neural network has seen every sample of the entire training data set, one epoch is finished. Generally, it takes many epochs to sufficiently train a neural network on the data set. In order to obtain the best results, the neural network architecture needs to be modified iteratively. These adjustments include the number of layers, the number of neurons in each layer, the activation function, the optimizer, and the input features.

2.4. Artificial Neural Network Strategy for Structural Response Prediction

The main goal of this study is to use artificial neural networks to decrease the computational effort and, therefore, to make the Monte Carlo method attractive for real-life applications. We focus on predicting the peak story drift ratio (PSDR) for the output instead

of predicting the complete response time history. The latter would require more complex neural network architectures, such as recurrent neural networks. Using feedforward architectures, there are several options, i.e., convolutions, to extract the features from the accelerogram [30]. However, the failure prediction is more accurate using a relatively small artificial neural network with fewer training samples. Therefore, this strategy is supposed to have the highest speed up compared to the full simulation procedure [43].

Carefully preprocessing the input data allows us to use a simple artificial neural network architecture. The challenge of so-called feature engineering is the generation of physically meaningful input parameters. In this context, in earthquake engineering, extensive research has been carried out to characterize seismic motion in terms of single magnitudes, such as earthquake intensities [44]. These intensity measures are used as input parameters for neural networks in this paper. Hereby, the Arias intensity I_A , the characteristic intensity I_c , and the cumulative absolute velocity (CAV) were already used for response predictions [45,46]. Besides using features directly from the accelerogram, one may extract additional spectral features. Therefore, the displacement, velocity, and acceleration response spectra as well as the spectral values at the first eigenfrequency of the structure are used to evaluate additional relevant input features. As a result of this, a meaningful combination of several intensity measures as input parameters was extensively studied [47]. For a steel frame structure, the combination of the effective peak ground acceleration (EPA), the spectral acceleration at the natural period $S_a(T_1)$, the velocity spectrum intensity $\int_{0.1}^{2.5} S_v$, the spectral displacement at the natural period $S_d(T_1)$, the cumulative absolute velocity (CAV), and the peak ground acceleration (PGA) revealed promising prediction results [29]. A set of relevant input parameters that can be calculated from an accelerogram are summarized in Table 1.

Table 1. Earthquake intensities used as neural network input.

Feature	Abbreviation	Formula
Acceleration spectrum intensity	$\int_{0.1}^{0.5} S_a$	$\int_{0.1}^{0.5} S_a(T) dT$
Arias intensity	I_a	$\frac{\pi}{2g} \int_0^t \ddot{x}_g dt$
Cumulative absolute velocity	CAV	$\int_0^t \ddot{x}_g dt$
Effective peak ground acceleration	EPA	$\frac{S_a(0.1, \dots, 0.5)}{2.5}$
Housner intensity	$\int_{0.1}^{2.5} S_d$	$\int_{0.1}^{2.5} S_d(T) dT$
Peak ground acceleration	PGA	$\max \ddot{x}_g $
Spectral acceleration at T_1	$S_a(T_1)$	$S_a(T_1)$
Spectral displacement at T_1	$S_d(T_1)$	$S_d(T_1)$
Spectral velocity at T_1	$S_v(T_1)$	$S_v(T_1)$
Velocity spectrum intensity	$\int_{0.1}^{2.5} S_v$	$\int_{0.1}^{2.5} S_v(T) dT$

We first calculate the set of artificial earthquake records \mathcal{S} , based on which we extract the corresponding intensity set \mathcal{I} that contains the intensity vectors

$$\mathbf{X} = [I_a, S_a(T_1), \int_{0.1}^{2.5} S_v, S_d(T_1), \text{CAV}, \text{PGA}]:$$

$$\mathcal{S} = \langle \ddot{x}_g^{(1)}(t), \dots, \ddot{x}_g^{(l)}(t) \rangle \Rightarrow \mathcal{I} = \langle \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(l)} \rangle. \tag{11}$$

Applying supervised learning, the neural network learns from training data. Therefore, at first, a subset of \mathcal{S} must be evaluated:

$$\mathcal{S}_t = \langle \ddot{x}_g^{(1)}(t), \dots, \ddot{x}_g^{(n)}(t) \rangle \Rightarrow \mathcal{R}_t = \langle \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)} \rangle. \tag{12}$$

The neural network learns the damage indicator directly, which is, in our case, the PSDR. The target set for the neural network training, \mathcal{O}_t , is calculated from the response set \mathcal{R}_t :

$$\mathcal{R}_t = \langle \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)} \rangle \Rightarrow \mathcal{O}_t = \langle \text{PSDR}^{(1)}, \dots, \text{PSDR}^{(n)} \rangle . \tag{13}$$

Using a neural network, the prediction of the sets are written as

$$\mathcal{I}_t = \langle \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)} \rangle \Rightarrow \hat{\mathcal{O}}_t = \langle \text{PS}\hat{\text{D}}\text{R}^{(1)}, \dots, \text{PS}\hat{\text{D}}\text{R}^{(n)} \rangle . \tag{14}$$

Once the neural network is trained and optimized in such a way that the predictions converge to the targets, the neural network can be used to predict the individual response values and, consequently, the full set:

$$\mathcal{I} = \langle \mathbf{X}^{(1)}, \dots, \mathbf{X}^{(l)} \rangle \Rightarrow \hat{\mathcal{O}} = \langle \text{PS}\hat{\text{D}}\text{R}^{(1)}, \dots, \text{PS}\hat{\text{D}}\text{R}^{(l)} \rangle . \tag{15}$$

2.5. The New Training Data Selection Method Based on Sample Intensities

Artificial neural networks can solve a wide variety of tasks in several fields of application. Once a neural network is properly trained, it can make predictions that are both accurately and efficiently. In general, neural networks will predict reliably as long as they have seen similar patterns before. Thus, one may imagine that the mean of the data set will be excellently predicted. However, in this study, we want to estimate the response of events near the failure region of the distribution, i.e., extreme events that happen very rarely. Therefore, we provide a new strategy to cover the entire domain of possible events during the training procedure. The major steps of the procedure to use machine learning are shown in Figure 6, while the appropriate choice of the training data is addressed in this paper.

