Resilient Cities and Structures 3 (2024) 20–43

Contents lists available at ScienceDirect

Resilient Cities and Structures

journal homepage: www.elsevier.com/locate/rcns

Full Length Article

Meta databases of steel frame buildings for surrogate modelling and machine learning-based feature importance analysis

Delbaz Samadian*, Imrose B. Muhit, Annalisa Occhipinti, Nashwan Dawood

School of Computing, Engineering, and Digital Technologies, Teesside University, Middlesbrough TS1 3BX, United Kingdom

ARTICLE INFO

Keywords:
Surrogate models
Meta database
Pushover analysis
Steel moment resisting frames
Sensitivity and explainability analyses
Machine learning

ABSTRACT

Traditionally, nonlinear time history analysis (NLTHA) is used to assess the performance of structures under future hazards which is necessary to develop effective disaster risk management strategies. However, this method is computationally intensive and not suitable for analyzing a large number of structures on a city-wide scale. Surrogate models offer an efficient and reliable alternative and facilitate evaluating the performance of multiple structures under different hazard scenarios. However, creating a comprehensive database for surrogate modelling at the city level presents challenges. To overcome this, the present study proposes meta databases and a general framework for surrogate modelling of steel structures. The dataset includes 30,000 steel moment-resisting frame buildings, representing low-rise, mid-rise and high-rise buildings, with criteria for connections, beams, and columns. Pushover analysis is performed and structural parameters are extracted, and finally, incorporating two different machine learning algorithms, random forest and Shapley additive explanations, sensitivity and explainability analyses of the structural parameters are performed to identify the most significant factors in designing steel moment resisting frames. The framework and databases can be used as a validated source of surrogate modelling of steel frame structures in order for disaster risk management.

1. Introduction

Structural engineering faces challenges in mitigating natural hazards like earthquakes and floods, requiring a comprehensive understanding of building vulnerability. Assessing the resilience of individual buildings is impractical due to the computational resources needed for continuous analysis. Disaster risk management requires a city-wide approach.

For example, performance-based earthquake engineering (PBEE) provides a systematic probabilistic approach to characterising earthquakes, assessing their impact on building response, and mapping engineering demand parameters to decision variables [1–3]. However, despite recent advancements in finite element modelling, nonlinear time history analyses (NLTHAs) still impose a significant computational burden. While studies such as Montuori et al. [4] and Lin and Miranda [5] proposed a simplified yet reliable approach utilising an equivalent single-degree-of-freedom (SDOF) system to estimate the maximum roof displacement of multistorey buildings, the assessment of the exact response of a collection of buildings at a discrete level through NLTHA remains challenging and excessively time-consuming. Moreover, within the framework of PBEE, there are inherent uncertainties stemming from modelling assumptions, material properties, section properties, record-to-record variation, and other factors [6–8]. Incorporating all of these uncertainties in the evaluation of engineering demand parameters (EDPs) adds to the computational demands of PBEE, making it even more challenging to assess the exact response of buildings accurately. Therefore, there is growing interest in adopting a versatile and efficient framework in civil engineering and risk management for accurate building analysis. Surrogate models have gained popularity as a statistical approach, offering various algorithms and strategies. They serve as a viable alternative to NLTHAs in probabilistic seismic risk assessment (PSRA). Surrogate models are trained using input-output parameters from simulation data, efficiently capturing structural responses. By mapping structural and ground motion parameters to EDPs, they streamline the analysis process without extensive NLTHA simulations for each scenario.

2. State-of-the-art on surrogate modelling in structural engineering

In recent years, structural and earthquake engineers have started utilising surrogate modelling techniques to assess resilience and analyse risks associated with assets exposed to different types of natural and even man-made hazards. This section presents a comprehensive review of previous studies on the use of metamodelling in the field of structural engineering. In this regard, several models have been proposed to address a single hazard scenario, where in most cases, earthquake is regarded as the most common natural hazard. As an example, Zhong

* Corresponding author.
E-mail address: D.samadian@tees.ac.uk (D. Samadian).

https://doi.org/10.1016/j.rcns.2023.12.001
Received 10 August 2023; Received in revised form 30 November 2023; Accepted 16 December 2023

Table 1
Literature review of the recent studies on surrogate modelling in the field of structural engineering.

Study	Year	Case study	Best Surrogate models	Hazard
Zhong et al. [9]	2022	12-storey building	PLoM	Earthquake
Sudret and May [10]	2013	3-storey building	Polynomial chaos expansion	Earthquake
Gidaris et al. [11]	2015	4-storey building	Kriging	Earthquake
Gidaris et al. [12]	2016	4-storey building	Kriging	Earthquake
Du and Padgett [13]	2020	Highway bridge	PLSR-ANN	Earthquake
Vaseghiamiri et al. [14]	2020	6- and 12-storey buildings	SDOF	Earthquake
Gudipati and Cha [15]	2021	6- and 4-storey buildings	ANN	Earthquake
Aristizábal and Caballero [16]	2019	A 9 m embankment	ANN	Earthquake
Dang-Vu et al. [17]	2021	Piloti-type low-rise building	DNN	Earthquake
Cavalagli et al. [18]	2019	Church	PC	Earthquake
Micheli et al. [19]	2022	40-storey building	Kriging	Wind
Micheli et al. [20]	2019	39-storey building	RBF	Wind
Micheli et al. [21]	2020	39-storey building	AWN	Wind
Javadian et al. [22]	2018	3-storey building	ANN	Vehicle collision
Zhang et al. [23]	2023	Train–bridge system	AS	Wheel load of train
Zheng et al. [24]	2023	42-storey building	BP-ANN	Earthquake and Wind
Xing et al. [25]	2022	A high-rise building	Kriging	Earthquake and Wind
Esteghamati and Flint [26]	2021	3–6 storey buildings	SVR	Earthquake and Global warming

et al. [9] employed a surrogate model called probabilistic learning on manifolds (PLoM) to estimate the seismic response of a 12-storey building. The surrogate model was able to incorporate factors such as ground motion uncertainty, design variability, and modelling variability to accurately account for parameters such as maximum storey drift ratio and peak floor acceleration. Sudret and May [10] employed a polynomial chaos expansion technique to estimate seismic fragility curves for a 3-storey building. The researchers used a surrogate model to approximate EDPs of interest and subsequently computed the corresponding structural responses under various seismic records. Gidaris et al. [11] and Gidaris et al. [12] employed kriging for surrogate modelling. The former tried to establish a connection between the structural response and ground motion parameters for the seismic risk assessment of a 4-storey building while the latter aimed at floor isolation systems (FIS) installed on the second floor to protect a computer server. In their study, Du and Padgett [13] investigated various surrogate models for the mean response hypersurface and different models for correlated model errors in the context of a typical three-span concrete-girder highway bridge in the United States. They compared different approaches and concluded that partial least squares regression (PLSR) and artificial neural network (ANN) were the most effective for capturing systematic trends in the mean response. Vaseghiamiri et al. [14] introduced a surrogate single-degree-of-freedom (SDOF) model specifically designed for special steel moment resisting frame (SMRF) buildings. This surrogate model offers a means to estimate the probability distribution of the roof drift ratio for SMRFs, providing accurate results while keeping computational costs low. Gudipati and Cha [15], Aristizábal and Lopez-Caballero [16], and Dang-Vu et al. [17] explored the use of different surrogate models and finally came up with the ANN and deep neural network (DNN) as the best surrogate models for their investigation, respectively. By using a polynomial chaos (PC) based surrogate model, Cavalagli et al. [18] were able to monitor the structural health of the Basilica of Santa Maria degli Angeli in Italy more efficiently and effectively. The model allowed for real-time or near-real-time assessments of the structural response, enabling early detection of potential issues or changes in the structural integrity.

In other studies, wind is considered the most dominant hazard. Micheli et al. [19] introduced a multi-surrogate model approach, which involves dividing the structural system into several sub-systems. Each sub-system is then assigned its own dedicated surrogate model, which handles a relatively small number of inputs and outputs. Micheli et al. [20] focused on quantifying the risk associated with a 39-storey building that was equipped with damping devices. They employed fragility functions to assess the building's vulnerability to different hazard levels. To efficiently estimate the responses required for the fragility functions, they utilised a metamodel known as a radial basis function (RBF). Micheli et al. [21] investigated the application of kriging surrogate and adaptive wavelet network (AWN) models as surrogate models for a 39-storey building equipped with semi-active friction devices under wind excitation. Javadian et al. [22] combined neural networks with the shifted integration-Gauss technique to develop a surrogate model for the collapse assessment of a steel-framed structure under vehicle impact loads. This integrated approach allowed for the evaluation of the collapse potential of the structure more efficiently. They leveraged neural networks and created a powerful computational tool capable of capturing the complex relationships between the structural response and vehicle impact loads. Zhang et al. [23] conducted a comparison between adaptive sampling (AS) and one-stage sampling (OS) surrogate models based on Gaussian process regression (GPR). The objective was to estimate the maximum dynamic response of a train-bridge system. The findings of the study demonstrated that the AS approach outperformed the OS approach in terms of accurately predicting the failure probability of trains on the bridge.

On the other hand, readers can find few studies on surrogate models for multi-hazard scenarios. Zheng et al. [24] utilised backpropagation (BP) and ANN techniques to develop a surrogate model for tall buildings that are subjected to the combined effects of earthquakes and winds. The researchers considered various uncertainties, ranging from aleatory to epistemic, in the development of these surrogate models. Xing et al. [25] developed a surrogate model via kriging metamodels for high-rise buildings equipped with different outrigger systems exposed to multi-hazard scenarios of earthquake and wind and the metamodel was then used to capture the fragility parameters of buildings. Esteghamati and Flint [26] focused on 720 mid-rise concrete office buildings in the USA. They aimed to assess the seismic resiliency and seismic-induced embodied carbon footprint of these buildings. To accomplish this, the researchers explored the use of five different surrogate models: multiple regression, random forest, extreme gradient boosting, support vector machine, and k-nearest neighbours. Table 1, tabulates the recent studies on surrogate modelling in civil engineering practices.

The current research was motivated by several factors. Firstly, it was observed from the existing literature that determining the input parameters and selecting the data points for constructing a surrogate model, also known as the design of experiments (DOE) or sampling, is a crucial step. The accuracy and precision of the surrogate model depend on the number of samples used. However, the process of evaluating these data points involves running expensive simulations, resulting in a significant computational burden. To address this challenge, appropriate sampling strategies are needed to maintain the quality of the surrogate model while minimizing the sampling cost [27]. In this study, the au-

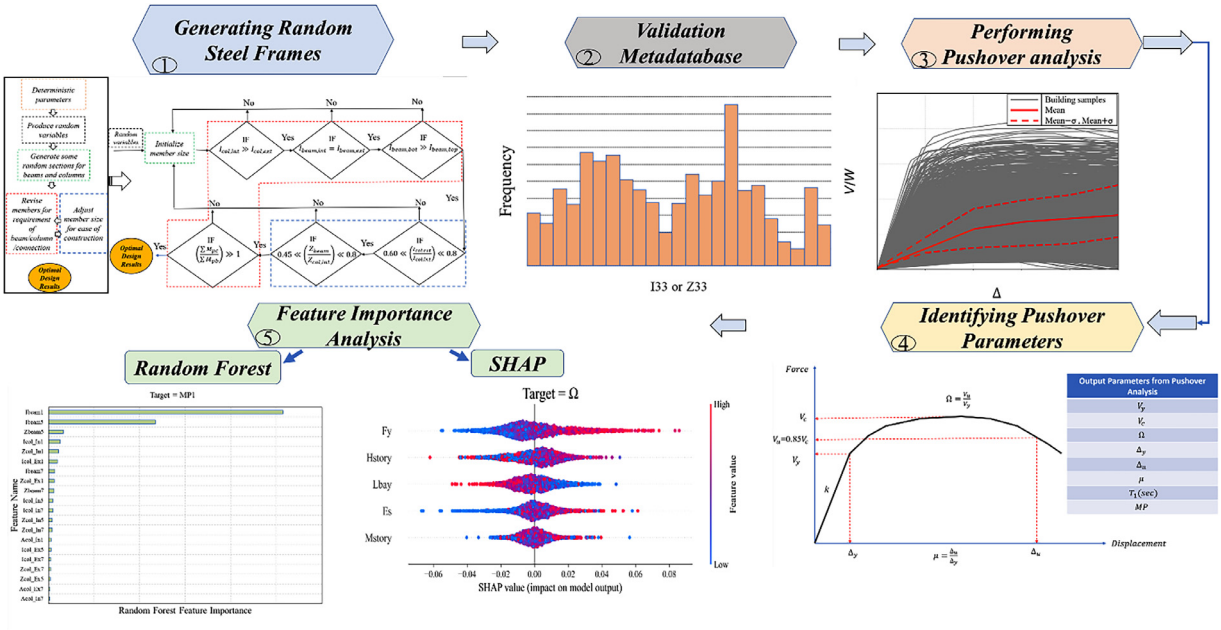


Fig. 1. Flowchart of methodology for generating meta databases and sensitivity analysis.

thors aimed to develop meta databases for surrogate models of special steel moment resisting frames (SMRFs) that could be applied to various regions with minimal modifications. Instead of designing a large number of SMRFs, which would require extensive computational resources, or relying on overly simplified representative models, this paper proposes a trade-off approach. The framework established preliminary criteria for designing buildings and then generated the DOE randomly on a large scale. This approach allowed for the construction of surrogate models without incurring excessive computational costs.

Another motivation for this research was the need to identify the important factors and eliminate redundant ones, particularly when dealing with a large feature set. Since the selection of input parameters significantly affects the accuracy and reliability of surrogate models, it is crucial to determine which parameters have a significant impact on the nonlinear response of the meta database. To address this issue, machine-learning-based sensitivity analysis techniques were employed. By applying these methods, the influence of different random parameters on the response of the meta database is assessed. This analysis allowed them to identify the key factors that played a vital role in shaping the nonlinear response behaviour.

The third motivation for this research was to move away from conventional deterministic assumptions for material, geometry, and capacity parameters in generating the meta database of SMRFs. Instead, almost all of the crucial parameters are considered random variables. This approach aimed to reduce uncertainties associated with the properties of buildings and enhance the robustness and reliability of the final surrogate model for end-users. By treating these parameters as random variables, the inherent variability and uncertainty present in real-world structural systems were adequately captured. This approach acknowledges the fact that material properties, geometric dimensions, and capacity parameters can exhibit significant variations in practice. By incorporating these random parameters into the meta database generation process, the resulting surrogate model will become more representative of the actual structural behaviour and can provide more reliable predictions.

Lastly, the inclusion of low-rise, mid-rise, and high-rise buildings in the meta database allows for the exploration of the full spectrum of structural responses, from simpler systems to more complex and tall structures. This not only enhances the versatility of the surrogate model but also improves its applicability to a wide range of practical scenarios.

The research presented in this paper provides a valuable framework and workflow for developing suitable surrogate models for SMRFs. Fig. 1 illustrates the workflow that was followed throughout this study. It serves as a template that can be adapted and applied to other types of structures and different locations of interest.

The first step is focused on the methodology employed to generate the meta database. This section describes the process of creating a comprehensive and diverse set of structural models, considering various random parameters to capture the uncertainty in the properties of the SMRFs. The meta database serves as the foundation for the subsequent analyses and surrogate model development. The second step tries to show the meta databases are validated and reliable enough to be used in further steps. In steps 3 and 4, the pushover analysis results and selected pushover parameters are discussed. These parameters are crucial in characterizing the nonlinear response behaviour of the SMRFs and are used as inputs in the sensitivity analysis conducted in the study. The pushover analysis provides valuable insights into the structural response under lateral loads and helps identify key performance indicators for further investigation. The final step presents the results of the sensitivity analysis. This analysis aims to identify the most influential parameters that govern the nonlinear response of the SMRFs. By employing machine-learning-based sensitivity analysis techniques including Random Forest (RF) and SHAP (SHapely Additive exPlanations), the study determines the relative importance of different random parameters in the meta database. This information is crucial for understanding the key factors that contribute to the variability and uncertainty in the structural response. Finally, in Section 4, the research findings are discussed, and conclusions are drawn based on the results obtained. This paper aligns with two United Nations Sustainable Development Goals (UNSDGs), namely Goal 9 and Goal 11. These goals are set to promote the establishment of resilient infrastructures and to enhance the safety, resilience, and sustainability of our cities and settlements.

3. Motivation for selecting SMRF and RBS connection

Experimental evidence has provided confirmation that SMRFs exhibit exceptional ductility as lateral load resisting systems when appropriately detailed [28]. To achieve this desirable ductile behaviour, SMRFs are typically designed to concentrate yielding in plastic hinges located at the end of the beams throughout the entire height of the building [29]. This design approach aims to prevent plastic hinging of the

columns, which could result in an undesirable soft storey. The remaining components of the SMRF are designed to remain elastic, either by utilising the capacity design criteria or by employing the over-strength seismic load. By adopting these measures, stable plastification can be achieved without compromising the structural integrity, thus minimising the risk of collapse [28].

However, past seismic events such as the 1994 Northridge and 1995 Kobe Earthquakes revealed unexpected brittle fractures occurring at the welded beam-to-column connections in SMRFs [30,31]. These fractures led to significant repair costs, amounting to a combined total of $280 billion [32,33], in order to rectify the damage inflicted upon the steel frames [34]. Consequently, in the aftermath of these earthquakes, the reduced beam section (RBS) connection emerged as a promising solution and was swiftly incorporated into American and European design standards [35]. The RBS connection is characterised by a reduction in the beam section, allowing for the formation of inelastic deformations in the reduced section. This connection type offers notable advantages, including increased rotation capacity and enhanced energy absorption capacity when subjected to seismic actions [36,37].

Moreover, prior research conducted by Guan et al. [38] has suggested that in terms of the archetype concept for seismic performance of building structures, SMRFs with RBS connections are allowed without any limitations in ASCE 7–16 [39] for buildings that exceed a height of 160 ft (48.77 m) in regions with high seismic activity. Another benefit of SMRFs with RBS connections is that they eliminate the need for diagonal braces, resulting in a simplified structural model in software and reduced forces on foundations compared to other structural systems. Consequently, this leads to more cost-effective sub-structure systems [40]. Due to these advantages, SMRFs can be extensively utilised in the construction of low- and mid-rise residential and commercial buildings, especially when software simulations require the analysis of numerous building cases to systematically capture variations in key structural characteristics. As a result, considering the aforementioned advantageous characteristics of SMRFs and RBS connections, this study chose these as the candidates for generating a meta database for buildings.

4. Generating meta databases

4.1. Description of reference buildings

The study introduces a framework for generating a comprehensive database of SMRFs, addressing the challenge of a limited number of samples. Three prototype buildings represent different types of SMRFs, serving as reference structures for database generation. The first 2-storey SMRF captures low-rise characteristics, the second 8-storey SMRF focuses on mid-rise design and performance, and the third 20-storey SMRF reflects taller structures.

They are all located in Los Angeles Downtown (LADT), USA. The buildings are obtained from federal emergency management agency guidelines for quantification of building seismic performance factors (FEMA P695) [41] and their designs are in line with the American Society of Civil Engineers (ASCE 7–16) [39], and the seismic provisions of American Institute of Steel Construction (AISC) (AISC 2010) [42], with seismic design category of D_{max}. The buildings mainly vary only in their height and have natural periods (T_1) of 0.92, 2.17, and 4.01 s with total heights of 8.40 m, 31.80 m, and 35 m for the 2-,8-, and 20-storey SMRFs, respectively. The extracted T_1 values are very close to those of the reference building specified by Table 6-4 of FEMA-P695 (0.91, 2.29, and 4.47 s for 2-, 8-, and 20-storey buildings, respectively). Also, for all SMRFs, the first storey's height is 4.5 m and other stories have a typical height of 3.90 m. Furthermore, the gravity loads include a dead load of 4.78 kN/m^2 and a live load of 2.38 kN/m^2 applied to all floors. The cladding load is applied as a perimeter load of 1.20 kN/m^2. Also, the size of beams and columns for three buildings are those in Tables 2–4. Fig. 2 illustrates the elevation and the plan view of the archetype structures. Other characteristics of the buildings can be found in [43].

Table 2
Member sizes for beams and columns for reference 2-storey building.

Storey	Beam Size	Exterior Column Size	Interior Column Size
1	W30 × 132	W24 × 131	W24 × 162
2	W16 × 31	W24 × 131	W24 × 162

Table 3
Member sizes for beams and columns for the 8-storey reference building.

Storey	Beam Size	Exterior Column Size	Interior Column Size
1	W30 × 108	W24 × 131	W24 × 162
2	W30 × 116	W24 × 131	W24 × 162
3	W30 × 116	W24 × 131	W24 × 162
4	W27 × 94	W24 × 131	W24 × 162
5	W27 × 94	W24 × 131	W24 × 131
6	W24 × 84	W24 × 131	W24 × 131
7	W24 × 84	W24 × 94	W24 × 94
8	W21 × 68	W24 × 94	W24 × 94

Table 4
Member sizes for beams and columns for the 20-storey reference building.

Storey	Beam Size	Exterior Column Size	Interior Column Size
1	W33 × 169	W14 × 426	W24 × 335
2	W33 × 169	W14 × 426	W24 × 335
3	W33 × 169	W14 × 426	W24 × 335
4	W33 × 169	W14 × 426	W24 × 335
5	W33 × 169	W14 × 398	W24 × 335
6	W33 × 169	W14 × 398	W24 × 335
7	W33 × 169	W14 × 370	W24 × 335
8	W33 × 169	W14 × 370	W24 × 335
9	W33 × 141	W14 × 311	W24 × 279
10	W33 × 141	W14 × 311	W24 × 279
11	W33 × 141	W14 × 283	W24 × 250
12	W33 × 141	W14 × 283	W24 × 250
13	W33 × 141	W14 × 233	W24 × 250
14	W33 × 141	W14 × 233	W24 × 250
15	W30 × 108	W14 × 159	W24 × 162
16	W30 × 108	W14 × 159	W24 × 162
17	W30 × 108	W14 × 132	W24 × 162
18	W30 × 108	W14 × 132	W24 × 162
19	W24 × 62	W14 × 132	W24 × 103
20	W24 × 62	W14 × 132	W24 × 103

4.2. Modelling reference buildings

The finite element models of the reference buildings are developed in the open-source software OpenSees [44] as per state-of-the-art considerations. The beam-column connections are considered RBS connections in accordance with the recommendation of AISC-341–16 [42] and AISC-358–16 [37]. In OpenSees, to capture hinge properties of RBS, uniaxialMaterial Bilin is borrowed. As per Fig. 3, the beam and column elements are modelled as linear elastic elements and the hysteretic behaviour of the beam-column connections is modelled using Ibarra-Medina-Krawinkler (IMK) bilinear model using nonlinear rotational springs provided at the ends of the elements [45]. The cyclic degradation in flexural strength and stiffness of steel components under cyclic loading can be effectively captured by the IMK bilinear model. To confine the modified IMK model, a backbone curve is introduced, as shown in Fig. 4(a). Fig. 4(b) presents the initial backbone curve in a simplified manner by eliminating the cyclic phase. The characteristics of this curve are defined by a set of rules that govern the hysteretic behaviour and deterioration rate. Within this figure, the effective yield strength and rotation are denoted by M_y and θ_y, respectively, while the effective stiffness is represented by $K_e = {M_y}/{\theta_y}$. For monotonic loading, the capping strength and its associated rotation are illustrated by M_c and θ_c, respectively. The pre-capping rotation capacity and post-capping rotation capacity are denoted by θ_p and θ_{pc}, respectively. Furthermore,

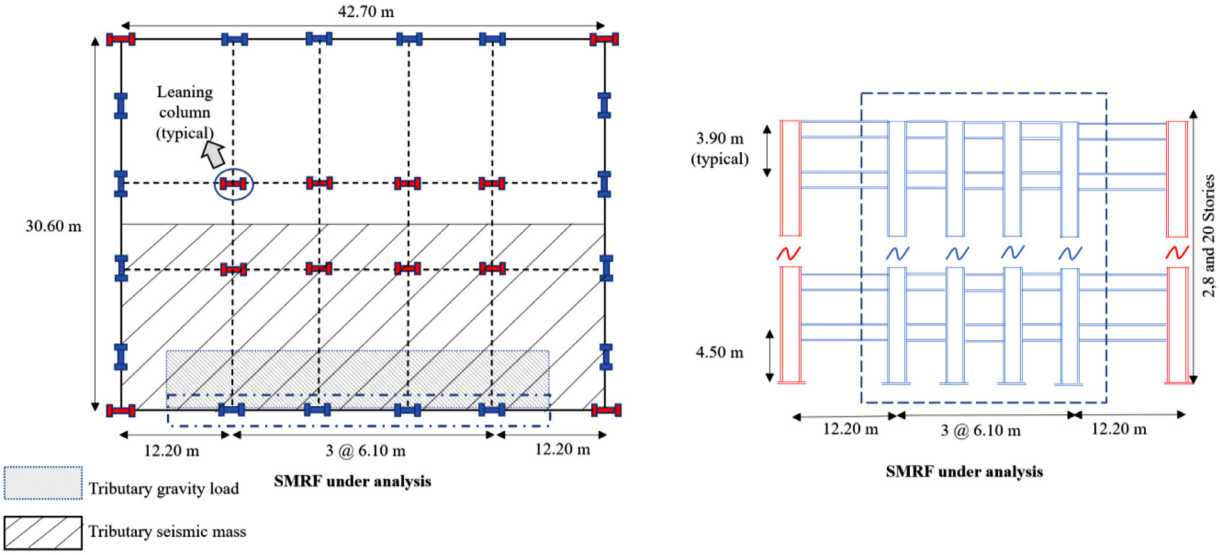


Fig. 2. Plan and elevation views for reference building of 2-storey, 8-storey, and 20-storey buildings.

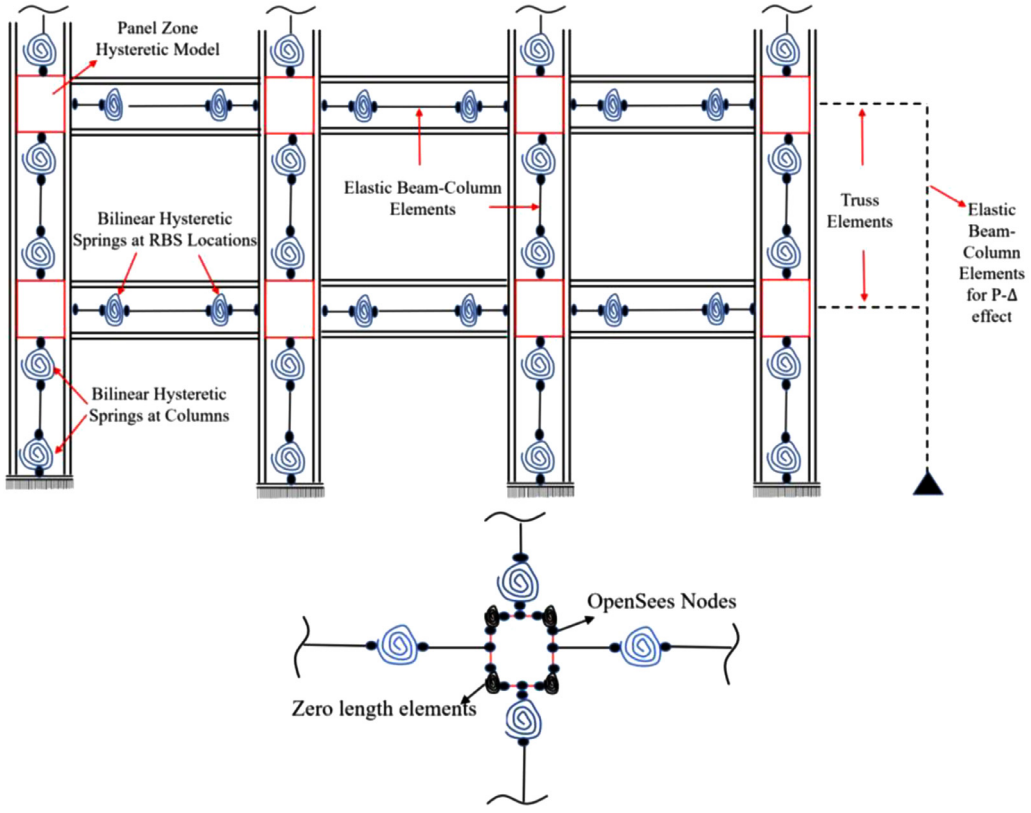


Fig. 3. Details of modelling in OpenSees.

the residual strength and ultimate rotation capacity are represented by M_r and θ_u, respectively. The rate of cyclic deterioration is taken into account through the parameter Λ. In this study, to determine these parameters for different elements such as beams and columns, regression equations calibrated on over 300 experiments on steel wide flange beams, as revealed by ASCE/SEI 41–17 [46] and Lignos et al. [47,48,49], are utilised. The panel zones are simulated as a hinge parallelogram assembly connected by rigid elements, having a nonlinear spring in one of the corners to capture shear distortion [50]. Finally, to consider P-Δ effects, a leaning column is modelled with Elastic Beam-Column Element and loaded with half of the plan-building's gravity on each floor, connected to the SMRFs via Truss elements [51]. Table 5. shows, also, the first three modes' participation factors and natural periods of the SMRFs. In order to validate the accuracy of the models created in OpenSees, the pushover curves of the reference buildings are compared with those provided by the NEHRP guidelines [41]. According to [41], a pushover analysis is performed for each archetype, utilising a lateral load distribution that aligns with the fundamental mode shape and mass distribution of the structure. The comparison is made in terms of the ratio of base shear to the total weight of the building (y-axis) against the drift ratio

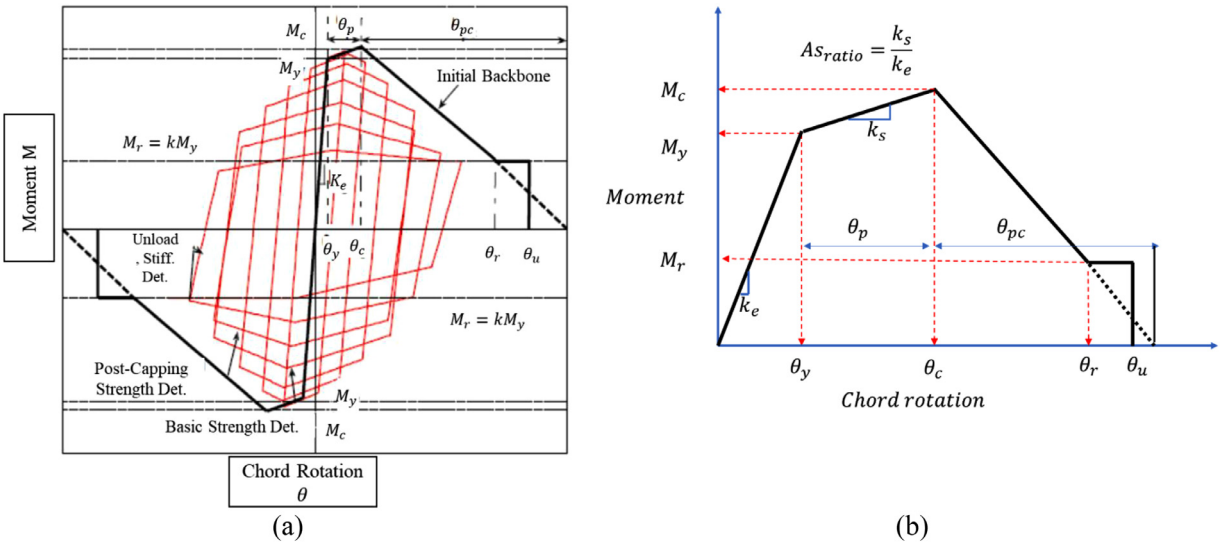


Fig. 4. Moment-rotation relationship for plastic hinges of moment frame: (a) with cyclic phase [adapted from [47]];(b) without cyclic phase.

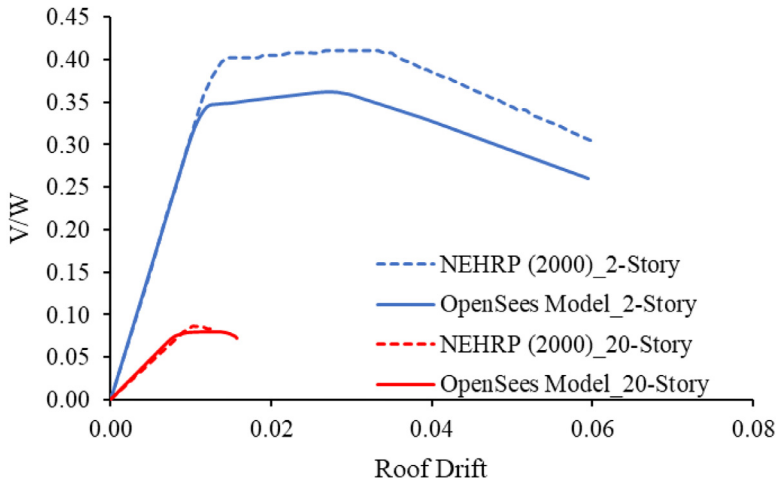


Fig.5. Comparison of pushover curves extracted by OpenSees with NEHRP (2000).

Table 5
SMRFs' modal properties of reference buildings.

	2-storey	8-storey	20-storey
First period (*sec*)	0.9161	2.17048	4.00521
Second period (*sec*)	0.2179	0.76007	1.452728
Third period (*sec*)	0.1477	0.41665	0.836197
First mode mass participation	94.57	81.35	80.56
Second mode mass participation	100	93.75	92.74
Third mode mass participation	100	97.34	96.56

of the building (x-axis). Fig. 5 illustrates the comparison results, demonstrating a close agreement between the pushover curves obtained from OpenSees and the ones extracted from [41] for both the 2-storey and 20-storey buildings. This finding indicates that the models developed in OpenSees accurately capture the behaviour of the reference buildings as depicted by [41]. However, it should be noted that the NEHRP guideline only provides pushover curves for 2-, 4-, and 20-storey buildings. As a result, the authors were unable to directly compare the pushover curves for the 8-storey reference building. Nonetheless, the good agreement observed for the 2-storey and 20-storey buildings enhances the confidence in the accuracy of the OpenSees models and their ability to simulate the structural response of the reference buildings.