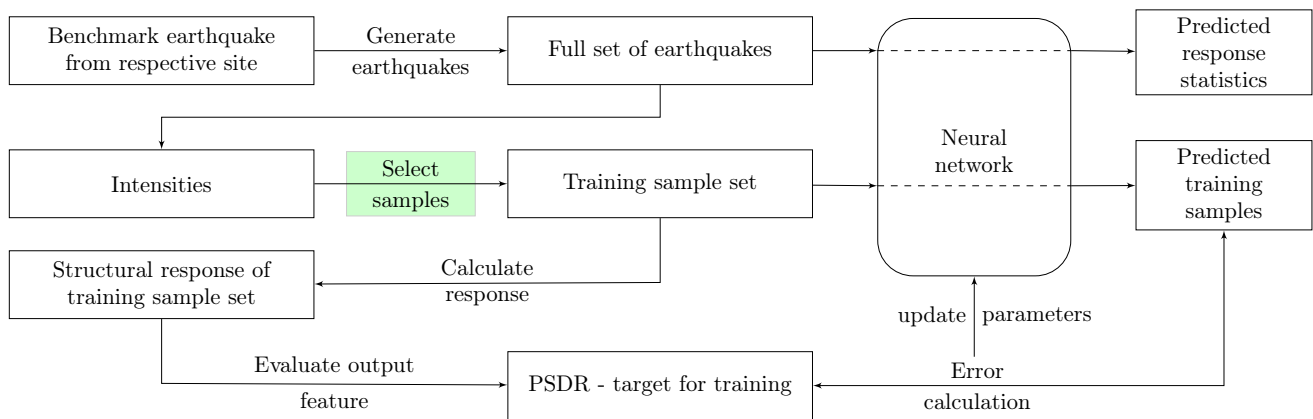


Figure 6. Flowchart of machine learning enhanced structural response statistics using the proposed strategy for the selection of the training samples during supervised learning.

In particular, Figure 7 shows all the steps of the new selection process. Instead of using an arbitrary training data set $\mathcal{I}_{t,rand}$, we select a training data set $\mathcal{I}_{t,sel}$ based on the intensities of the generated artificial earthquake samples. Therefore, we split each intensity range into evenly distributed intervals. Dependent on the number of intensity measures considered for the selection, we choose from an i -dimensional grid. If the selection is based on one intensity only, we pick one sample from each bin. In this case, the number of selected samples is equal to the number of bins or slightly smaller if empty bins exist. This procedure will likely give a better-distributed training set than a randomly picked set, especially, if the chosen intensity correlates with the output quantity. However, it will probably not cover the whole domain of all intensity measures chosen as input parameters. One may interpret this strategy as an adaption of the Latin-Hypercube sampling approach. The difference is

that we perform the sampling beforehand based on non-stationary Gaussian random white noises and, subsequently, filter the data.

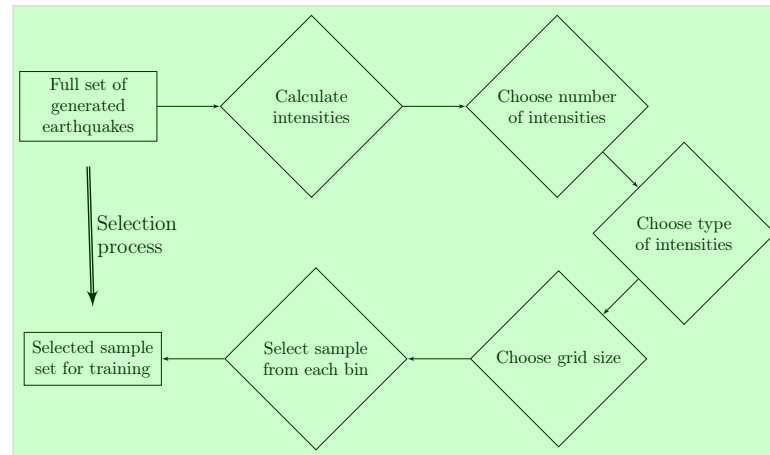


Figure 7. Flowchart of the proposed selection strategy for the training samples.

The domain is represented better considering two intensities. The procedure can be easily shown for this setup. In Figure 8, we present the correlation between velocity spectrum intensity $\int_{0.1}^{2.5} S_v$ and the spectral displacement $S_d(T_1)$. In particular, Figure 8a shows all samples from set \mathcal{I} , while Figure 8b shows only 450 samples from this set, in other words a randomly chosen subset used for the training, $\mathcal{I}_{t,rand}$. As one can see, most samples with high intensities are not included in this set, although they are the most important samples that allow for the neural network to learn more around the tail region of the distribution, as the intensities have a certain correlation with the structural response. In Figure 8c, we also present the samples chosen from \mathcal{I} using this strategy. Both Intensities were cut into 34 intervals, which resulted in the selection of 438 samples. In doing so, we achieve that the selected training set $\mathcal{I}_{t,sel}$ covers the whole parameter space of both intensities. Therefore, the neural network will learn from very frequent and from very rare events to the same extent.

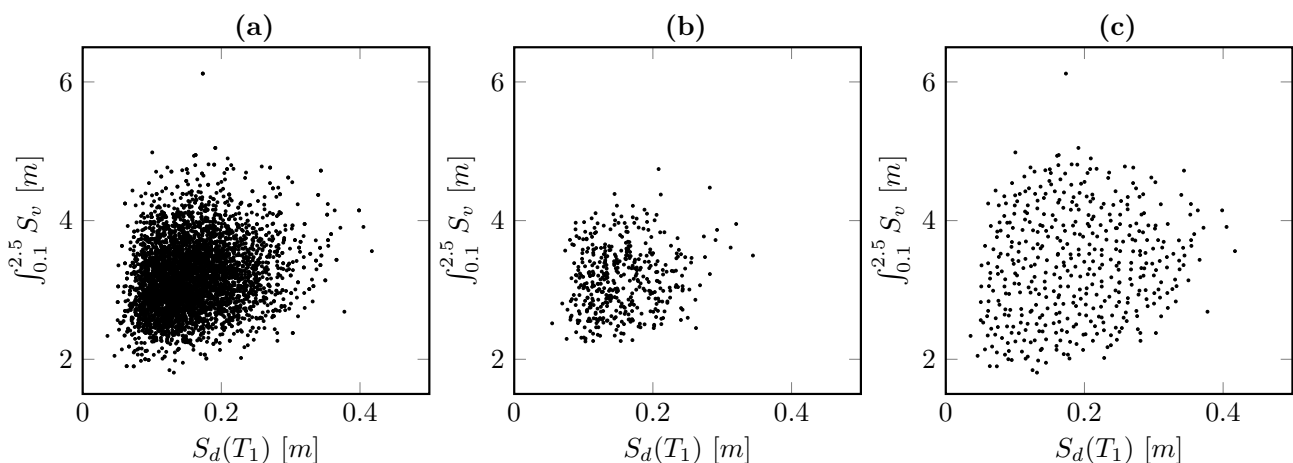


Figure 8. Two-dimensional selection strategy: selection of samples from (a) the full set \mathcal{I} using the velocity spectrum intensity $\int_{0.1}^{2.5} S_v$ and the spectral displacement at the natural period $S_d(T_1)$, (b) randomly chosen samples for the training set $\mathcal{I}_{t,rand}$, and (c) selected samples for the training set $\mathcal{I}_{t,sel}$.

The crucial part of this strategy is a meaningful choice in the number of the intensities considered and their type to perform the selection procedure. Furthermore, the grid size

must be adjusted to get a certain amount of samples, and the reduction of the grid size results in a higher number of training samples. In particular, if several intensity measures are considered, the grid size must be chosen large enough to select a small share of the set.