4.3. Types of random and deterministic parameters

To generate meta databases for steel frame buildings, the first step involves defining different parameters for the sampling process. These parameters can be categorised into three types: random, deterministic, and casual random parameters.

Random parameters: they include beam-sections and column-sections; yield strength of steel (F_y); elastic module (E_s); height of typical stories ($H_{story}(typical)$); mass value ($Mass$); ratio of length of bays to the height of first storey ($\frac{L_{bay}}{H_{story}}$). Table 6 provides the nominal values and distribution functions for these random parameters. It should be noted that for H_{story}, its upper and lower bounds are selected based on engineering judgment.

Deterministic Parameters: Deterministic parameters consist of the number of stories and the number of bays. These parameters remain constant during the sampling process for each reference building.

Casual Random Parameters: Casual random parameters are those that change alongside random parameters. They include the height of the first storey ($H_{first-story}$); pre-peak rotation (θ_p); post-peak rotation (θ_{pc}); ratio of elastic stiffness to the pre-peak stiffness (As_{ratio}) based on the Modified Ibarra-Medina-Krawinkler (IMK) hysteretic model shown in Fig. 4(b); flange width (b_f); flange thickness (t_f); web height (h_w);

Table 6
Random parameters for the sampling process.

Random Variable	Distribution	Mean/Nominal or lower and upper bond	Coefficient of Variation	Reference
$Mass$	Normal	1.05	0.10	Fayaz and Zareian [52]
F_y	Log-normal	1.18	0.13	Ellingwood et al. [53]
E_s	Log-normal	1.18	0.13	Ellingwood et al. [53]
$\frac{L_{bay}}{H_{story}}$	Uniform	Lower bound=1.25 Upper bound=2	—	Vaseghiamiri et al. [14]
H_{story}	Uniform	Lower bound= 0.75 × nominal Upper bound= 1.25 × nominal	—	Engineering judgment

Table 7
Geometric, material, and spring parameters for 2-storey reference buildings.

Input Parameters		
Geometric Parameters	Material Parameters	Spring Parameters
H_{story}	E_s	$\theta_p\_Col_{Ex1}$, $\theta_p\_Col_{Ex2}$
$\frac{L_{bay}}{H_{story}}$	F_y	$\theta_{pc}\_Col_{Ex1}$, $\theta_{pc}\_Col_{Ex2}$
Col_{Ex1}, Col_{Ex2}	$Mass$	$As_{ratio}\_Col_{Ex1}$, $As_{ratio}\_Col_{Ex2}$
$Acol_{Ex1}$, $Acol_{Ex2}$	–	$\theta_p\_Col_{Int1}$, $\theta_p\_Col_{Int2}$
$Icol_{Ex1}$, $Icol_{Ex2}$	–	$\theta_{pc}\_Col_{Int1}$, $\theta_{pc}\_Col_{Int2}$
Col_{Int1}, Col_{Int2}	–	$As_{ratio}\_Col_{Int1}$, $As_{ratio}\_Col_{Int2}$
$Acol_{Int1}$, $Acol_{Int2}$	–	$\theta_p\_Beam_1$, $\theta_p\_Beam_2$
$Icol_{Int1}$, $Icol_{Int2}$	–	$\theta_{pc}\_Beam_1$, $\theta_{pc}\_Beam_2$
$Beam_1$, $Beam_2$	–	$As_{ratio}\_Beam_1$, $As_{ratio}\_Beam_2$
$Abeam_1$, $Abeam_2$	–	–
$Ibeam_1$, $Ibeam_2$	–	–
Number of stories	–	–
Number of bays	–	–

web thickness (t_w); the moment of inertia (I_{sec}); plastic section module (Z_{sec}); cross sectional area (A_{sec}). These parameters can be categorised into geometric, material, and spring parameters. By modifying the main random parameters, new buildings can be generated. When the main random parameters change, the associated cross-sectional properties and hysteretic parameters of beams or columns are updated accordingly, resulting in new parameters known as casual random parameters. These casual random parameters implicitly change throughout the process and can be considered as the children of the main random parameters.

To illustrate the parameters involved in generating random buildings, Table 7 provides an example for 2-storey buildings. In the case of random buildings, it is assumed that for all three types of buildings (low-rise, mid-rise, and high-rise), the exterior columns at each storey level have the same cross-sectional dimensions, as the interior columns. Similarly, all the beams have the same properties at each storey level. For mid-rise and high-rise buildings, the columns and beams are grouped every two stories. Therefore, "Col_{Ex1}" refers to the exterior columns at the first storey, while "Col_{Int1}" represents the interior columns at the same level.

4.4. Python-based module for generating random buildings

After defining random and deterministic parameters, a Python-based module is developed to generate meta databases from the three reference buildings. These datasets will be used for developing surrogate models. The module is designed to generate random buildings that satisfy specific criteria. Two main criteria are considered for generating steel moment resisting frame (SMRF) buildings, ensuring that the final buildings are highly validated for any usage in surrogate modelling. The first criterion is the resiliency criteria, which are as follows:

1) $(\frac{\sum M_{pc}}{\sum M_{pb}}) \geq 1$: $\sum M_{pc}$ the sum of the moments in the column above and below the joint at the intersection of the beam and column centerlines and $\sum M_{pb}$the sum of moments in the beams at the intersection of the beam and column centerlines). According to AISC 341 Chapter E3 [42], the column-beam moment ratio should be greater than 1.0 to meet the strong-column-weak-beam (SCWB) design criterion. Therefore, this criterion must be satisfied at each intersection of beams and columns in the buildings during the generation process.

2) $I_{col,int} \geq I_{col,ext}$: It is a common assumption in building design that the interior columns should have greater strength than the exterior columns [38].

3) $I_{beam,int} = I_{beam,ext}$: FEMA-P695 [41] recommends that interior and exterior beams at each storey level can have the same section size.

4) $I_{beam,bot} \geq I_{beam,top}$ and 5) $I_{column,bot} \geq I_{column,top}$: According to the FEMA-P695 [41] guidelines, the sizes of beams and columns in the lower stories should not be smaller than those in the upper stories. This ensures that the structural members have sufficient strength and capacity as the building progresses vertically.

6) $I_{col,int} = I_{col,int}$ and $I_{col,ext} = I_{col,ext}$: In the design of steel frame buildings, it is common practice to adopt the same size for two exterior columns at each storey level, as well as for two interior columns at each storey level. Similarly, in mid-rise and high-rise buildings, the same member size is typically used in every two adjacent stories. Additionally, it is common to have deeper columns in the lower stories to accommodate splice connections, while the beams at the lower floor levels are typically deeper and stronger compared to the beams in the upper stories. These design considerations ensure proper structural integrity and load distribution throughout the building [39,42,29].

The second set of criteria, known as the **Practically criteria**, further refines the preliminary design of the SMRF buildings by imposing additional constraints on the beam and column section sizes. These criteria are as follows:

The ratio of the plastic section modulus of the beam (Z_{beam}) to the plastic section modulus of the interior column ($Z_{col,int}$) should be between 0.45 and 0.8 [38,29].

The ratio of the moment of inertia of the exterior column ($I_{col,ext}$) to the moment of inertia of the interior column ($I_{col,int}$) should be between 0.6 and 0.8 [54].

4.5. Workflow of generating random SMRFs

Fig. 6 illustrates the workflow for generating random SMRFs based on the reference buildings. The process involves several steps to ensure the generation of structurally valid and compliant buildings. The workflow can be summarised as follows:

Selection of reference buildings: The appropriate reference buildings (2-storey, 8-storey, and 20-storey) are chosen as representatives of low-rise, mid-rise, and high-rise structures, respectively.

Initialization of beam and column sizes: The sizes of beams and columns are initially assigned based on the selected reference buildings.

Checking and verification of sizes: The sizes of beams and columns, as well as their connections, are checked to ensure they meet the relevant strength requirements and satisfy preliminary design criteria. If any discrepancies or issues are found, the member sizes are revised accordingly.

Combination with random parameters: The final columns and beams, which have been verified and adjusted, are combined with random parameters. These random parameters may include variations in material properties, geometric dimensions, or other relevant factors.

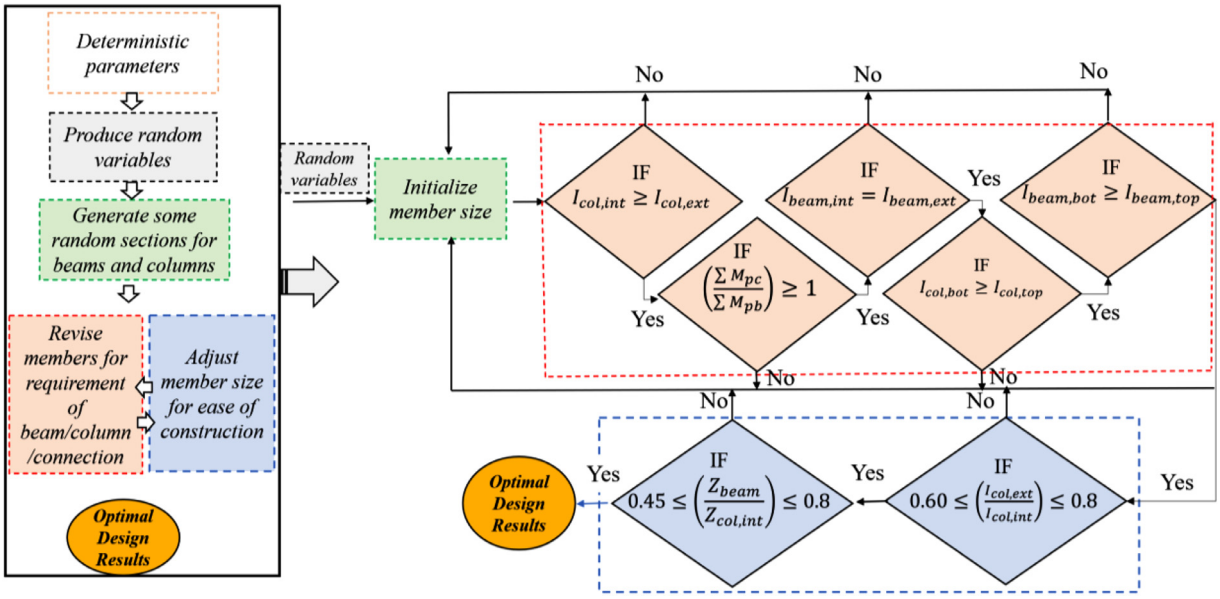


Fig. 6. Framework for generating random SMRF samples from reference buildings.

Creation of the final random SMRF building: By combining the revised member sizes with the random parameters, the final random SMRF building is generated. This building is a representative example of a structurally valid SMRF that satisfies the defined criteria and can be used for further analysis and surrogate modelling.

Each of these steps plays a crucial role in ensuring the generation of random SMRF buildings that are both valid and suitable for subsequent analyses and modelling. The details of each step are elaborated further in the following sections of the study.

Steps 1 and 2: To facilitate the selection of SMRF beam and column sizes, an electronic database of wide flange sections provided by the American Institute of Steel Construction (AISC) is utilised. This database contains a wide range of section sizes for beams and columns. In this study, the section sizes for beams and columns are chosen to be one order higher and lower than those of the reference building. For example, if the reference building requires a minimum beam size of W21×68 and a maximum beam size of W30×108 (as shown in Table 2), the selected beams for random buildings will fall within the range of W18 to W33 according to the AISC database. Therefore, only beam sections from the AISC database within this range are considered as potential options for the SMRF column or beam sections. Furthermore, the adoption of RBS connections imposes additional requirements on the beam and column sections. Specifically, the section depth, weight, and flange thickness of beams must be less than W36, 300 lb/ft, and 1.75 inches, respectively. The column section depths must also be less than W36. Based on these requirements, the original database is filtered to create two sub-databases: one for beam sections and another for column sections. The section sizes in each sub-database are listed in ascending order of the moment of inertia (I_{sec}) and the plastic section modulus (Z_{sec}). An index is assigned to each section, starting from zero and incremented by one, to denote the order of section sizes (with zero representing the weakest section). If two sections have the same I_{sec}, the section with the minimum Z_{sec} is placed first in the sub-database. During the generation of random buildings, the algorithm selects a random index from the sub-database for each beam and column at each storey level. This ensures that the selected beam and column sizes satisfy the criteria outlined by the AISC database and the RBS connection requirements.

Step 3: At this step, a sub-algorithm is employed to optimize the member sizes in order to meet the relevant preliminary requirements, including the resiliency and practical criteria outlined in Fig. 6. Two important coefficients, namely the moment of inertia ratio between the exterior and interior columns and the plastic section modulus ratio between the beam and interior column, are defined as the practical criteria. These criteria are based on a review of industry-generated SMRF designs and are intended to ensure appropriate proportions and compatibility between beam and column sizes. The typical range for $(\frac{I_{col,ext}}{I_{col,int}})$ is 0.6 to 0.8 [54], while the range for $(\frac{Z_{beam}}{Z_{col,int}})$ varies depending on the building height. For buildings with less than 10 stories, the typical range is 0.7 to 0.8, while taller buildings typically have a range of 0.45 to 0.7 [38,29]. These criteria help to ensure structural stability and performance in the designed SMRF buildings. In addition to the practical criteria, all the resiliency criteria mentioned earlier must be satisfied. If any of the criteria are not met at each storey level or each node (intersection of beams and columns), the algorithm will restart by randomly selecting another index for the beams and columns until all the resiliency and practical criteria are fulfilled.

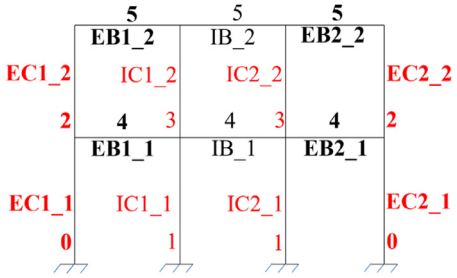


Fig. 7. Indexing columns and beams for a 2-storey reference building.

Step 4: At this stage, the algorithm has successfully generated member sizes for the SMRF that satisfy all design requirements. For example, in Fig. 7, the left exterior column at the first storey ($EC1\_1$), shown with solid red colour, is assigned with the number 0. Accordingly, the right exterior column ($EC2\_1$) should also be assigned the same number to maintain symmetry. To ensure a reasonable design, the left interior column ($IC1\_1$), shown with red colour, must have an equal or higher number than $EC1\_1$ and $EC2\_1$ to indicate a size that is equal or stronger. Similarly, for the second storey, the left exterior column ($EC1\_2$) and the right exterior column ($EC2\_2$) should have the same or stronger sizes compared to $EC1\_1$ and $EC2\_1$, respectively. In this case, they are assigned the number 2 to indicate their size. This numbering

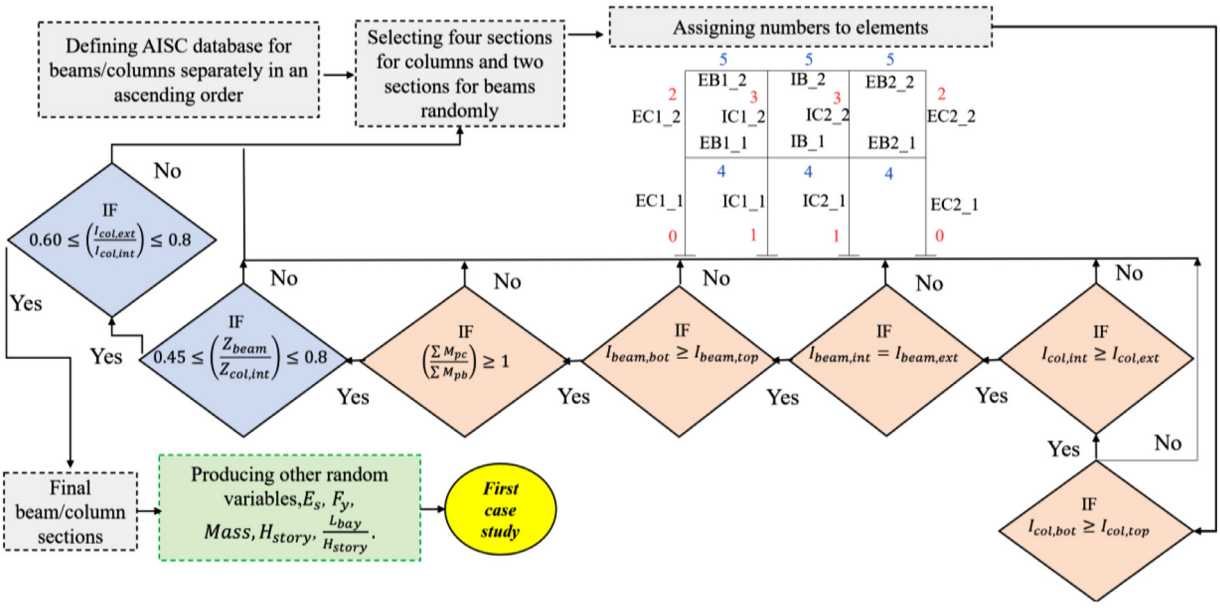


Fig. 8. Exemplifying the first sample SMRF.

scheme ensures that the member sizes are properly assigned and follow a logical progression within the SMRF building. It allows for consistency and maintains the required design strength across different stories and columns.

Step 5: Once the sizes of beams and columns have been finalised, the next step is to generate random parameters as listed in Table 6. These random parameters will be generated randomly and combined with the section sizes of beams and columns to construct 2-dimensional frame structures. In Fig. 8, the generated random parameters are applied to the frame structure, resulting in the complete representation of the SMRF building. This includes the dimensions, connections, and other relevant characteristics of the beams and columns in the structure.

To ascertain the beam and column sizes an electronic database from AISC is utilised instead of a probability density. The sizes of beam and column sections for 10,000 buildings are randomly selected from the AISC database, with the sizes being one order higher and lower than those of the reference building designed by FEMA-P 695 and other guidelines. Then, the original database is filtered to create two sub-databases: one for beam sections and another for column sections. The section sizes in each sub-database are listed in ascending order based on the moment of inertia (I_{sec}) and the plastic section modulus (Z_{sec}). The framework will assign an index to each section, starting from zero and incrementing by one, to indicate the order of section sizes. Next, the algorithm selects a random index from the sub-database for each beam and column at each storey level. This ensures that the chosen beam and column sizes meet the criteria specified by the AISC database and the requirements of RBS connections.

If we envisage the entire process as dicing a dice, each time a dice is rolled, a column section and beam section are selected from the database list of AISC. Then, their associated casual random parameters (I_{sec}; Z_{sec}; A_{sec}) are checked in a loop to determine if they meet all the resiliency and practical criteria. If they do, the selected column and beam sections are finalised, and a new SMRF is generated, as shown in Fig. 9. This process is repeated to generate 10,000 buildings for each reference building.

The same procedure as described earlier is followed for generating random SMRF buildings for 8-storey and 20-storey structures. The main difference is that in these cases, the columns and beams are the same for every two-storey level, whereas, in the 2-storey building, they may vary for each storey level. Similarly, the sub-databases of column and beam sizes are selected based on the corresponding reference buildings. This ensures that the generated random buildings for each structure type adhere to the specific size ranges and design requirements. The final databases for the 2-storey, 8-storey, and 20-storey buildings, which include the selected column and beam sizes along with the associated random parameters, can be accessed here for further analysis and utilization.

5. Validation of database

After generating sample buildings and a meta database for each reference building, to ensure the suitability of the created database for developing surrogate models, several parameters are evaluated beyond resilience and practicality criteria, aiming to maximize the validation of the database. To adhere to page limits, only the figures and tables associated with 2-storey building are presented here; however, the discussion of the additional scenarios (8-storey and 20-storey buildings) is included as usual. Figs. 10 and 11, as the representative figures, dis-

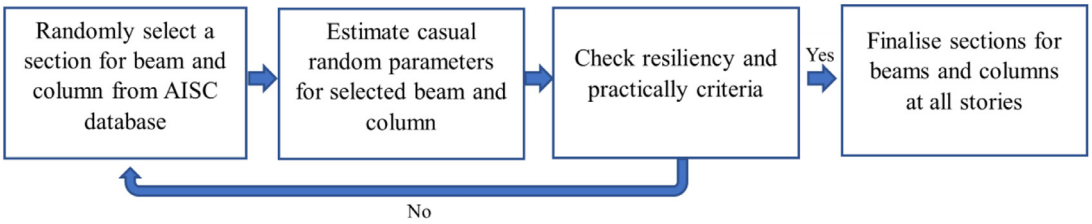


Fig. 9. The process of generating random buildings from AISC database by estimating casual random parameters.

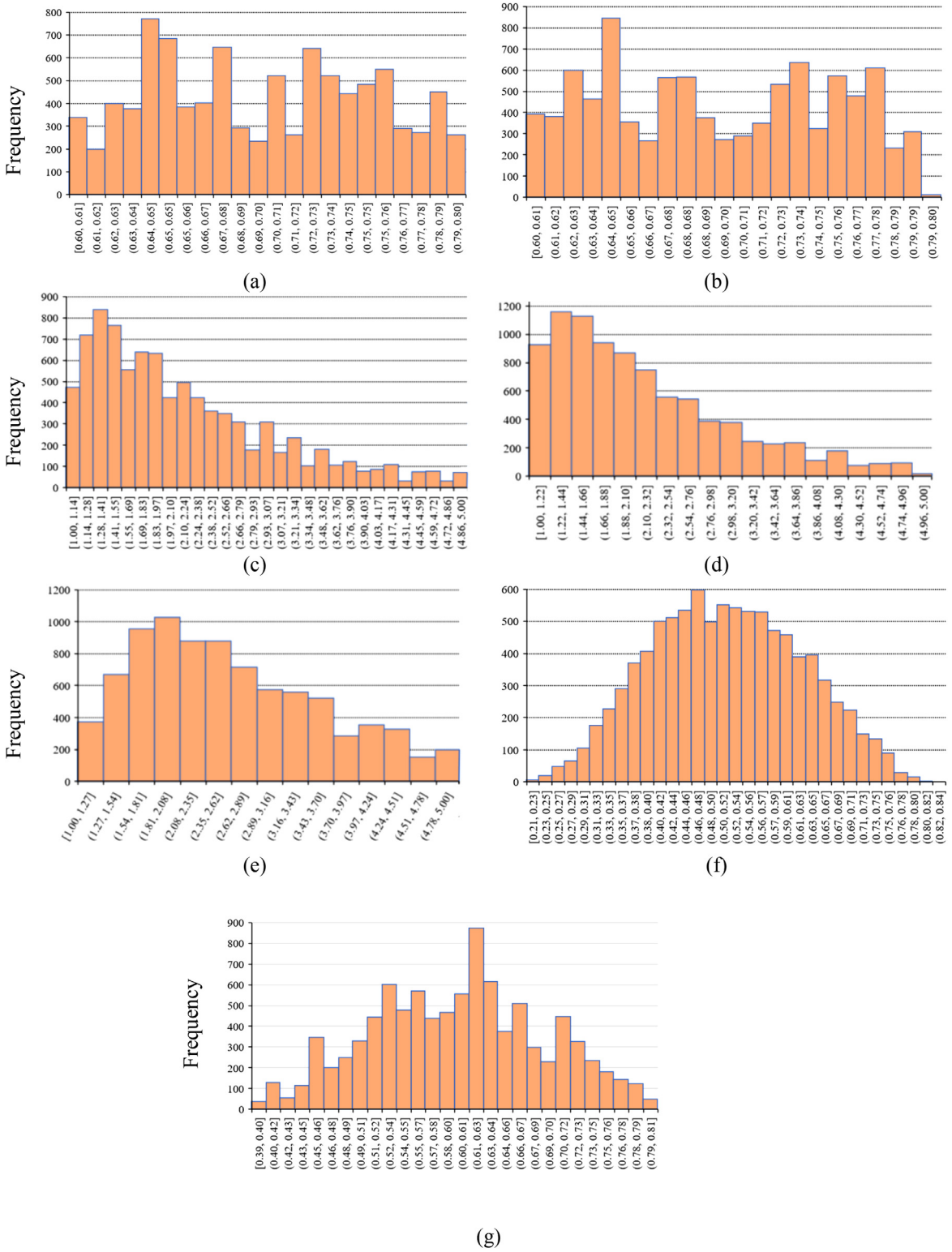


Fig. 10. Distribution of I_{33} for the database of 2-storey reference building, considering various ratios: a) EC1_1/ IC1_1; b) EC1_2/ IC1_2; c) IC1_1/ IC1_2; d) EC1_1/ EC1_2; e) EB1_1/EB1_2;f) EB1_1/(EC1_1+ EC1_2); g) EB1_2/ EC1_2.

play histograms depicting different ratios of I_{33} and Z_{33} for 2-storey reference buildings. The results in Fig. 10(a) demonstrate that the ratio of I_{33} for $EC1\_1$ to $IC1\_1$ falls between 0.6 and 0.8, satisfying the practical criteria. A similar observation can be made for Fig. 10(b), which shows the ratio of $(\frac{I_{33EC1\_2}}{I_{33IC1\_2}})$. Both practical criteria are well met for the 2-storey reference building. Furthermore, Fig. 10(c) illustrates that the internal column at the first storey, $IC1\_1$, is stronger than the one at the second storey, $IC1\_2$. However, for most sample buildings in this database, there is not a significant difference in size between the columns at the first and second stories, confirming that the final buildings are not overdesigned. This finding is also supported by the results of Fig. 10(d) for external columns at the first and second stories.

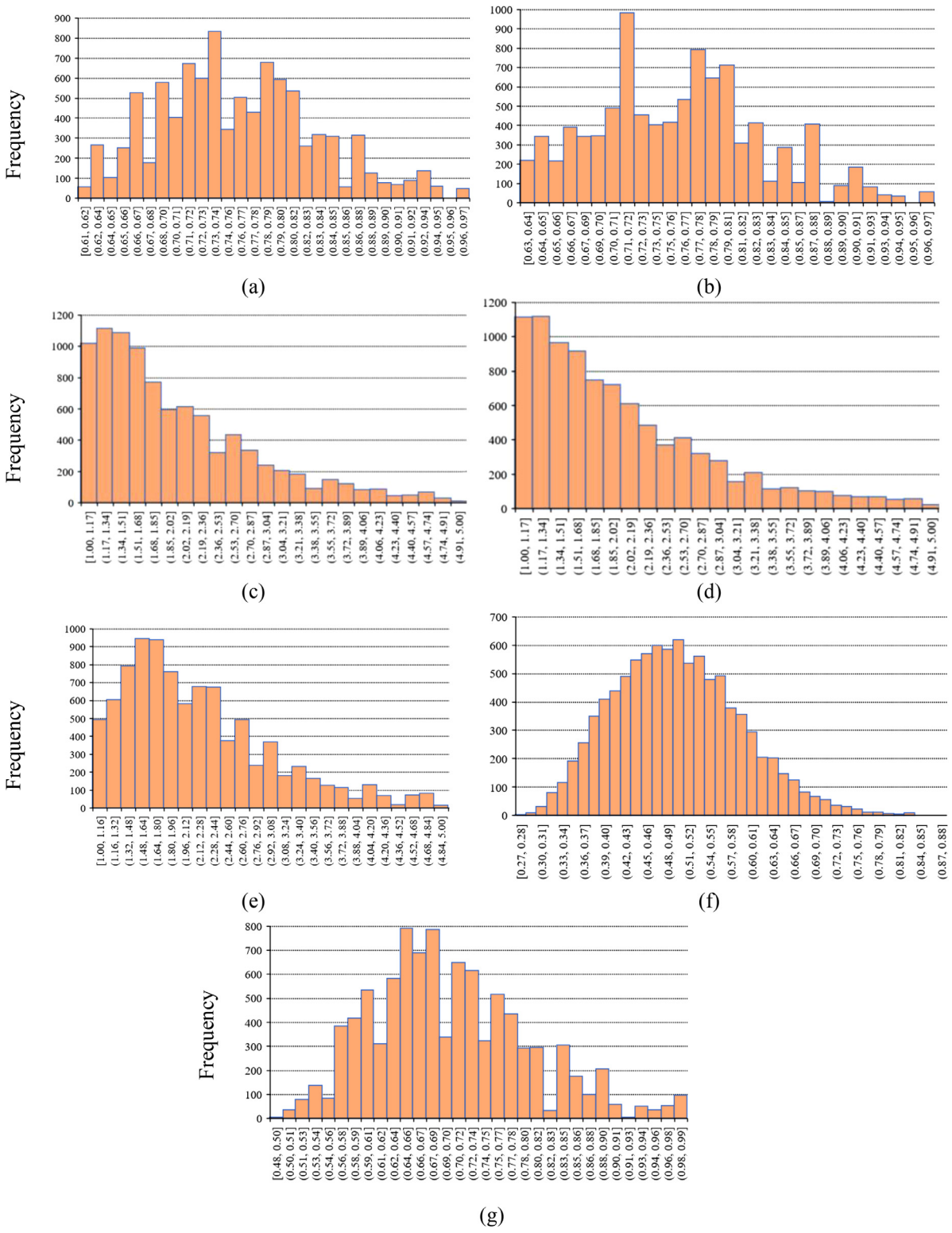


Fig. 11. Distribution of Z_{33} for the database of 2-storey reference building, considering various ratios: a) EC1_1/ IC1_1; b) EC1_2/ IC1_2; c) IC1_1/ IC1_2; d) EC1_1/ EC1_2; e) EB1_1/EB1_2; f) EB1_1/(EC1_1+ EC1_2); g) EB1_2/ EC1_2.

Regarding the comparison of I_{33} ratios for beams, Fig. 10(e) clearly shows that the dimensions of the beams in the first storey are equal to or greater than those in the second storey. For the majority of sample buildings, the beams at the bottom storey are not significantly larger than those at the top storey. Finally, the ratio of I_{33} at the intersection of beams and columns at the second storey is analysed and depicted in Figs. 10 (f-g), where Fig. 10(f) displays this value at the bottom intersection and Fig. 10(g) shows the ratio at the top intersection, confirming that the condition of SC-WB from the resilience criteria is satisfied for the second storey.

Fig. 11 presents similar results for Z_{33} in 2-storey buildings. The results show that histograms for 20-storey buildings are less distributed

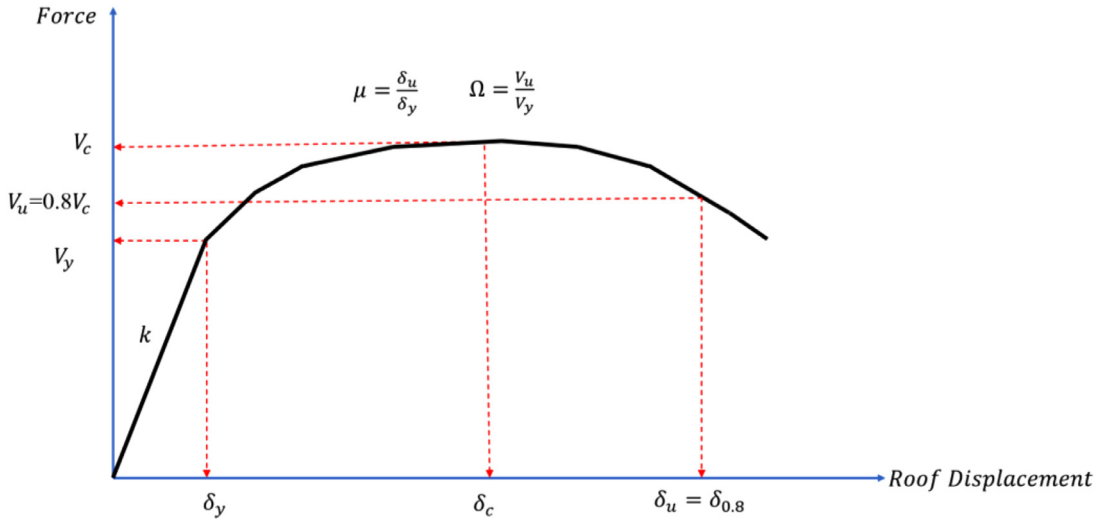


Fig. 12. Typical Force-Roof displacement curve and pushover parameters.

across different intervals compared to the 2-storey and 8-storey buildings. This can be attributed to the increased number of conditions for resilience and practicality criteria at each intersection of beams and columns. As a result, the candidates for beams and columns at the first storey level that satisfy all the criteria become limited, leading to restricted column and beam sizes for each storey level and, consequently, fewer ranges for I_{33} and Z_{33} in 20-storey buildings.

6. Pushover results

The pushover curve of a multi-degree-of-freedom (MDOF) system establishes a relationship between the base shear of a structure and its roof displacement. Previous studies [24,55] have highlighted the wide utilization of the pushover method and experiments in developing capacity models. As per Fig. 12, the pushover parameters including V_c (maximum strength); V_y (yield strength); Ω (overstrength factor); δ_y (displacement corresponding to yield strength); δ_u (displacement corresponding to ultimate strength); μ (ductility capacity); T_1 (fundamental period of structure), and MP_1 (modal participation ratio at the fist mode of vibration) are extracted from Force-Roof Displacement curve. For each 2-storey, 8-storey, and 20-storey SMRFs database, pushover analysis is performed on nearly 10,000 building models, employing a lateral load pattern proportional to the fundamental mode of vibration. The pushover curves are displayed in Fig. 13, with the vertical axis normalised by the total weight of the building and the horizontal axis representing roof drift values. The grey curves represent the pushover curves of randomly generated frames, while the red solid and dashed lines represent the median and the 16th and 84th percentiles of the results, respectively.

This figure indicates a decrease in both ductility capacity and maximum normalised strength as the number of stories increases. The decrease in ductility capacity is attributed to the significant P-Δ effect observed in taller buildings. As anticipated, taller buildings necessitate lower design base shears to ensure structural stability.

During the verification process of the design of 30,000 buildings, the criteria of resiliency and practicality are taken into consideration. Besides, to ensure the validity of the design, the acceptable design range of random buildings is evaluated as depicted in Figs. 10 and 11.

Fig. 14 displays a box plot that represents the first and third quartiles, as well as the median values of T_1 for 10,000 buildings in each database. The mean value is indicated by a cross mark within the box plot. The vertical lines, known as whiskers, extend from the ends of the box to illustrate the minimum and maximum values. According to Fig. 14, the mean values of T_1 for all reference buildings (1.10, 2.74, and 4.56 s for 2-, 8-, and 20-storey buildings, respectively) are very close to those of the reference building specified by FEMA-P695 (0.91, 2.29, and 4.47 s for 2-, 8-, and 20-storey buildings, respectively).