The selection process, shown in Figure 9, is based on three intensities: the velocity spectrum intensity $\int_{0.1}^{2.5} S_v$, the spectral displacement $S_d(T_1)$, and the spectral acceleration at the natural period $S_a(T_1)$. Each intensity is split into 19 bins resulting in a selection of 565 samples. Figure 9 shows that the selected data points for the three intensities spread evenly over the entire domain of interest.

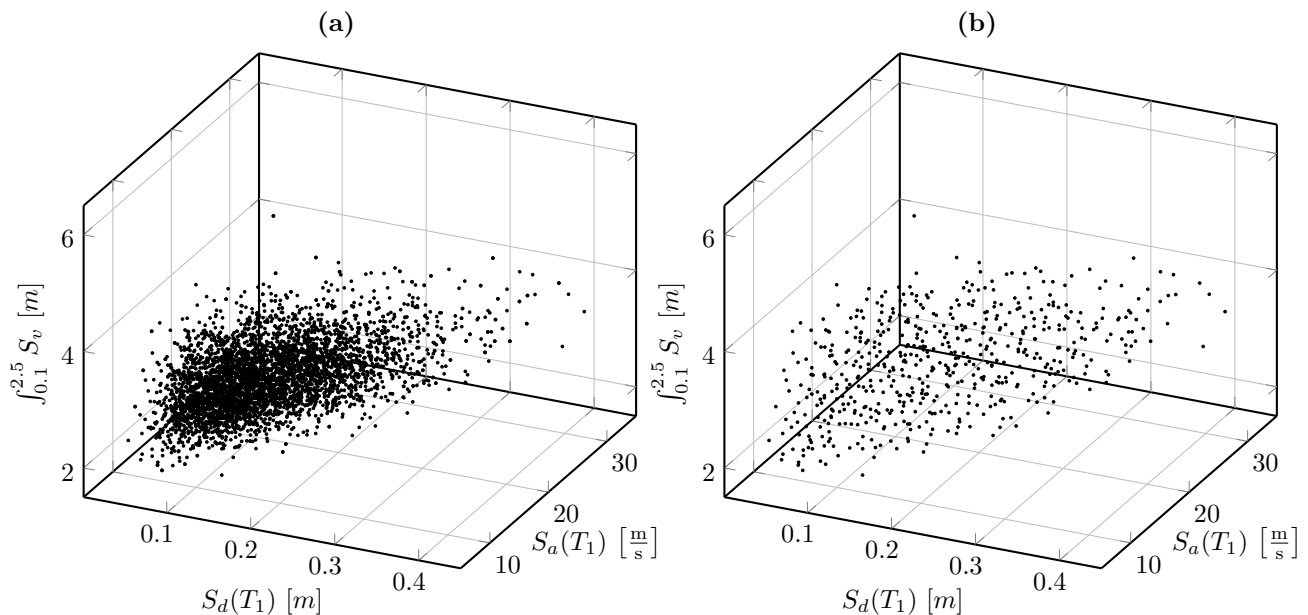


Figure 9. Three-dimensional selection strategy: Selection of samples from (a) the full set \mathcal{I} using the velocity spectrum intensity $\int_{0.1}^{2.5} S_v$, the spectral displacement at the natural period $S_d(T_1)$ and the spectral acceleration at the natural period $S_a(T_1)$, and (b) selected samples for the training set $\mathcal{I}_{t,sel}$.

3. Numerical Example

In this section, we provide a numerical demonstration of the new strategy. We chose the peak story drift ratio (PSDR) as the damage indication criterion. Thus, the evaluation of the PSDR was used for the entire training data set and the reference solution using the crude Monte Carlo simulation approach. We used an in-house finite element tool written in C++ and Python to calculate the nonlinear structural response. This tool has been used in several previous studies and has been validated using commercial finite element codes [48]. For the implementation of artificial neural network architectures, we used the Python library TensorFlow [49]. Three different structures are subjected to synthetic seismic excitations to present the new strategy. Furthermore, we provide the neural network predictions based on different numbers of intensities chosen to select the training data.

3.1. Nonlinear Frame Structures

The three structures are modeled by fiber beam elements using an elastoplastic material law with kinematic hardening, as shown in Figure 10. We chose a Young's modulus of $E_1 = 210$ GPa and a material density of $\rho = 7850 \frac{\text{kg}}{\text{m}^3}$. The yielding limit of the material was set as a value of $\sigma_y = 235$ MPa. The post-yielding stiffness was chosen as $E_2 = 21$ GPa, which is 10% of initial Young's modulus. The beams and columns of the frames were modeled using the same beam element formulation, shown in Figure 10. For the fiber-beam elements, we used hollow cross sections with different widths, heights, and thicknesses.

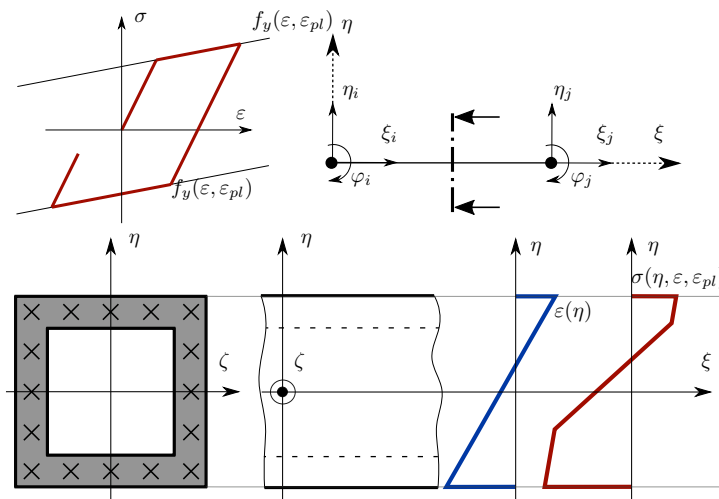


Figure 10. Beam elements with a hollow cross section and elastoplastic material behavior with kinematic hardening.

The geometrical inputs for all structures used for the numerical experiments are summarized in Tables 2 and 3. Additionally, the structures are presented in Figure 11. To account for a realistic mass distribution, all structures have additional point masses. The first natural period of the structures A, B and C are $T_{1,A} = 0.74$ s, $T_{1,B} = 0.40$ s and $T_{1,C} = 0.88$ s, respectively.

Table 2. Dimensions of the structures.

Dimension	Structure A	Structure B	Structure C
w_{bay}	6 m	5.5 m	6 m
h_{s1}	4.5 m	3.5 m	4.5 m
h_{s2}	4 m	3.0 m	4 m
h_{s3}	4 m		4 m
h_{s4}			4 m
h_{s5}			4 m

Table 3. Dimensions of the cross sections used for the beams and columns.

Dimension	P ₁	P ₂	P ₃	P ₄
w_b	0.3 m	0.35 m	0.3 m	0.35 m
h_b	0.3 m	0.35 m	0.4 m	0.45 m
t_b	0.03 m	0.03 m	0.03 m	0.03 m

The PSDR for the first two stories is depicted in Figure 11 for structure C. For these particular problems, we observed that a story drift ratio of $\sim 4.5\%$ corresponds to a full plastification of the frame corners in the first stories. These assumptions agree with the estimation of the collapse of steel structures [50]. However, lower story drift ratios can already cause damage to the structure, i.e., story drift ratios above 2% lead already to plastification of the frame corners. More elaborate models could also include damage and softening effects and may improve the failure criterion [51], which is material for future research. However, in this study, we focus on the general approach of the strategy with the PSDR as representative output quantity and an elastoplastic material law from the underlying finite element model.

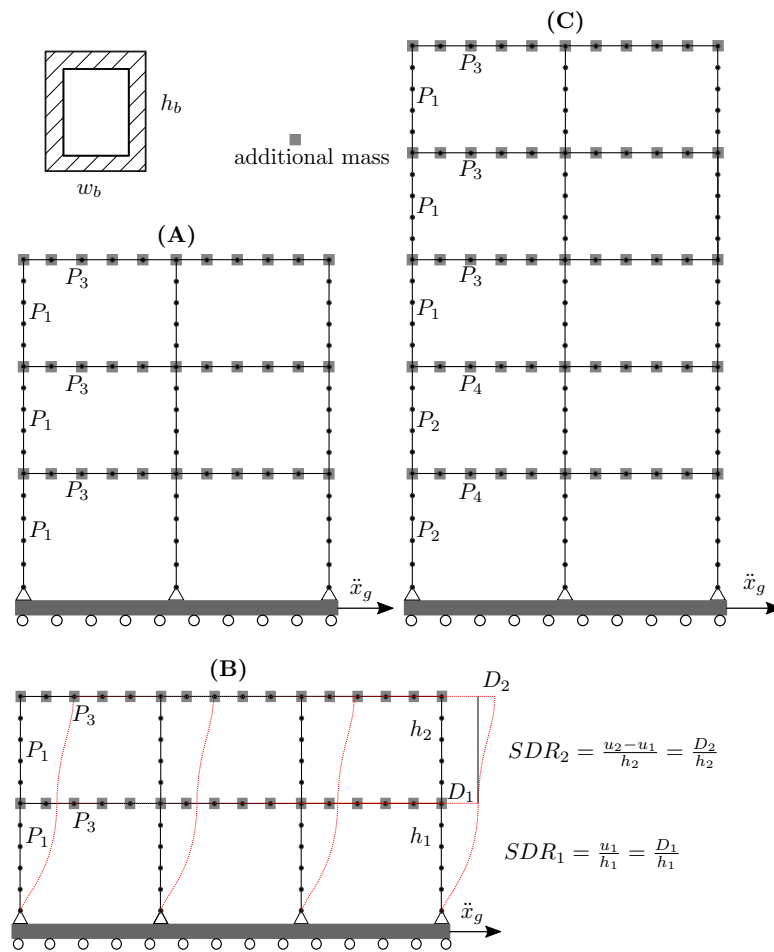


Figure 11. Numerical example to demonstrate the strategy: (A) Two-bay-tree-story frame structure; (B) three-bay-two-story frame structure; (C) two-bay-five-story frame structure.

3.2. Training Data Selection from Generated Earthquake Samples

Figure 12 reveals the distribution of the intensities of the training set using the violin plot format. To compare the statistics of each set, all intensities are scaled between zero and one by min-max normalization based on the entire set. In Figure 12a, the violin plots of the full set are shown. Comparing this set with a randomly chosen sample set, shown in Figure 12b, the issue becomes clearer. Observing the randomly chosen training distribution, one immediately realizes that the most severe earthquakes are not covered. However, using our new selection strategy enables covering a broad range of data that include a large share of high- and low-intensity values, as shown in Figure 12c. Therefore, these events are covered more likely in the training set. The presented distributions in Figure 12c are based on the selection using two intensities, and Figure 12d is based on the selection using three intensities. The strategy clearly shows the anticipated effect on the training set. The distributions based on two or three dimensions do, however, not change significantly.

Increasing the number of intensities that contribute to the selection process, the number of bins needs to be decreased. Otherwise, the number of training samples increases and the computational benefit of the whole strategy vanishes. Using six intensities for the selection process, the probability density function of each intensity shows an uneven distribution, as shown in Figure 13. The peaks of these distributions correlate with the number of the selected bins, and the share of the higher selected Intensities decreases except for the peak ground acceleration. Therefore, we propose to select training data from a relatively small number of intensities, i.e., two or three different features.