7. Validation of the number of building samples

Sensitivity analysis is done on the number of required samples for the meta databases to make sure that the generated number of samples is sufficient for developing surrogate models. In this regard, four random parameters including fundamental period of structure (T_1), mass of building (M_{Story}), overstrength (Ω), and normalised base shear ($\frac{V_y}{W}$) are considered and their Mean and standard deviation ($Std.dev$) are extracted for different numbers of building samples. As per Fig. 15, there is no considerable fluctuation in the mean and standard deviation after 10,000 buildings, leading us to select 10, 000 as the sufficient number of samples for generating random buildings.

8. Feature importance study

Previous studies have employed machine learning techniques to explore the interrelationships between different parameters. For instance, Kim et al. [56] considered various design parameters as random variables and highlighted the significance of beam yield strength for moment-resisting frame buildings, while column yield strength was found to be crucial for dual system buildings. Zhang et al. [57] examined the influence of uncertain parameters on the fire resistance of reinforced concrete slabs, illustrating their sensitivity using linear regression, random forest, gradient-boosting decision trees, and extreme gradient boosting. Nguyen and Dang [58] conducted sensitivity and reliability analysis on random variables across different stages of a horizontal steel frame in a one-storey industrial building with a crane in Vietnam. Javadian and Kim [59] developed a fuzzy sensitivity model to investigate the impact of input parameter uncertainty on output uncertainty, employing a truss structure and a four-storey reinforced concrete framed structure. Rodríguez et al. [60] performed sensitivity analysis on steel moment-resisting frames with semi-rigid bolted connections, assessing the sensitivity of progressive collapse to design criteria and column-loss scenarios. Tipu et al. [61] utilised random forest to determine feature importance and identify the influence of different concrete components on compressive strength. SHAP (SHapely Additive exPlanations) was employed by Somala et al. [62] to examine the impact of individual input variables on predicting peak ground motion parameters in New Zealand, with the magnitude being the most important feature for peak ground velocity estimation and rupture distance for peak ground

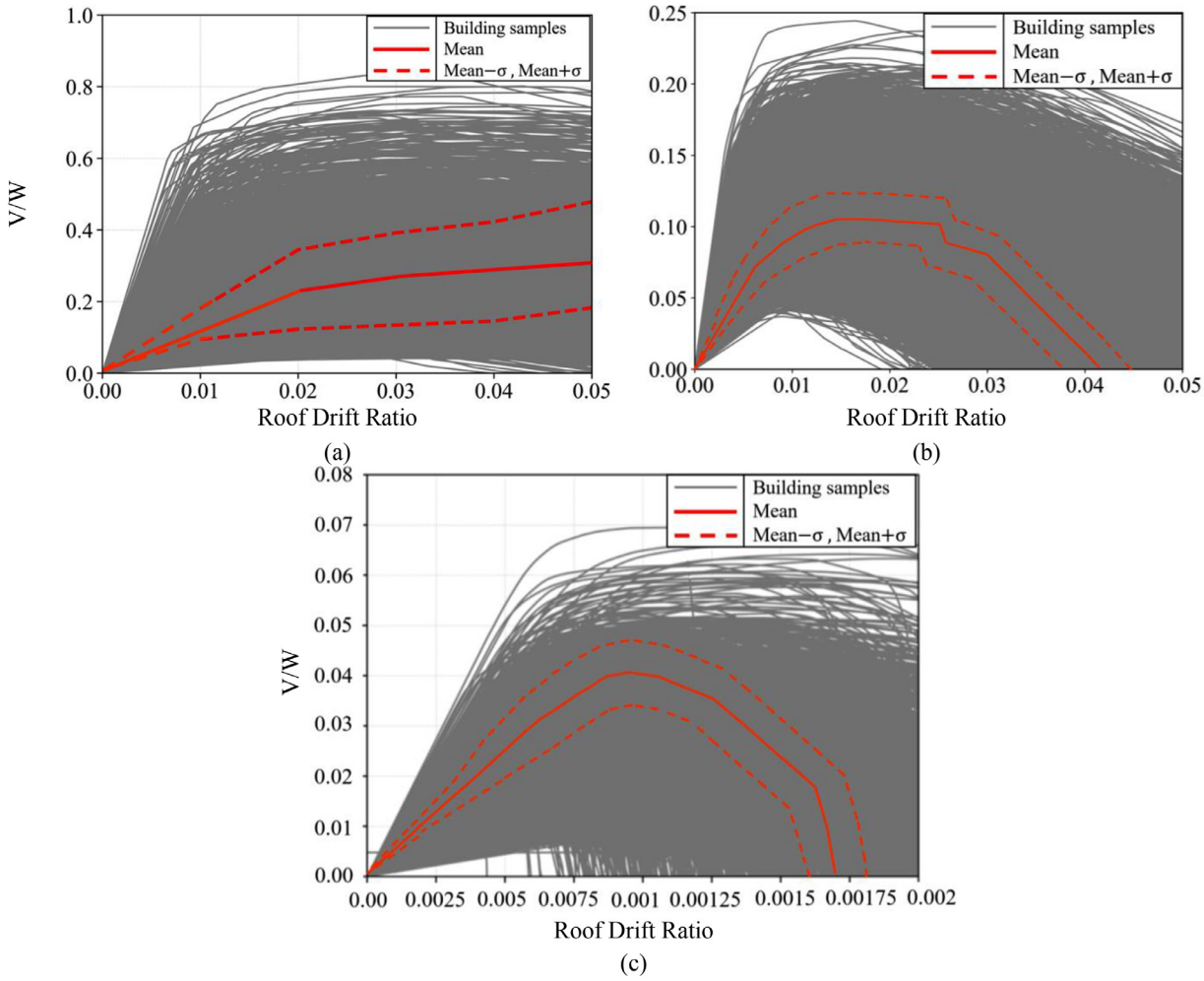


Fig. 13. Pushover curves for: a) 2-storey buildings database; b) 8-storey buildings database; c) 20-storey buildings database.

acceleration. Le et al. [63] analysed the relationship between input parameters affecting the shear strength of FRP-RC beams and the output (shear strength) using feature importance analysis and SHAP values. Similarly, Truong et al. [64] identified the most critical factor, effective footing depth, for predicting punching shear strength of reinforced concrete (RC) column footings through feature importance analysis.

These studies demonstrate the application of machine learning techniques to assess the influence of input parameters on output parameters across various structural systems and phenomena. This research also explores the impact of various input parameters on output parameters. However, there is a challenge in this regard as there are numerous methods available in the literature for conducting feature importance analysis. Rajbahadur et al. [65] have classified feature importance analyses into two main categories: (1) classifier-specific (CS) methods such as Random Forest, Neural Networks, Regression Trees, Logistic Regression, etc., and (2) classifier-agnostic (CA) methods like permutation importance (Permutation) and SHapley Additive ExPlanations (SHAP). The CS methods typically utilise a given classifier's internals to calculate the feature importance scores, while CA methods determine the importance of a feature by treating the classifier as a "black-box," i.e., without using any classifier-specific details. They also noted that different feature importance methods may result in different feature importance ranks even for the same dataset and classifier.

On one hand, the authors chose RF due to its lower computational costs compared to other methods. Additionally, Saarela and Jauhiainen [66] discussed the advantages of RF and concluded that RF outperformed other linear methods like logistic regression. On the other hand, SHAP is one of the more recent global feature importance methods that is theoretically guaranteed to produce optimal feature importance ranks [67] compared to permutation, which is one of the oldest CA methods. Although SHAP was proposed by Lundberg and Lee [68] only in 2017, it has already garnered over 2000 citations. Furthermore, Hooker et al. [69] demonstrated that while permutation importance is a very attractive choice for model interpretation, it has several problems, especially when working with correlated features.

As a result of this research, RF and SHAP have been chosen as the representatives of CS and CA techniques, respectively. The aim is to analyse how distinct feature importance methods lead to varying crucial parameters for each SMRF building database.

8.1. Random forest (RF)

The random forest regressor (RFR) is an ensemble learning algorithm that builds upon the classification and regression tree (CART) method. It was first introduced by Breiman [70] and has been widely used in the field of machine learning. By combining multiple CARTs, random forests enhance the accuracy of regression and prediction tasks. In the RFR model, each decision tree operates independently, while all the CARTs work together to generate the final output. The working principle of RFR involves creating multiple CART estimators, and the final output is obtained by averaging the outcomes of these estimators. This approach ensures improved prediction performance for each individual CART. In a random forest, each tree is constructed using a random subset of ob-

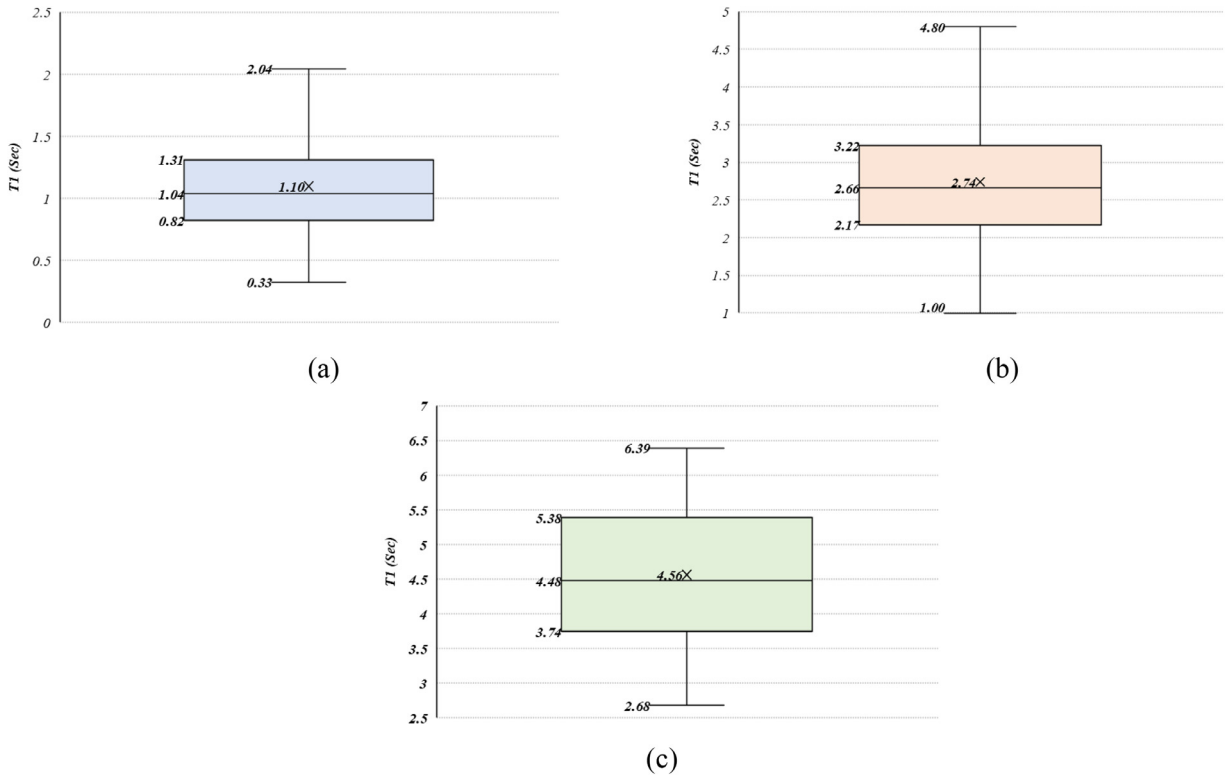


Fig. 14. The box and whisker plots of T_1 (Sec) for: a) 2-storey database; b) 8-storey database; c) 20-storey database.

servations, and splits within each tree are based on random subsets of candidate variables [71,72].

The process of RFR can be summarised into three main steps: first, K regression trees are constructed using bootstrapped subsets, which are generated by randomly sampling from the database; second, at each node of each tree, K segmentation variables are randomly selected, and the best branching criterion is used to determine the splits; finally, each regression tree starts from the top and proceeds with top-to-bottom branching until the termination criterion for segmentation is met. Mathematically, the RFR can be represented by Eq. (1):

$$\hat{y} = \frac{1}{B}\sum_{d=1}^{B} f_d(x) \tag{1}$$

where $\hat{y}$ represents the output of the random forest regressor, and $f_d(x)$ represents the output of an individual decision tree and B is the number of trees.

In the RF model, feature importance is determined by evaluating the impurity at each node. The importance of a feature is measured by assessing its contribution to reducing impurity across all the trees in the forest. The mean squared error (MSE) is used as the loss function in the RF model to quantify the error. To calculate the importance of each node individually, Eq. (2) is employed, where each feature's importance is estimated from the forest's weighted average of all the trees. Subsequently, the importance of each feature is determined by applying Eq. (3).

$$N_{i,j} = \frac{ni}{N}\cdot\left[MSE - \left(\frac{n_{i,left}}{N}\times left\ MSE\right) - \left(\frac{n_{i,right}}{N}\times right\ MSE\right)\right] \tag{2}$$

$$F_i = \frac{\sum N_j}{\sum N_{ij}} \tag{3}$$

In Eq. (2), $N_{i,j}$ is the i^{th} node importance for j^{th} feature, $n_{i,left}$ and $n_{i,right}$ are the samples at the left and right node, respectively, N refers to the total number of samples. Also in Eq. (3), F_i represents the importance of the i^{th} feature, N_j represents the importance of the j^{th} node, and N_{ij} denotes the importance of all nodes that have the i^{th} feature. This equation quantifies the contribution of each feature to the overall importance within the random forest

In this study, the RF algorithm is applied in three different scenarios. Firstly, it is utilised to determine the key cross-sectional properties that significantly impact the spring parameters of beams and columns, such as θ_p, θ_{pc}, and as_{ratio}. Secondly, RF is employed to examine the influence of only random parameters on the output parameters. In this case, the focus is solely on understanding the impact of random factors. Thirdly, the RF algorithm is extended to include both random and non-random parameters, allowing for the identification of the most important factors overall. By considering all input and output parameters, the RF model can provide insights into the crucial role they play in the structural performance of buildings.

According to Fig. 16, when analysing spring parameters in beams of 2-storey buildings, Z_{33} emerges as the most influential parameter among the various spring parameters. However, according to the results, for 8-storey and 20-storey buildings, parameter A takes on the role of being the most critical factor. Furthermore, *A. remains*of utmost importance for columns across all building heights, including 2-storey, 8-storey, and 20-storey structures.

In the feature importance analysis involving all random and casual random parameters, different trends emerge for the target parameters of 2-storey buildings. As shown in Fig. 17, for parameters such as δ_y, δ_c, V_c, Ω, $\delta_{-0.8}$ and μ, the mass of the building takes the top spot in terms of importance, while material properties rank second. It should be noted that due to the large number of parameters involved, only the top twenty most important parameters are plotted in the figure. However, for the target parameter MP_1, the moment of inertia of beams at the second storey ($IBeam_2$) plays the most significant role, while the cross-sectional properties of columns at the first storey have a relatively lower importance. For V_y, the parameter E_s takes the lead as the most important factor, followed by the mass of the buildings. Lastly, for T_1,

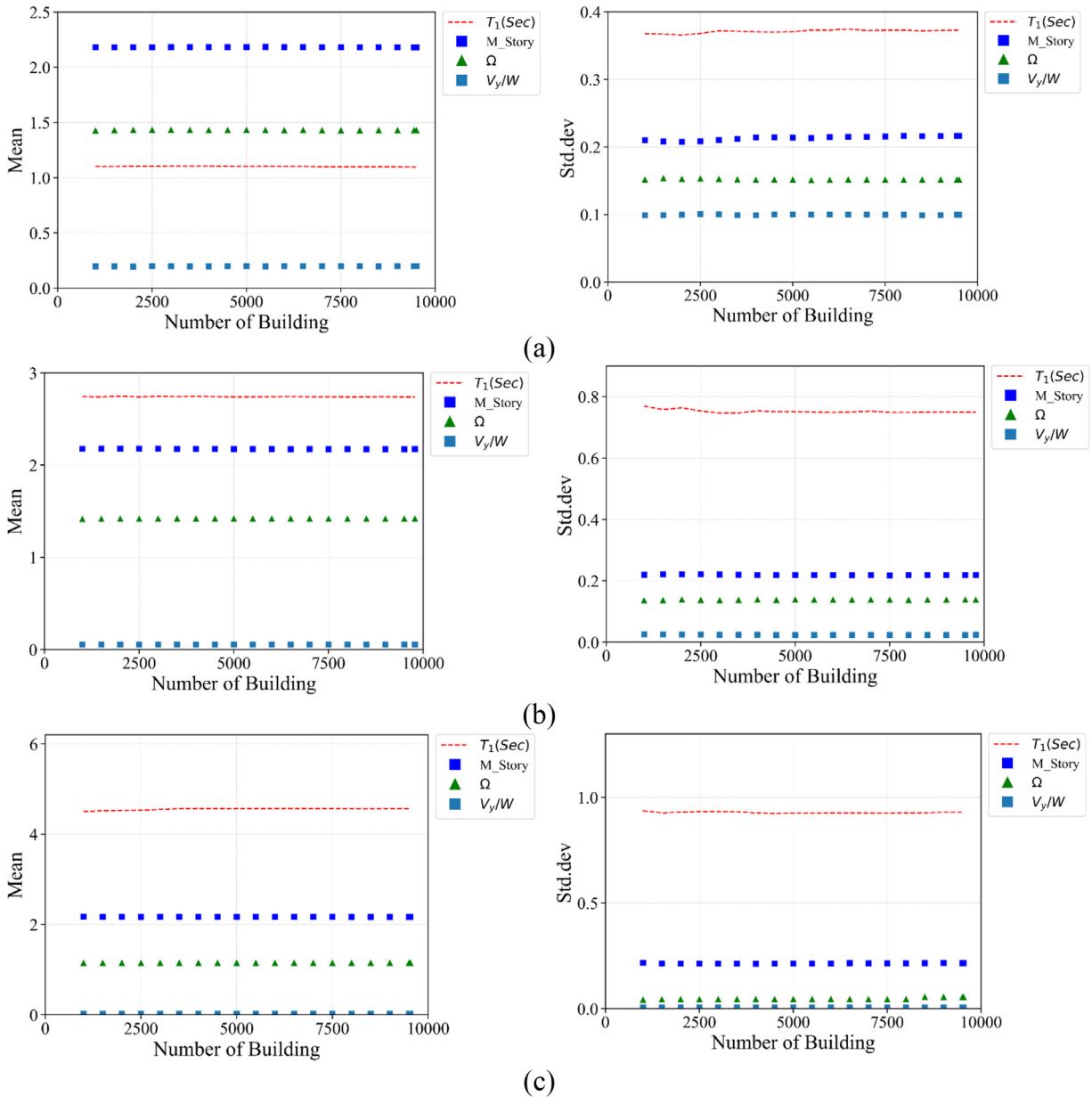


Fig. 15. Sensitivity analysis on the number of building samples for: (a) 2-storey building database; (b) 8-storey building database; (c) 20-storey building database.