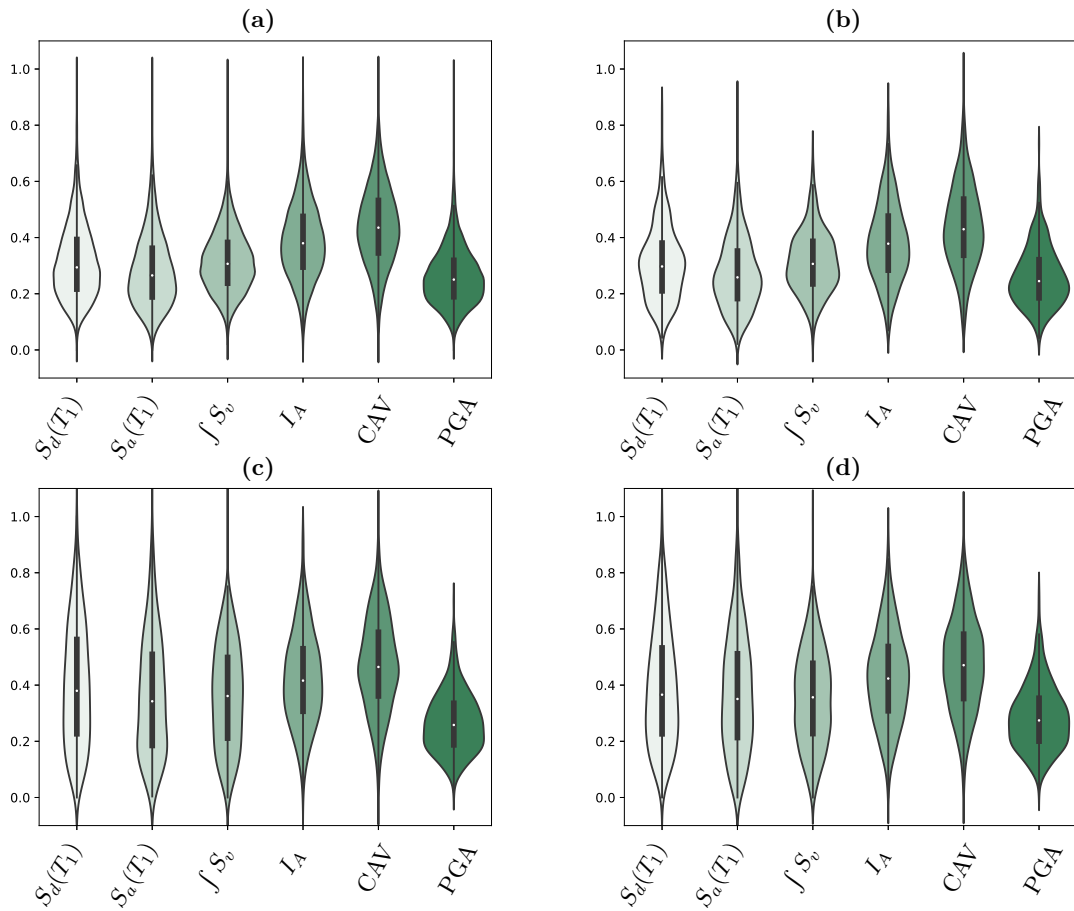


Figure 12. Violin plots [52,53] show the distribution of the intensities, which are used as inputs for the neural network; (a) full set, 10,000 samples; (b) 500 randomly chosen samples from the full set; (c) 438 samples are chosen as training data based on the proposed strategy using two intensities (velocity spectrum intensity $\int_{0.1}^{2.5} S_v$ and the spectral displacement $S_d(T_1)$) with 34 bins each; (d) 565 samples are chosen as training data based on the proposed strategy using three intensities (velocity spectrum intensity $\int_{0.1}^{2.5} S_v$, the spectral displacement $S_d(T_1)$ and the spectral acceleration $S_a(T_1)$) with 19 bins each.

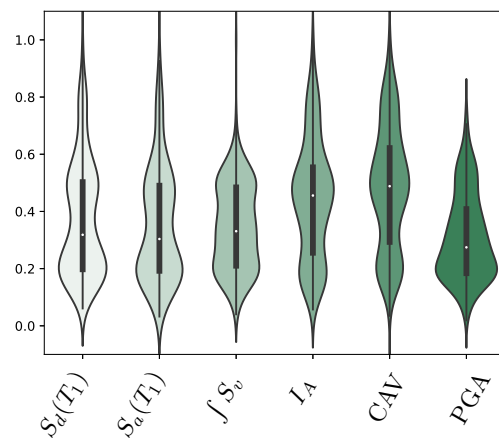


Figure 13. Violin plots [52,53] show the distribution of the intensities, which are used as inputs for the neural network; 291 samples are chosen as training data based on the proposed strategy using the shown six intensities with 5 bins each.

3.3. Predicted Response Statistics

The response statistics using the Crude Monte Carlo method were performed calculating $l = 10^4$ artificially generated earthquakes, which constitutes for all simulations the entire benchmark set \mathcal{S} . The peak story drift ratio set \mathcal{O} was calculated using the finite element method by evaluating the reference solutions through numeric time integration using the Newmark method [54]. We applied the new strategy to the structures A, B, and C (see Figure 11) using the proposed sample selection for two, three, and six intensities, respectively.

For each selection, an artificial neural network was trained. The input vector \mathbf{X} consists of six features; therefore, the input layer also had six neurons. We used three hidden layers, consisting of between 11 and 14 neurons, using rectified linear activation functions. The output layer consisted of one neuron only, which produced the PSDR prediction. We used the ADAM algorithm to update the learning rate during training [55]. This configuration was used for all neural network architectures presented in this study.

The statistics of the predictions are shown in Figure 14. For the structures, A, B, and C, the prediction of the PSDR is shown in the top, middle, and bottom plots of Figure 14, respectively. The probability density function (PDF) is shown on the left-hand side, while the complementary cumulative distribution function (CCDF) is shown on the right-hand side.

For the two-dimensional selection process, the samples were selected based on the velocity spectrum intensity $\int_{0.1}^{2.5} S_v$ and the spectral displacement $S_d(T_{1,A})$ using 29 bins each. The selected set consists of $n = 333$ samples, which was used for the training. The validation data was taken from the full set and has a size of 20% of the entire training set. Furthermore, considering the spectral acceleration $S_a(T_{1,A})$, the 3D selection method used 15 bins, which resulted in 336 samples for the training. The six-dimensional selection method used five bins, which resulted in 291 samples. The PDF of the predicted PSDR of structure A is shown in Figure 14a and shows the expected agreement with the finite element solution. In Figure 14b, the corresponding CCDF is shown. On the one hand, we can see that the prediction of randomly selected data cannot predict the tail region correctly. On the other hand, all neural networks trained by the new selected data approach predict this region well.

We performed the same procedure for structure B. The two-dimensional selection process was based on $\int_{0.1}^{2.5} S_v$ and $S_d(T_{1,B})$, using 27 bins for each resulted in $n = 318$ training data samples. Considering the spectral acceleration $S_a(T_{1,B})$ for the 3D selection, we used 15 bins, which resulted in $n = 325$ samples for the training set $S_{t,sel}$. We used only five bins for the six-dimensional selection, which resulted in $n = 274$ samples. Independent of the training data, the PDF of the predicted PSDR of structure B, shown in Figure 14c, matches the finite element solution. Figure 14d shows the corresponding CCDF. Compared to the previous example using structure A, we observe that the prediction of randomly selected data performs better in the tail region. This structure shows higher resistance to earthquakes and the neural network can predict the results of the extreme event better in this case, although the prediction overshoots the finite element solution slightly. The neural networks trained by the selected data approach can predict this region well.

Structure C shows the highest share of samples exceeding a PSDR of 4.5%. The two dimensional selection process based on $\int_{0.1}^{2.5} S_v$ and $S_d(T_{1,C})$, with 28 bins each, resulted in $n = 331$ selected samples for training. Considering $S_a(T_{1,C})$ for the three dimensional selection and using 14 bins, resulted in $n = 309$ selected samples. The six-dimensional selection method resulted in $n = 274$ samples if five bins were used for each intensity. The PDF of the PSDR using structure C is shown in Figure 14e. All artificial neural networks can predict the statistics in the mean region well. Figure 14f shows the CCDF. However, the artificial neural network trained by randomly selected data fails to predict the statistics in the tail region of the distribution. The results of the selected training procedure show that it is beneficial to use two or three intensities during the selection instead of using all

six intensities. Generally, the extreme events are not covered properly if large bin sizes are chosen.