$\theta_{pc}\_Col\_In1$ is identified as the most critical parameter, with the height of the storey being the second most important. Additionally, according to the results in Fig. 17(f), it is observed that for the parameter θ_{pc}_Col, the cross-sectional area holds the highest importance, leading A_Col_In1 to influence the value of T_1 significantly.

In the case of 8-storey buildings, as per the results, the importance of parameters varies for different target parameters. For δ_c and T_1, the height of the first storey plays the most crucial role. On the other hand, for V_c, V_y, $\delta_{\_0.8}$, the cross-sectional properties of interior columns at the third storey hold significant importance. For the parameters δ_y, Ω and μ, the cross-sectional properties of beams at the third storey are of great importance. In the case of MP_1, the moment of inertia of beams at the first storey ($IBeam_1$) is the most important parameter, indicating its strong influence on the target parameter.

In the case of 20-storey buildings, when assessing RF feature importance analysis, it is observed that for most parameters, except T_1 and MP_1, the mass of the buildings has the greatest impact on the output parameters. The mass of the buildings is a significant factor influencing the structural performance. For the parameter T_1, the height of the first storey plays the most important role. This suggests that the height of the initial storey significantly affects the fundamental period of the building. Regarding MP_1, $\theta_p\_Col\_In1$ emerges as the most important parameter. On the other hand, we know that the cross-sectional area of an interior column is the most important factor for θ_p_Col, implying that the cross-sectional properties, specifically the area of the interior column (A_{col_In1}), have a strong influence on the modal participation ratio of the 20-storey buildings.

During RF analysis considering only random parameters, the following findings are observed for 2-storey buildings as per Fig. 18. (1) For parameters such as δ_y, $\delta_{\_0.8}$, Ω, and μ, the mass of the buildings is identified as the most important factor. It has the highest impact on these output parameters and material properties rank second in terms of importance. (2) For δ_c, V_c, and V_y, the parameter E_s is identified as the most important. (3) The height of the first storey has the greatest influence on the parameter T_1 and in the case of MP_1, the length of the bay emerges as the most important parameter.

For 8-storey buildings, when examining parameters such as δ_c, δ_y, μ,T_1, V_c, V_y, it became evident that the height of each storey played the

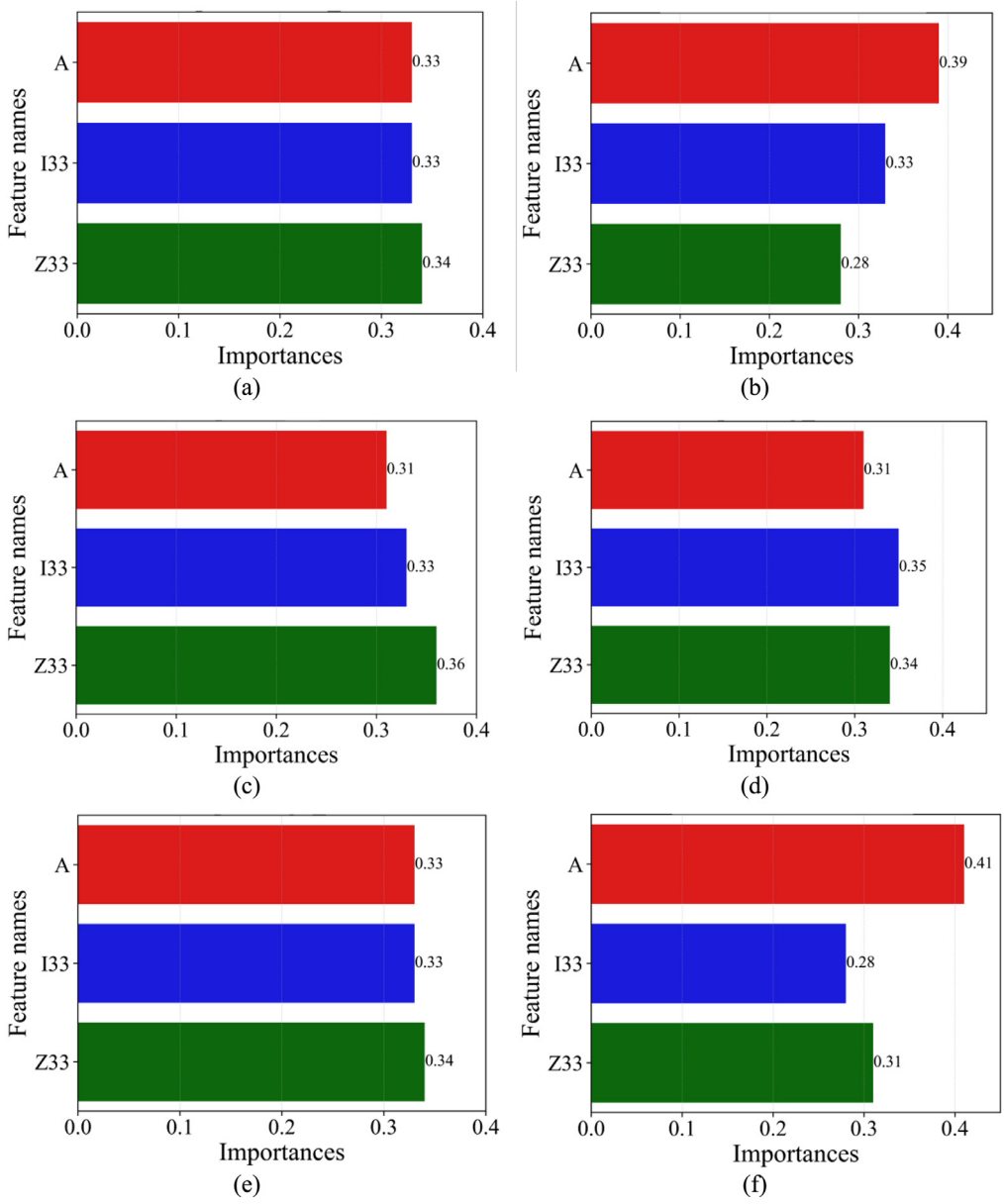


Fig.16. Random Forest feature importance for spring parameters of 2-storey buildings for target parameter:(a) as_{ratio}_Beam; (b) as_{ratio}_Column; (c) θ_p_Beam; (d) θ_p_Column; (e) θ_{pc}_Beam; (f) θ_{pc}_Column.

most pivotal role. Interestingly, in the case of Ω, it was the elastic modulus of the material, E_s, that emerged as the most significant determinant. Furthermore, when investigating the parameter $\delta\_0.8$, it was discovered that the yield strength itself (F_y) held the utmost importance. Lastly, focusing on the modal participation ratio (MP_1), it was revealed that the mass of the buildings emerged as the most critical parameter.

The results of this study show that for 20-storey buildings, the analysis reveals that for the parameter $\delta_{\_0.8}$, the most significant factor is F_y. Conversely, for parameters such as δ_c, δ_y, V_c, V_y, the mass of buildings emerges as the most influential parameter. When it comes to Ω and T_1, the height of the storey proves to be the critical factor impacting the results. Additionally, for the parameter MP_1, it is the length of the bay that takes precedence, while for μ, the parameter E_s demonstrates the greatest importance.

8.2. SHAP (SHapely additive exPlanations)

With the advancement of explainable techniques like SHAP, it is now possible to explain the prediction of machine learning algorithms and investigate the impact of each feature on the model's prediction. To achieve an effective model, two levels of explainability are necessary: global and local [63]. Global explainability refers to understanding the effect of each variable on the prediction model, while local explainability helps us comprehend why a particular prediction is made. SHAP provides insights into the impact of each input variable individually, thus contributing to global interpretability [73,74]. Additionally, SHAP values can measure the importance of features for individual samples at a local level, enabling local interpretability. SHAP can also explain the overall importance of each feature (input parameter) on the complete dataset. It accomplishes this by employing a linear model of coalitions, as shown in Eq. (4), to make the model interpretable. SHAP value serves as an effective tool for investigating the output of a machine learning model, with the ultimate goal of enhancing its understandability. By assigning weights to each feature, SHAP can be used to explain the prediction outcomes of any machine learning model. It represents the output model as a linear sum of the input variables, allowing each observation to obtain its corresponding SHAP value. SHAP can also determine whether an input variable has a positive or negative effect on the out-

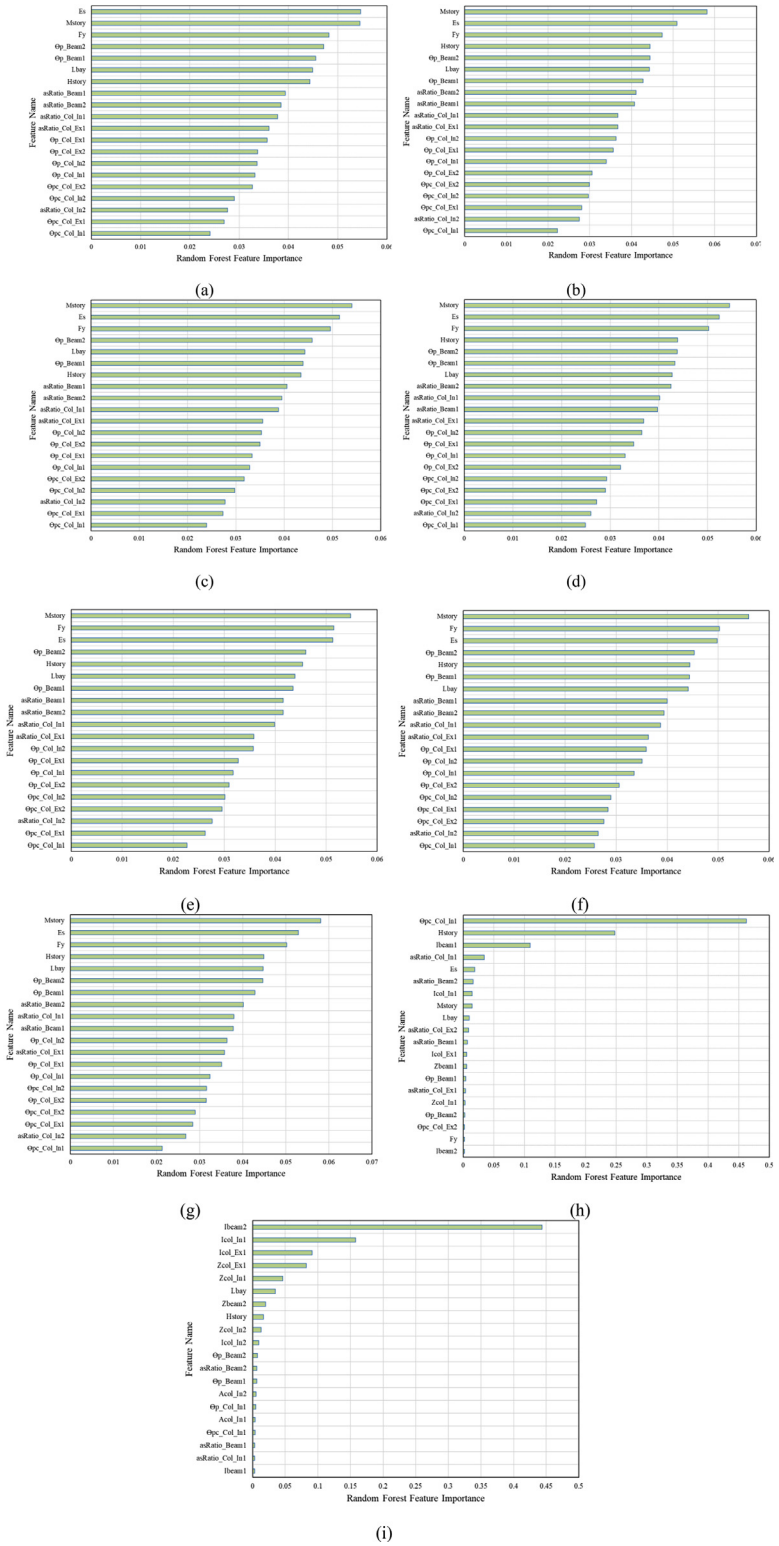


Fig. 17. Random Forest feature importance considering all random and casual random parameters of 2-storey buildings considering target parameter of: (a) V_y; (b) δ_y; (c) V_c; (d) δ_c;(e) Ω; (f) $\delta_{-0.8}$; (g) μ; (h) T_1; (i) MP_1.

put. While there are alternative methods for estimating SHAP values, this research utilises the Tree SHAP method, which is optimised for tree-based machine learning models such as decision trees, random forests, and gradient-boosted trees. Tree SHAP evaluates tree-based models in conjunction with an NxM dimensional input dataset (X), generating an $N \times M$-dimensional matrix containing the SHAP values (N represents the number of samples). The use of SHAP interaction values ensures consistent explanations for interaction effects on individual predictions. One of the benefits of SHAP is the ability to compute global interpretation by calculating the Shapley values for the entire dataset and combining them. For a more detailed explanation of SHAP and its associated proof, interested readers are encouraged to refer to [71,75].

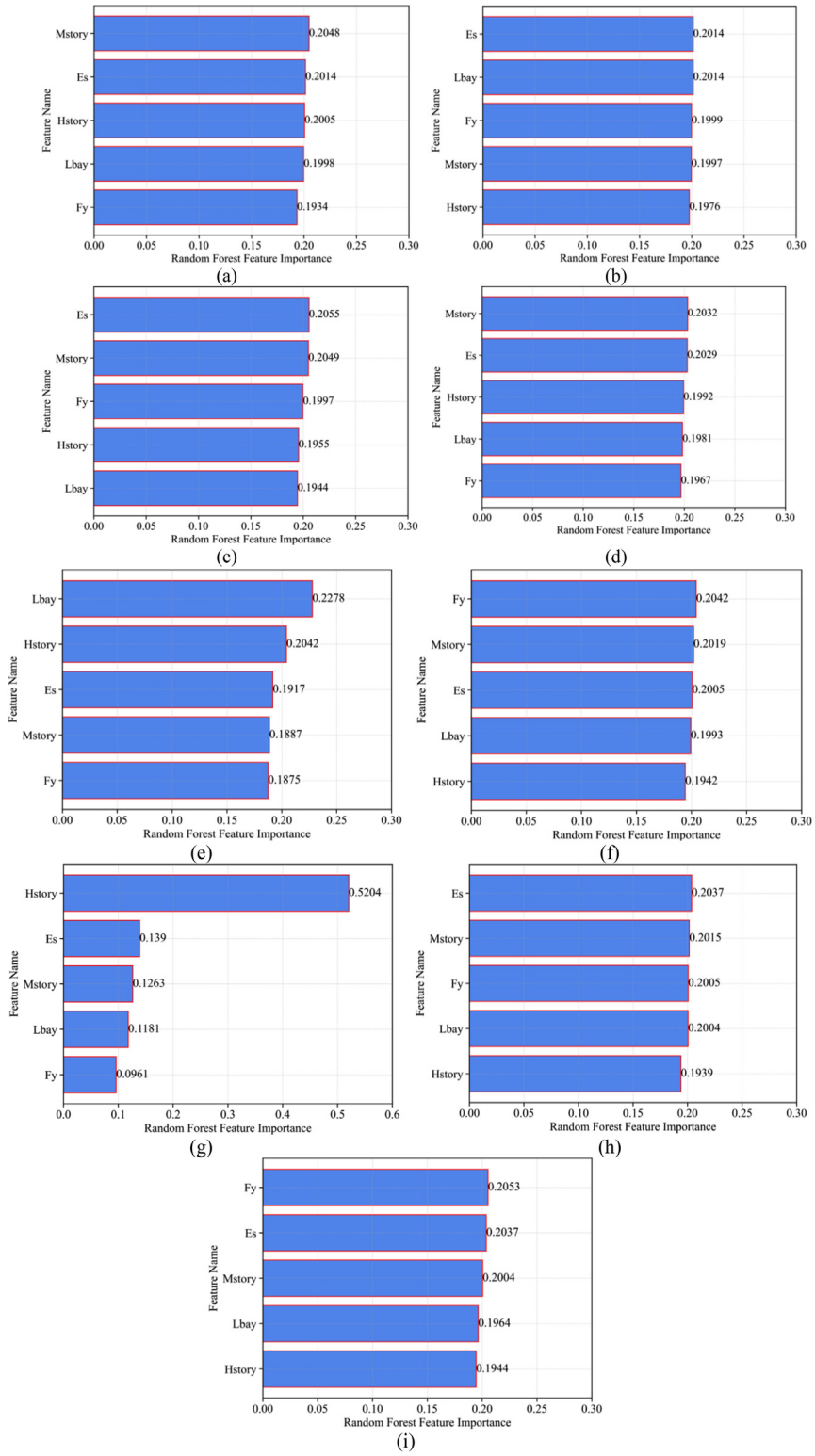


Fig. 18. Random Forest feature importance considering only random parameters of 2-storey buildings for target parameter of: (a) $\delta_{-0.8}$; (b) δ_c; (c) δ_y; (d) μ; (e) MP_1; (f) Ω; (g) T_1; (h) V_c; (i) V_y..