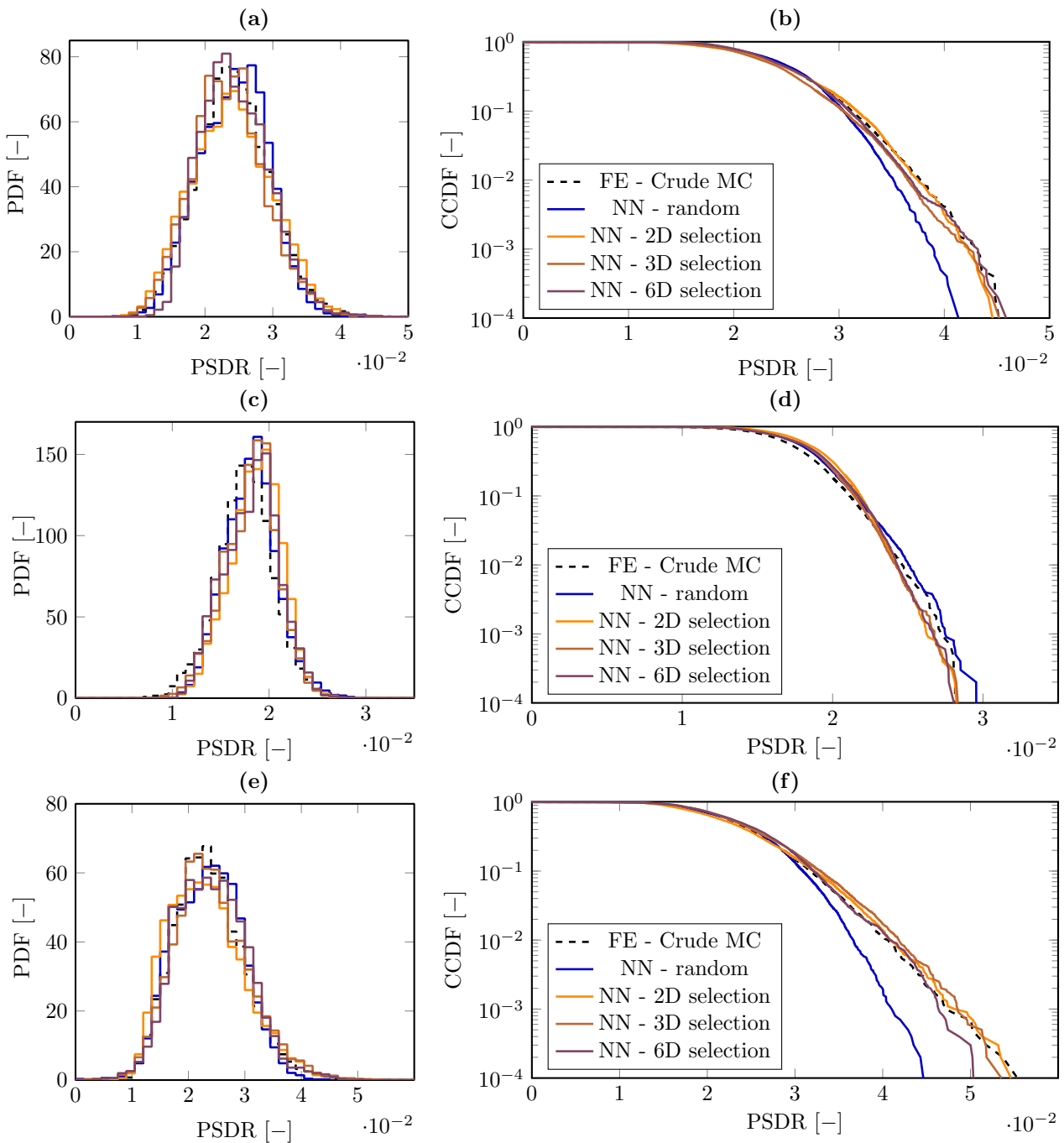


Figure 14. Response statistics. Structure A: (a) Probability density function and (b) complementary cumulative distribution function; Structure B: (c) Probability density function and (d) complementary cumulative distribution function; Structure C: (e) Probability density function and (f) complementary cumulative distribution function. Color scheme: Finite element solution (black, dashed), predicted solution from neural network with randomly chosen data (blue), predicted solution from neural network with selected data using selection process based on $S_d(T_1)$ and $\int_{0.1}^{2.5} S_v$ (orange), predicted solution from neural network with selected data using selection process based on $S_d(T_1)$, $\int_{0.1}^{2.5} S_v$ and $S_d(T_1)$ (brown), predicted solution from neural network with selected data using selection process based on all six intensities (dark ruby).

4. Discussion

Our three numerical studies reveal that the neural networks trained by the new data selection can predict the tail region. In particular, the response of structure B is predicted well by each neural network. Regarding the taller structures A and C, the neural network trained by randomly selected data fails to predict rare events, while the statistics are provided accurately by selected training data. Hereby, it turns out that the selection strategy, using two or three intensities, namely, the velocity spectrum intensity $\int_{0.1}^{2.5} S_v$, the spectral displacement $S_d(T_{1,A})$, and the spectral acceleration $S_a(T_{1,A})$, provides the most stable solution. The selection process using all six intensities does not ensure the inclusion of extreme events, as the bin size must be chosen rather large.

The predictions of artificial neural networks using randomly selected data during training are unstable. Even though the mean region is predicted well, the results in the tail region give inconsistent results. The easiest way to tackle this problem is to increase the training data set. The predictions will be less sensitive to each randomly selected point, and it is likely to perform better on the task. However, the computational benefit of the entire strategy would suffer. While preserving the computational efficiency, we can reduce the instability using the proposed selection strategy. We propose to use two or three intensity measures for the selection to make this procedure efficient. Using the new selection strategy, the predicted response statistics significantly improve in the tail region of the distribution. However, the neural network prediction using the same training data can still vary slightly if the influence of the randomness, such as the initialization of the weights and biases of the neural networks, is not removed.

The proposed procedure increases the share of rare events from the distribution in the training data without increasing the computational cost. Another strategy to provide extreme events in the training data is to use scaling methods during the generation procedure of the synthetic earthquake [29]. The generation of another set of earthquakes results in additional computational effort. Furthermore, these excitations must be generated carefully regarding factorization; otherwise, the prediction on a differently generated set can lead to unsatisfactory results. The new proposed strategy does not need to create extended training data. Instead, the data are taken from the already generated set. Thus, compared with the strategy of extended sets, this selection process benefits in its applicability.

The prediction of the structural response of a single earthquake might be improved by choosing more complex neural network architectures, such as convolutional neural networks. Advanced machine learning algorithms would learn from the unprocessed input data, i.e., the entire time history of the generated earthquakes. The downside of employing more sophisticated architectures, such as convolutional neural networks, is that the simulation requires higher computational effort during training. The neural network must learn more parameters, and the number of training samples that need to be evaluated for the training data set increases significantly.

The computational cost for training the network parameters and predicting the proposed neural network is small compared to the cost of evaluating the training data. Therefore, the computational speed-up can be roughly estimated by dividing the total number of samples l of the entire set S by the number of samples n of the training set S_t . Using 333 samples for training and 20% for validation leads to a speed-up of around 25. Considering the calculation of the intensities, this speed-up will decrease slightly. Overall, the numerical experiments have shown that the new procedure results in a speed-up factor of above 20.

Limitations and Potential Advancements

The approach provides a reliable prediction of the tail end of the response statistics. Yet, we tested this strategy on two-dimensional models only. Three-dimensional structures will be of high interest for this strategy as the number of degrees of freedom drastically increases and, consequently, the computational cost. In this case, the accuracy of the proposed approach must be tested. Furthermore, we used a rather simple assumption

for the criterion of structural failure. Incorporating more elaborate material models that include material damage and softening will result in a significantly higher computational cost. It is interesting to evaluate the computational speed up in this case since we assume that the neural network prediction will not be affected by this extension. However, in this study, we used a bi-linear material model and analyzed the behavior in the cross section with the highest stresses and focused on the level of plastification. Evaluating static cyclic test calculations one observes that, if the story drift ratio exceeds the chosen failure criterion of a maximum PSDR the most affected cross section, i.e., the cross section in the first frame corner, fully plastifies.

5. Conclusions

This paper proposes a new training data selection strategy for neural network-enhanced Monte Carlo simulations. The training data selection enables one to reliably predict the whole response statistic domain, while revealing a high level of computational efficiency. Based on a synthetic earthquake set that is generated taking into account the properties of a specific region, this strategy provides a powerful technique to predict the probability of failure, even in the tail end of the distribution.

Although the application of the proposed selection strategy reduces the sensitivity of classically trained neural network architectures, the prediction quality is still sensitive to the neural network parameters. Thus, future studies are necessary to find the best neural network architecture and stabilizing measures for consistent output quality. However, we found a way to effectively choose training data to apply artificial neural networks for seismic response statistics.

The use of more than only one benchmark earthquake record can be incorporated into the data generation process, which is of high interest for future research. Furthermore, the applicability to more realistic structures could reveal even more significant benefits of the new strategy in the future, as more complicated structures do not necessarily require more sophisticated neural network architectures, which would, finally, reveal an even higher speed up in computation time.

Author Contributions: Conceptualization, D.T. and F.B.; methodology, D.T., L.E. and F.B.; software, D.T., L.E. and F.B.; validation, D.T., L.E. and F.B.; formal analysis, L.E. and D.T.; investigation, D.T.; resources, B.M. and F.B.; data curation, D.T. and F.B.; writing—original draft preparation, D.T.; writing—review and editing, F.B. and B.M.; visualization, D.T.; supervision, F.B. and B.M.; project administration, F.B. and B.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CAV	Cumulative absolute velocity
CCDF	Complementary cumulative distribution function
CDF	Cumulative distribution function
EPA	Effective peak ground acceleration
PDF	Probability density function
PGA	Peak ground acceleration
PSDR	Peak story drift ratio
SDOF	Single degree of freedom system

References

1. Crozet, V.; Politopoulos, I.; Chaudat, T. Shake table tests of structures subject to pounding. *Earthq. Eng. Struct. Dyn.* **2019**, *48*, 98–8847. [[CrossRef](#)]
2. Furinghetti, M.; Yang, T.; Calvi, P.M.; Pavese, A. Experimental evaluation of extra-stroke displacement capacity for curved surface slider devices. *Soil Dyn. Earthq. Eng.* **2021**, *146*, 106752. [[CrossRef](#)]
3. Ghezelbash, A.; Beyer, K.; Dolatshahi, K.M.; Yekrangnia, M. Shake table test of a masonry building retrofitted with shotcrete. *Eng. Struct.* **2020**, *219*, 110912. [[CrossRef](#)]
4. Furinghetti, M.; Lanese, I.; Pavese, A. Experimental assessment of the seismic response of a base isolated building through hybrid simulation technique. *Front. Built Environ.* **2020**, *6*, 33. [[CrossRef](#)]
5. Bucher, C. *Computational Analysis of Randomness in Structural Mechanics*; Taylor & Francis: London, UK, 2009.
6. Bucher, C. Asymptotic sampling for high-dimensional reliability analysis. *Probabilistic Eng. Mech.* **2009**, *24*, 504–510. [[CrossRef](#)]
7. Au, S.K.; Beck, J.L. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Eng. Mech.* **2001**, *16*, 263–277. [[CrossRef](#)]
8. Bamer, F.; Bucher, C. Application of the proper orthogonal decomposition for linear and nonlinear structures under transient excitations. *Acta Mech.* **2012**, *223*, 2549–2563. [[CrossRef](#)]
9. Bamer, F.; Markert, B. An efficient response identification strategy for nonlinear structures subject to non-stationary generated seismic excitations. *Mech. Based Des. Struct. Mach.* **2017**, *45*, 313–330. [[CrossRef](#)]
10. Bamer, F.; Amiri, A.; Bucher, C. A new model order reduction strategy adapted to nonlinear problems in earthquake engineering. *Earthq. Eng. Struct. Dyn.* **2017**, *46*, 537–559. [[CrossRef](#)] [[PubMed](#)]
11. Stoffel, M.; Bamer, F.; Markert, B. Artificial neural networks and intelligent finite elements in non-linear structural mechanics. *Thin-Walled Struct.* **2018**, *131*, 102–106. [[CrossRef](#)]
12. Stoffel, M.; Bamer, F.; Markert, B. Neural network based constitutive modeling of nonlinear viscoplastic structural response. *Mech. Res. Commun.* **2019**, *95*, 85–88. [[CrossRef](#)]
13. Stoffel, M.; Gulakala, R.; Bamer, F.; Markert, B. Artificial neural networks in structural dynamics: A new modular radial basis function approach vs. convolutional and feedforward topologies. *Comput. Methods Appl. Mech. Eng.* **2020**, *364*, 112989. [[CrossRef](#)]
14. Stoffel, M.; Bamer, F.; Markert, B. Deep convolutional neural networks in structural dynamics under consideration of viscoplastic material behaviour. *Mech. Res. Commun.* **2020**, *108*, 103565. [[CrossRef](#)]
15. Sun, H.; Burton, H.V.; Huang, H. Machine learning applications for building structural design and performance assessment: State-of-the-art review. *J. Build. Eng.* **2021**, *33*, 101816. [[CrossRef](#)]
16. Mignan, A.; Broccardo, M. Neural Network Applications in Earthquake Prediction (1994–2019): Meta-Analytic and Statistical Insights on Their Limitations. *Seismol. Res. Lett.* **2020**, *91*, 2330–2342. [[CrossRef](#)]
17. Kostinakis, K.; Morfidis, K. Application of Artificial Neural Networks for the Assessment of the Seismic Damage of Buildings with Irregular Infills' Distribution. In *Seismic Behaviour and Design of Irregular and Complex Civil Structures III*; Springer: Cham, Switzerland, 2020; pp. 291–306.
18. Salkhordeh M, Mirtaheri M, Soroushian S. A decision-tree-based algorithm for identifying the extent of structural damage in braced-frame buildings. *Struct. Control Health Monit.* **2021**, *28*, e2825. [[CrossRef](#)]
19. Zhang, Y.; Burton, H.V.; Sun, H.; Shokrabadi, M. A machine learning framework for assessing post-earthquake structural safety. *Struct. Saf.* **2018**, *72*, 1–16. [[CrossRef](#)]
20. Zhang, Y.; Burton, H.V. Pattern recognition approach to assess the residual structural capacity of damaged tall buildings. *Struct. Saf.* **2019**, *78*, 12–22. [[CrossRef](#)]
21. Thaler, D.; Bamer, F.; Markert, B. A machine learning enhanced structural response prediction strategy due to seismic excitation. *Proc. Appl. Math. Mech.* **2021**, *20*, e202000294. [[CrossRef](#)]
22. Muin, S.; Mosalam, K.M. Structural Health Monitoring Using Machine Learning and Cumulative Absolute Velocity Features. *Appl. Sci.* **2021**, *11*, 5727. [[CrossRef](#)]
23. Gao, Y.; Mosalam, K.M.; Chen, Y.; Wang, W.; Chen, Y. Auto-Regressive Integrated Moving-Average Machine Learning for Damage Identification of Steel Frames. *Appl. Sci.* **2021**, *11*, 6084. [[CrossRef](#)]
24. Karami-Mohammadi, R.; Mirtaheri, M.; Salkhordeh, M.; Hariri-Ardebili, M. A cost-effective neural network-based damage detection procedure for cylindrical equipment. *Adv. Mech. Eng.* **2019**, *11*, 1–10. [[CrossRef](#)]
25. Salkhordeh, M.; Govahi, E.; Mirtaheri, M. Seismic fragility evaluation of various mitigation strategies proposed for bridge piers. *Structures* **2021**, *33*, 1892–1905. [[CrossRef](#)]
26. Sediek, O.A.; El-Tawil, S.; McCormick, J. Seismic Debris Field for Collapsed RC Moment Resisting Frame Buildings. *J. Struct. Eng.* **2021**, *147*, 04021045. [[CrossRef](#)]
27. Siam, A.; Ezzeldin, M.; El-Dakhkhni, W. Machine learning algorithms for structural performance classifications and predictions: Application to reinforced masonry shear walls. *Structures* **2019**, *22*, 252–265. [[CrossRef](#)]
28. Huang, H.; Burton, H.V. Classification of in-plane failure modes for reinforced concrete frames with infills using machine learning. *J. Build. Eng.* **2019**, *25*, 100767. [[CrossRef](#)]
29. Thaler, D.; Stoffel, M.; Markert, B.; Bamer, F. Machine-learning-enhanced tail end prediction of structural response statistics in earthquake engineering. *Earthq. Eng. Struct. Dyn.* **2021**, *50*, 2098–2114. [[CrossRef](#)]