In Fig. 19, the overall SHAP values are presented. In this figure, the red colour indicates a positive impact, while the blue colour denotes a negative influence. A positive impact signifies an increase in prediction with an increase in the corresponding input variable. From Figs. 19 (b,d,f), it can be observed that for the spring properties of columns, the cross-sectional area has the most positive impact. On the other hand, for beams, the second-order moment of inertia has the most positive influence on θ_p_$Beam$, as per Fig. 19(c). For other spring properties of beams, the cross-sectional area is identified as the most important parameter, as per Figs. 19 (a,e). This observation holds true for the 8- and 20-storey buildings as well.

On the other hand, when considering random parameters for SHAP analysis, the results are as follows: For 2-storey buildings and $\delta_{-0.8}$, Ω, MP_1, V_c, the key parameter is the length of the bay, as per Fig. 20.

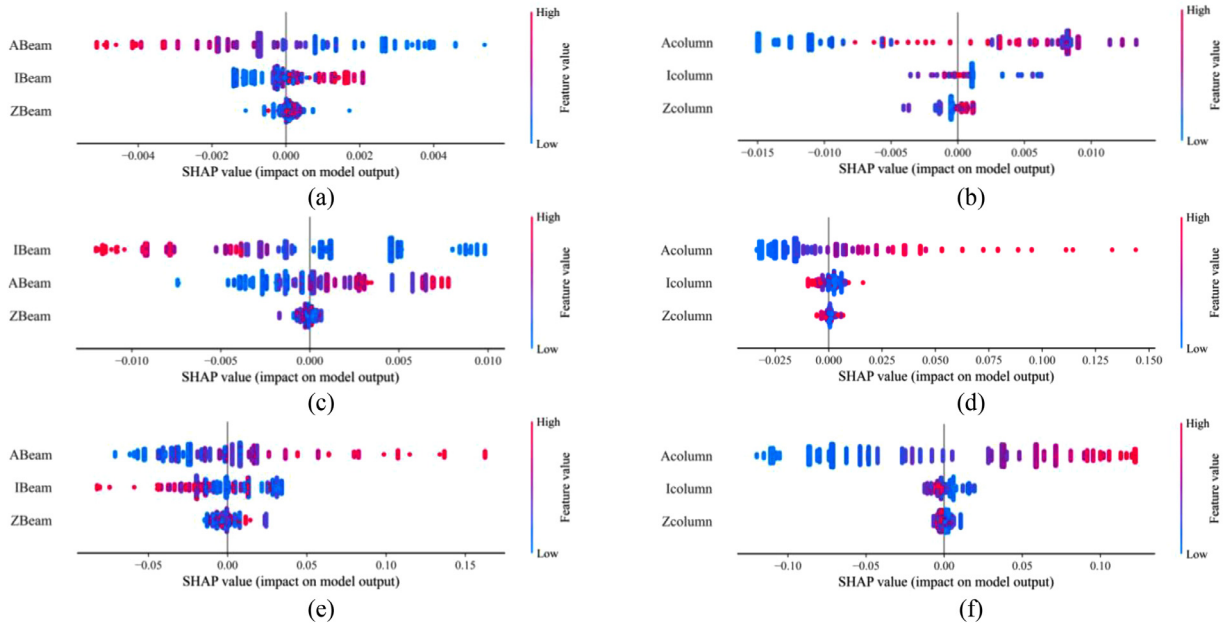


Fig. 19. SHAP values in predicting the plastic hinges parameters for 2-storey buildings for target parameter of: (a) $asRatio\_Beam$; (b) $asRatio\_Column$; (c) θ_p_Beam; (d) θ_p_Column; (e) θ_{pc}_Beam; (f) $\theta_{pcColumn}$.

For δ_y and V_y, the most important parameter is the mass of the storey. However, for μ and T_1, the height of the storey is crucial. Lastly, for δ_c, the most important parameter is E_s.

For 8-storey buildings except for $\delta_{\_0.8}$ and Ω, where F_y is the most important parameter, the height of the storey plays a key role in all other random parameters. When it comes to 20-storey buildings, for T_1, $\delta_{\_0.8}$, and δ_c, the height of the storey has the most significant impact. For μ, Ω, δ_y, the most important factor is E_s, while for V_y and V_c, F_y is the key parameter. Finally, for MP_1, the length of the bay plays the most important role.

Finally, when all random and casual random parameters are included in the SHAP analysis for the 2-storey building database, the results indicate the following:

- For T_1 and V_y, $\theta_{pc}\_Col\_In1$ is the most important parameter, as per Fig. 21. Fig. 19(f) illustrates that the cross-sectional area of the column (A_{Column}) plays the greatest role for $\theta_{pc}\_Col\_In1$, making it the most crucial factor for T_1 and V_y.
- For $\delta_{\_0.8}$ and δ_c, $asRatio\_Col\_Ex1$ and $asRatio\_Col\_In1$ are the most important parameters, respectively. Again, referring to Fig. 19(b), it can be observed that A_{Column} has the greatest influence, making it the most critical factor for $\delta_{\_0.8}$ and δ_c.
- For μ and V_c, the mass of the building is the most important parameter.
- For δ_y, the most important parameter is θ_p_Beam2 or I_{Beam2}, as shown in Fig. 19(c).
- Similarly, for MP_1, I_{Beam2} has the greatest impact, as depicted in Fig. 21(e).
- For Ω, $asRatio\_beam2$ is the most important parameter, and Fig. 19(a) confirms that A_{Beam2} is the key factor for Ω.

For 8-storey buildings, the crucial factors vary depending on the specific parameter:

- For δ_c and T_1, the height of the storey is the most important factor.
- For μ and δ_y, $asRatio\_Beam3$ plays the most significant role. A_{Beam3} can be concluded as the most important parameter for μ and δ_y
- $\theta_{pc}\_Col\_In3$ is the most important parameter for V_y and V_c. A_{Column} is the crucial factor for $\theta_{pc}\_Col\_In3$.
- For Ω, MP_1 and $\delta_{\_0.8}$, the most important parameters are A_{Beam3}, I_{beam1}, and Z_{beam5}, respectively.

For 20-storey buildings, for μ and V_c, E_s is the most important parameter. For δ_c, and δ_y, $asRatio\_Beam17$ with its crucial factor being A_{Beam}, is the most important parameter. For $\delta_{\_0.8}$, $asRatio\_Col\_In1$ (or A_{Column}) is the most crucial factor. For MP_1, A_{Col}_In1 indirectly plays the most important role, while for Ω, I_{Beam13} is the most important factor indirectly. Similarly to 2- and 8-storey buildings, for 20-storey buildings, the height of the storey remains the most important factor for T_1. Finally, for V_y, A_{Column} at the 19th storey is indirectly the most important parameter.

Discussions

To summarize the results presented in this work, and to understand straightforwardly which random or casual random parameters are of utmost importance for each database, for Fig. 17 and for RF results of 8-storey and 20-storey buildings, a cut-off value of 0.05 is defined for feature importance values and those feature names whose their feature importance value is greater or equal than 0.05 are identified for each structural parameter of $\delta_{\_0.8}$, δ_c, δ_y, μ, MP_1, Ω, T_1, V_c, V_y, and then, a weighted average is applied to find out the top feature name for each database. As reported in Fig. 22(a) and for 2the -storey database, among all random and casual random parameters, $\theta_{pc}\_Col\_In1$, plays the most important role when considering all structural parameters of $\delta_{\_0.8}$, δ_c, δ_y, μ, MP_1, Ω, T_1, V_c, V_y combined. On the other hand, from the results of RF for spring parameters as shown in Fig. 16(f), we know that for $\theta_{pc}\_Col\_In1$, cross-sectional area of column influence the most, leading parameter A_Col_In1 to be the most important parameter for the database of 2-storey building. For the 8-storey database, however, the moment inertia of interior columns at the third storey can be regarded as the most crucial feature name, as reported in Fig. 22(b). The results in Fig. 22(c) show that the height of the first storey comes at the top feature for the 20-storey database.

On the other hand, Table 8 compares the results extracted by RF and SHAP analyses for four main structural parameters of μ, Ω, MP_1, and T_1. It is interesting that for three databases of 2-storey, 8-storey, and 20-storey buildings, both RF and SHAP lead to the almost same results. For the 2-storey building database, mass of building is the most important factor for the ductility parameter, while for the modal participation ratio, the moment inertia of beams at the second storey is of great

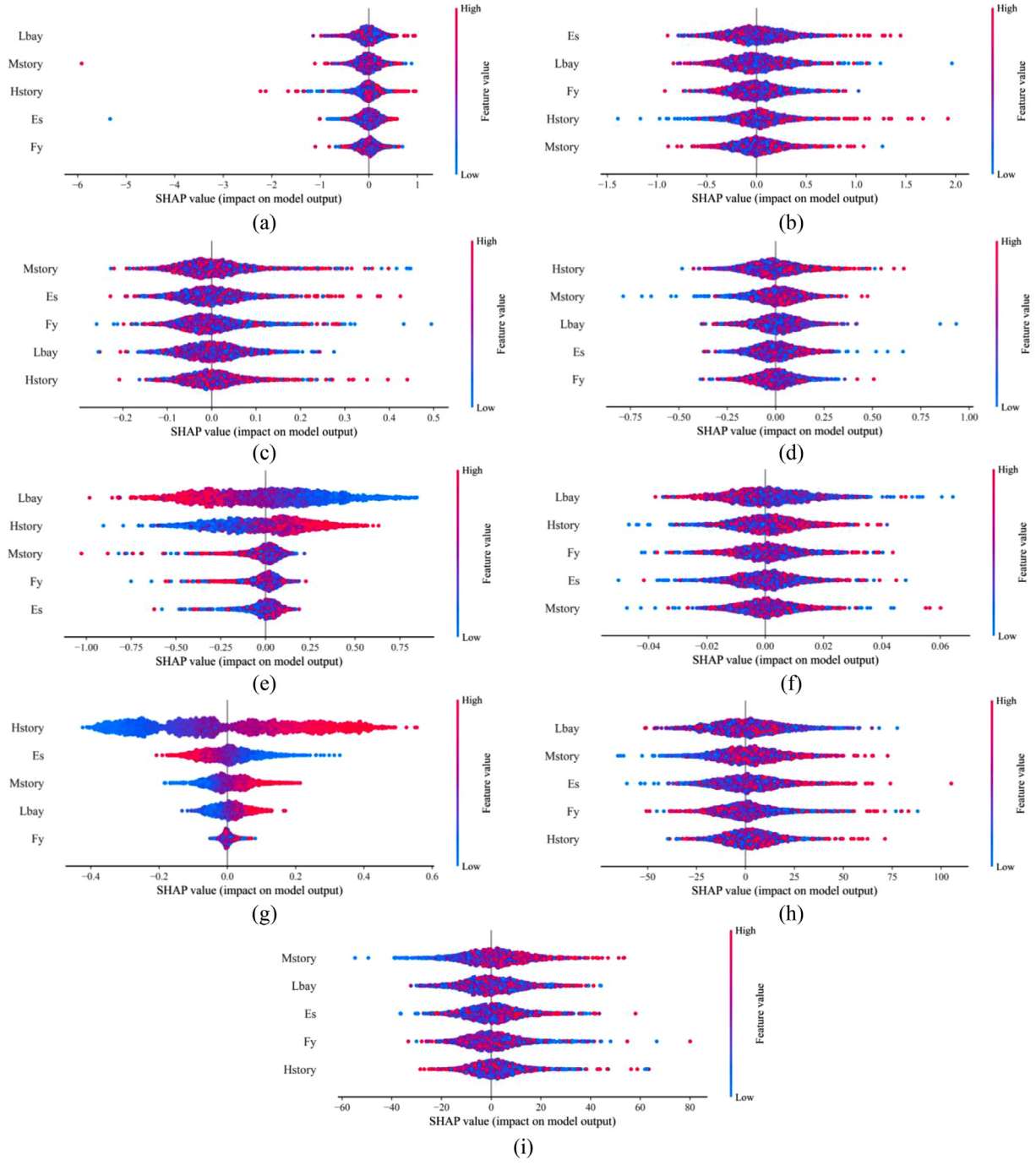


Fig. 20. SHAP values in predicting only random parameters for 2-storey buildings for target parameter of: (a) $\delta_{-0.8}$; (b) δ_c; (c) δ_y; (d) μ; (e) MP_1; (f) Ω; (g) T_1; (h) V_c; (i) V_y..

Table 8
Comparison between RF and SHAP analyses for four main structural parameters for low-rise, mid-rise, and high-rise buildings' databases.

	μ		MP_1		Ω		T_1	
	RF	SHAP	RF	SHAP	RF	SHAP	RF	SHAP
2-storey database	$Mass$	$Mass$	I_{Beam2}	I_{Beam2}	$Mass$	A_{Beam2}	A_{col_In1}	A_{col_In1}
8-storey database	A_{Beam3}	A_{Beam3}	$Beam3$	I_{beam1}	A_{Beam3}	A_{Beam3}	$History$	$History$
20-storey database	$Mass$	E_s	A_{col_In1}	A_{col_In1}	$Mass$	I_{Beam13}	$History$	$History$

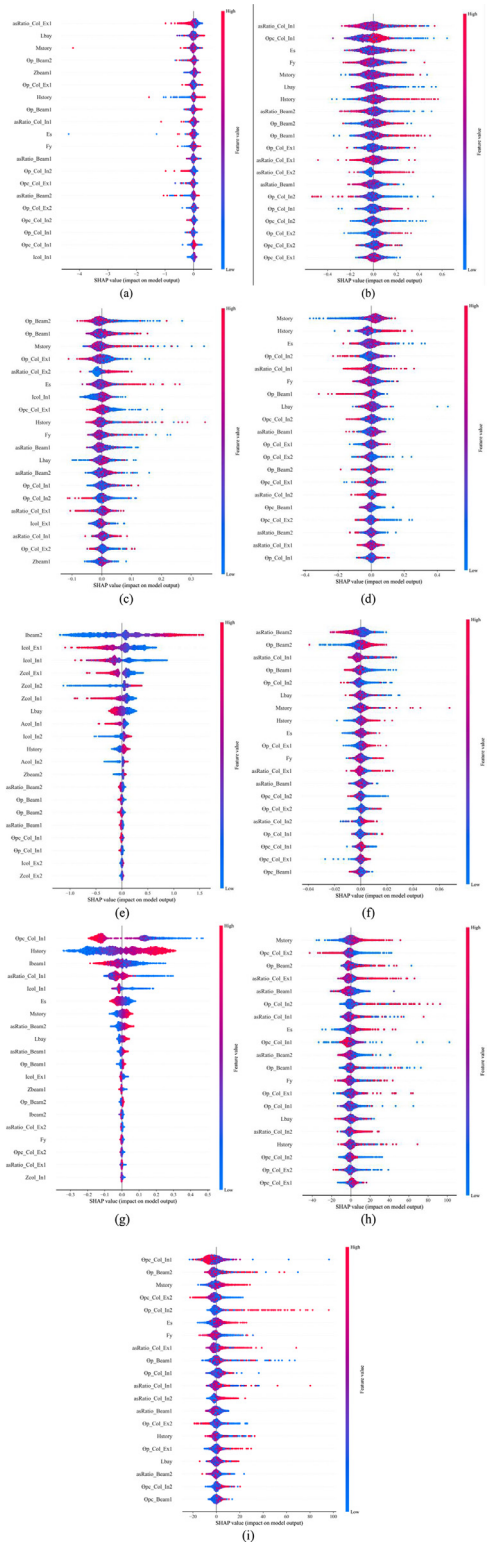


Fig. 21. SHAP values in predicting all random parameters for 2-storey buildings for target parameter of: (a) $\delta_{-0.8}$; (b) δ_c; (c) δ_y; (d) μ; (e) MP_1; (f) Ω; (g) T_1; (h) V_c; (i) V_y..

importance. The interior columns at the first storey influence the fundamental period of buildings the most. RF leads mass of building to be the most important parameter for the overstrength parameter whereas SHAP results A_{Beam_2} as the key factor for Ω. For the 8-storey database, it can be concluded that cross-sectional properties of beams at the third storey play the greatest role for μ, Ω, and MP_1. For T_1, however, height of first storey is the crucial factor. For the 20-storey building's database, similar to the 8-storey database, height of first storey plays a pivotal role in T_1. For MP_1, cross-sectional area of interior column at the first storey emerges as the most important parameter. While in RF, the mass of building is the most important factor for μ, and Ω, explainability analysis regards E_s and I_Beam13 to be the most influential parameter for μ and Ω, respectively.