30. Bamer, F.; Thaler, D.; Stoffel, M.; Markert, B. A Monte Carlo Simulation Approach in Non-linear Structural Dynamics Using Convolutional Neural Networks. *Front. Built Environ.* **2021**, *7*, 53. [[CrossRef](#)]
31. Kanai, K. Semi-Empirical Formula for the Seismic Characteristics of the Ground. *Bull. Earthq. Res. Inst. Univ. Tokyo* **1957**, *35*, 309–325.
32. Tajimi, H. A Statistical Method of Determining the Maximum Response of a Building Structure during an Earthquake. In Proceedings of the 2nd World Conference on Earthquake Engineering, Tokyo, Japan, 11–18 July 1960; pp. 781–798.
33. Fan, F.; Ahmadi, G. Nonstationary Kanai-Tajimi models for El Centro 1940 and Mexico City 1985 earthquakes. *Probabilistic Eng. Mech.* **1990**, *5*, 171–181. [[CrossRef](#)]
34. Rofooei, F.; Mobarake, A.; Ahmadi, G. Generation of artificial earthquake records with a nonstationary Kanai-Tajimi model. *Eng. Struct.* **2001**, *23*, 827–837. [[CrossRef](#)]
35. Center for Engineering Strong Motion Data. Northridge, California 1994-01-17 12:30:55 UTC. Virtual Data Center. Available online: www.strongmotioncenter.org/ (accessed on 1 October 2021).
36. Rezaeian, S.; Der Kiureghian, A. A stochastic ground motion model with separable temporal and spectral nonstationarities. *Earthq. Eng. Struct. Dyn.* **2008**, *37*, 1565–1584. [[CrossRef](#)]
37. Rezaeian, S.; Der Kiureghian, A. Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq. Eng. Struct. Dyn.* **2010**, *39*, 1155–1180. [[CrossRef](#)]
38. Bamer, F.; Markert, B. A nonlinear visco-elastoplastic model for structural pounding. *Earthq. Eng. Struct. Dyn.* **2018**, *47*, 2490–2495.
39. Bamer, F.; Strubel, N.; Shi, J.; Markert, B. A visco-elastoplastic pounding damage formulation. *Eng. Struct.* **2019**, *197*, 109373. [[CrossRef](#)]
40. Fogel, D.B.; Liu, D.; Keller, J.M. *Fundamentals of Computational Intelligence: Neural Networks, Fuzzy Systems, and Evolutionary Computation*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2016.
41. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: London, UK, 2016.
42. Rumelhart, D.; Hinton, G.; Williams, R. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
43. Thaler, D.; Bamer, F.; Markert, B. A comparison of two neural network architectures for fast structural response prediction. *Proc. Appl. Math. Mech.* **2021**, *21*, e202100137. [[CrossRef](#)]
44. Mollaioli, F.; Lucchini, A.; Cheng, Y.; Monti, G. Intensity measures for the seismic response prediction of base-isolated buildings. *Bull. Earthq. Eng.* **2013**, *11*, 1841–1866. [[CrossRef](#)]
45. Jough, F.K.G.; Şensoy, S. Prediction of seismic collapse risk of steel moment frame mid-rise structures by meta-heuristic algorithms. *Earthq. Eng. Eng. Vib.* **2016**, *15*, 743–757. [[CrossRef](#)]
46. Mitropoulou, C.C.; Papadrakakis, M. Developing fragility curves based on neural network IDA predictions. *Eng. Struct.* **2011**, *33*, 3409–3421. [[CrossRef](#)]
47. Morfidis, K.; Kostinakis, K. Seismic parameters' combinations for the optimum prediction of the damage state of R/C buildings using neural networks. *Adv. Eng. Softw.* **2017**, *106*, 1–16. [[CrossRef](#)]
48. Bamer, F.; Shi, J.; Markert, B. Efficient solution of the multiple seismic pounding problem using hierarchical substructure techniques. *Comput. Mech.* **2018**, *62*, 761–782. [[CrossRef](#)]
49. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: tensorflow.org (accessed on 1 October 2021).
50. FEMA 356. *Prestandard and Commentary for the Seismic Rehabilitation of Buildings*; Federal Emergency Management Agency: Washington, DC, USA, 2000.
51. Sediek, O.A.; Wu, T.Y.; McCormick, J.; El-Tawil, S. Collapse Behavior of Hollow Structural Section Columns under Combined Axial and Lateral Loading. *J. Struct. Eng.* **2020**, *146*, 04020094. [[CrossRef](#)]
52. Waskom, M.L. Seaborn: Statistical data visualization. *J. Open Source Softw.* **2021**, *6*, 3021. [[CrossRef](#)]
53. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
54. Bamer, F.; Shirafkan, N.; Cao, X.; Oueslati, A.; Stoffel, M.; Saxcé, G.D.; Markert, B. A Newmark space-time approach in structural mechanics. *Proc. Appl. Math. Mech.* **2021**, *20*, e202000304. [[CrossRef](#)]
55. Kingma, D.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.