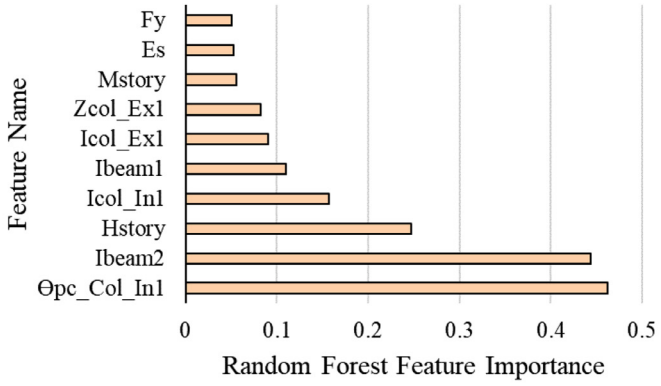
Conclusions

The framework presented in this study can address the challenge of obtaining a validated yet easily generatable database of steel frame structures by striking a balance between accuracy and computational efforts. It achieves this by considering the preliminary design of SMRF and implementing specific criteria to validate the database to the greatest extent possible. Additionally, all the significant parameters, excluding the number of bays and stories, are treated as random variables, reducing uncertainties associated with the material, geometry, and spring properties of the buildings. Once the framework is validated, the meta database for low-rise, mid-rise, and high-rise SMRF buildings are created and following this, pushover analysis is performed. Next, feature importance analysis is accomplished via RF and SHAP on various random and casual random structural parameters from the geometry of buildings to plastic hinges parameters to find out which ones are of great importance for SMRF buildings.

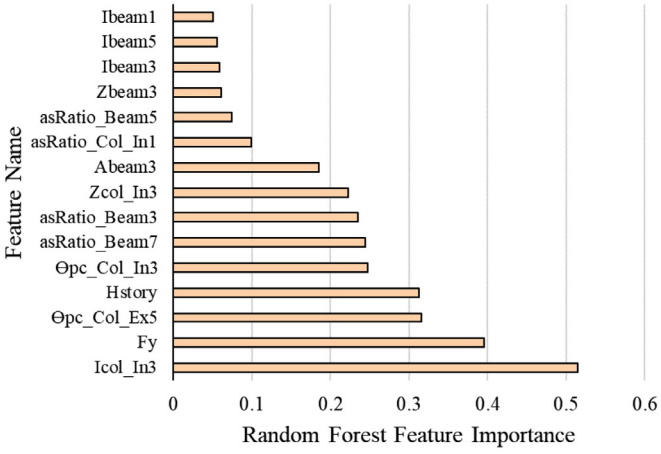
The main results from the feature importance analysis are as follows:

- When all random and casual random parameters are taken into consideration:
  - For the database of 2-storey reference building, cross-sectional area of interior columns at the first storey, A_Col_In1, is supposed to be the most important parameter.
  - For the 8-storey database, the moment inertia of interior columns at the third storey plays the most crucial random parameter.
  - For the 20-storey database, the height of first storey accommodates the top feature.
- Considering four main structural parameters of μ, Ω, MP_1, and T_1, the results show that RF and SHAP come up with almost the same results for low-rise, mid-rise and high-rise building databases.
  - For 2-storey building's database, mass of building and the moment inertia of beams at the second storey are the most important factors for μ and MP_1, respectively.
  - The interior columns at the first storey influence T_1 the most. RF and SHAP result mass of building and A_{Beam_2} as the most important parameters for Ω, respectively.
  - For both databases of 8-storey and 20-storey buildings, the height of first storey is the crucial factor for T_1.
  - For the 8-storey building database, cross-sectional properties of beams at the third storey plays the most pivotal role for μ, Ω, and MP_1.
  - For the 20-storey building database and for MP_1, cross-sectional area of interior column at the first storey emerges as the most important parameter.
  - While RF will result the mass of building to be the most important factor for μ and Ω, SHAP regards E_s and I_Beam13 to be the most influential parameter for μ and Ω, respectively.
- Deploying RF and SHAP for plastic hinge parameters shows that cross-sectional area can be regarded as almost the most important parameter for all three databases.

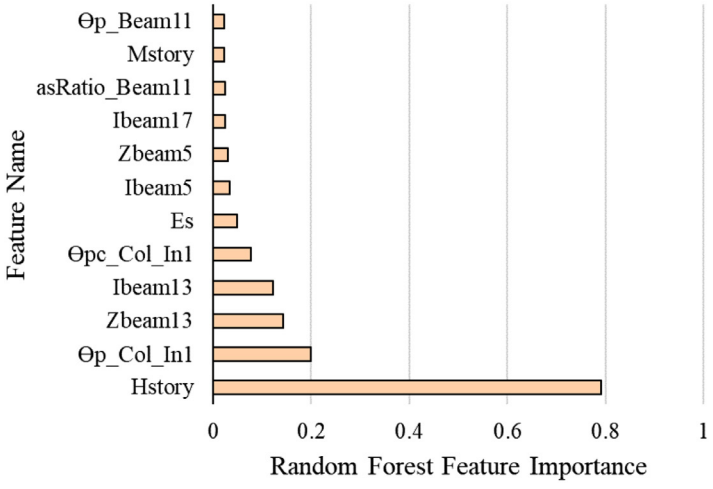
From the results of this investigation and as a practical application of this study, it can be clarified that for designing low-rise and mid-rise buildings, designers will be encouraged to devote paramount importance to the interior columns at the bottom level of buildings. Moreover,



(a)



(b)



(c)

Fig. 22. Weighted average feature importance extracted by RF for: (a) 2-storey database; (b) 8-storey database; (c) 20-storey database.

for high-rise buildings, choosing a proper range for the height of first storey can influence the overall performance of structures markedly.

It might be interesting for readers that the present study has some limitations. First, the framework for generating random buildings is only designed for steel frame structures with RBS connections. Second, the preliminary databases for the dimensions of beams and columns are extracted from AISC electronic database and one should modify the database for buildings designed for other locations. To extend beyond the findings of this investigation, future studies can be accomplished to develop meta databases for other types of buildings such as RC and masonry buildings or structures with other occupancies such as bridges, hospitals, schools, etc. The next step of this research involves utilising the meta databases generated here for developing surrogate models via different techniques considering earthquakes and floods, separately as

the dominant natural hazards. Besides, to account for multi-hazard scenarios, a separate surrogate model will be developed for meta databases of low-rise, mid-rise, and high-rise buildings combined.

Relevance to resilience

Resilience aims to aid decision-makers in planning for societies that can withstand natural disasters like earthquakes, floods, and man-made events such as terrorist attacks. Quantifying resilience for physical and organisational systems has been explored. However, structural engineering faces challenges in dealing with inevitable hazards like earthquakes and floods in terms of quantifying resilience on a city scale. To cope with the challenge of NLTHA, surrogate models are proposed that have gained increasing attention, offering statistical approaches for probabilistic seismic risk assessment (PSRA). This investigation introduces a meta database and a framework for surrogate modelling of steel structures. This approach necessitates structural geometry details like stories, bay lengths, beam, and column dimensions. Leveraging machine learning algorithms, sensitivity analyses were conducted on structural parameters to pinpoint the most influential factors in designing steel moment-resisting frames.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work resported in this paper.

Acknowledgments

The authors acknowledge the financial support from Teesside University to support the Ph.D. programme of the first author.

Data availability

The results for 8-storey and 20-storey buildings are available via this link. All the other data regarding 2-story SMRF buildings that support the findings of this study are available from the authors upon reasonable request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rcns.2023.12.001.

References

[1] Ghobarah A. Performance-based design in earthquake engineering: state of development. Eng Struct 2001. doi:10.1016/S0141-0296(01)00036-0.
[2] Porter KA. An overview of PEER's performance-based earthquake engineering methodology. 9th Int Conf Appl Stat Probab Civ Eng 2003;273:973–80.
[3] Bertero RD, Bertero VV. Performance-based seismic engineering: the need for a reliable conceptual comprehensive approach. Earthq Eng Struct Dyn 2002;31:627–52. doi:10.1002/eqe.146.
[4] Montuori R, Nastri E, Piluso V, Todisco P. A simplified performance based approach for the evaluation of seismic performances of steel frames. Eng Struct 2020;224:111222.
[5] Lin YY, Miranda E. Estimation of maximum roof displacement demands in regular multistorey buildings. J Eng Mech 2010;136(1):1–11.
[6] Applied Technology Council. Quantification of building seismic performance factorsTechnical report FEMA 695. Washington, DC: Federal Emergency Management Agency; 2009.
[7] Applied Technology Council. Seismic performance assessment of buildings volume 1 - methodologyTechnical report FEMA P58-1. Washington, DC: Federal Emergency Management Agency; 2012.
[8] Applied Technology Council. Seismic performance assessment of buildings volume 2 - implementation guideTechnical report FEMA P58-2. Washington, DC: Federal Emergency Management Agency; 2012.
[9] Zhong K, Navarro JG, Govindjee S, Deierlein GG. Surrogate modelling of structural seismic response using probabilistic learning on manifolds. Earthq Eng Struct Dyn 2023;52:2407–28. doi:10.1002/eqe.3839.
[10] Sudret B, Chu VM. Computing seismic fragility curves using polynomial chaos expansions. 11th International Conference on Structural Safety and Reliability (ICOSSAR 2013); 2013. Eidgenössische Technische Hochschule Zürich.
[11] Gidaris I, Alexandros AT, Mavroeidis GP. Kriging metamodelling in seismic risk assessment based on stochastic ground motion models. Earthq Eng Struct Dyn 2015;44(14):2377–99.
[12] Gidaris I, et al. Multi-objective risk-informed design of floor isolation systems. Earthq Eng Struct Dyn 2016;45(8):1293–313.
[13] Du A, Padgett JE. Investigation of multivariate seismic surrogate demand modelling for multi-response structural systems. Eng Struct 2020;207:110210. doi:10.1016/j.engstruct.2020.110210.
[14] Vaseghiamiri S, Mahsuli M, Ghannad MA, Zareian F. Surrogate SDOF models for probabilistic performance assessment of multistorey buildings: methodology and application for steel special moment frames. Eng Struct 2020;212:110276. doi:10.1016/j.engstruct.2020.110276.
[15] Gudipati VK, Cha EJ. Surrogate modelling for structural response prediction of a building class. Struct Saf 2021;89:102041. doi:10.1016/j.strusafe.2020.102041.
[16] Aristizabal C, Lopez-caballero F. Comparison between two surrogate models for embankment earthquake-liquefaction-induced settlements prediction. 13th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP13); 2019.
[17] Dang-Vu H, Nguyen QD, Chung T, Shin J, Lee K. Frequency-based Data-driven Surrogate Model for Efficient Prediction of Irregular Structure's Seismic Responses. J Earthq Eng 2022;26:7319–36. doi:10.1080/13632469.2021.1961940.
[18] Cavalagli N, Pepi C, Gioffrè M, Gusella V, Ubertini F. Surrogate models for earthquake-induced damage detection and localization in historic structures using long-term dynamic monitoring data: application to a masonry dome. In: Proceeding of the 7th international conference on computing methods structure dynamics earthquake engineering (COMPDYN 2015), Athens: Institute of Structural Analysis and Antiseismic Research School of Civil Engineering National Technical University of Athens (NTUA) Greece; 2019. p. 1329–43. doi:10.7712/120119.7001.19117.
[19] Micheli L, Alipour A, Laflamme S. Multiple-surrogate models for probabilistic performance assessment of wind-excited tall buildings under uncertainties. ASCE-ASME J Risk Uncertainty Eng Syst, Part A: Civil Eng 2020;6(4):04020042.
[20] Micheli L, Alipour A, Laflamme S. Data-driven risk-based assessment of wind-excited tall buildings. Structures congress 2019: Blast, impact loading, and research and education. Reston, VA: American Society of Civil Engineers; 2019.
[21] Micheli L, et al. Surrogate models for high performance control systems in wind-excited tall buildings. Appl Soft Comput 2020;90:106133.
[22] Javidan MM, Kang H, Isobe D, Kim J. Computationally efficient framework for probabilistic collapse analysis of structures under extreme actions. Eng Struct 2018;172:440–52. doi:10.1016/j.engstruct.2018.06.022.
[23] Zhang R, Wang D, Qu C. Selection and modification of ground motion records using a weighted scaling method based on the Newmark-Hall target spectrum. Structures 2022;44:1546–64. doi:10.1016/j.istruc.2022.08.088.
[24] Zheng XW, Hong-Nan L, Sh Zhong-Qi. Hybrid AI-Bayesian-based demand models and fragility estimates for tall buildings against multi-hazard of earthquakes and winds. Thin Walled Struct 2023;187:110749.
[25] Xing L, Gardoni P, Zhou Y. Kriging metamodels for the dynamic response of high-rise buildings with outrigger systems and fragility estimates for seismic and wind loads. Resil Cities Struct 2022;1:110–22. doi:10.1016/j.rcns.2022.04.003.
[26] 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. doi:10.1016/j.engstruct.2021.112971.
[27] Bhosekar A, Ierapetritou M. Advances in surrogate based modelling, feasibility analysis, and optimization: a review. Comput Chem Eng 2018;108:250–67.
[28] Bruneau M, Uang CM, Sabelli R. Ductile design of steel structures. New York; Toronto: McGraw-Hill; 2011 http://www.myilibrary.com?id=335745 [Accessed 1 Feb 2023].
[29] American Institute of Steel Construction (AISC) Specification for structural steel buildings. ansi/aisc 360-16. Chicago, IL: American Institute of Steel Construction; 2016.
[30] SAC Joint Venture Technical Report November; 2000.
[31] SAC Joint Venture Technical report; 2000.
[32] Sawada Y, Shimizutani S. How do people cope with natural disasters? Evidence from the great Hanshin-Awaji (Kobe) earthquake in 1995. J Money, Credit Bank 2008;40(2–3):463–88.
[33] Rose A, Lim D. Business interruption losses from natural hazards: conceptual and methodological issues in the case of the Northridge earthquake. Environ Hazards 2002;4(1):1–14.
[34] Engelhardt M, Sabol T. Seismic-resistant steel moment connections: developments since the 1994 Northridge earthquake. Prog Struct Mater Eng 1997;1(1):68–77.
[35] Horton TA, Hajirasouliha I, Davison B, Ozdemir Z, Abuzayed I. Development of more accurate cyclic hysteretic models to represent RBS connections. Eng Struct 2021;245:112899.
[36] Stylianidis P, Bellos J. Survey on the Role of Beam-Column Connections in the Progressive Collapse Resistance of Steel Frame Buildings. Buildings 1696;13(7):2023.
[37] American Institute of Steel Construction, prequalified connections for special and intermediate steel moment frames for seismic applications, ANSI/AISC. 358–16.
[38] Guan Xingquan, et al. Seismic drift demand estimation for steel moment frame buildings: from mechanics-based to data-driven models. J Struct Eng 2021;147(6):04021058.
[39] ASCE 7-16 Minimum design loads and associated criteria for buildings and other structures. RestonVA: American Society of Civil Engineers (ASCE); 2016.
[40] Hamburger RO, Malley JO. Seismic design of steel special moment frames. NIST GCR 2009:09–917.

[41] Applied Technology CouncilQuantification of building seismic performance factors, US Department of Homeland Security. FEMA 2009;P695.
[42] American Institute of Steel Construction (AISC), Seismic provisions for structural steel buildings, ANSI/AISC. 2016:341–16. Chicago, IL, USA, 2010
[43] Kircher, C., Deierlein, G., Hooper, J., Krawinkler, H., Mahin, S., Shing, B., & Wallace, J., Evaluation of the FEMA P-695 methodology for quantification of building seismic performance factors, 2010.
[44] McKenna F, Fenves GL, OpenSees ScottMH. Open system for earthquake engineering Simulation. Pac Earthq Eng Res Center, Univ Calif 2006. http://OpenseesBerkeleyEdu.
[45] Ibarra LF, Medina RA, Krawinkler H. Hysteretic models that incorporate strength and stiffness Deterioration. Earthq Eng Struct Dyn 2005;34(12):1489–511.
[46] ASCE/SEI, ASCE 41-17 Seismic evaluation and retrofit of existing buildings. Reston, Virginia: American Society of Civil Engineers; 2017.
[47] Lignos DG. Sidesway collapse of deteriorating structural systems under seismic excitations. Stanford university; 2008.
[48] Lignos DG, Krawinkler H. Deterioration modelling of steel components in support of collapse prediction of steel moment frames under earthquake loading. J Struct Eng-Reston 2011;137(11):1291–302.
[49] Lignos DG, Krawinkler H. A database in support of modelling of component deterioration for collapse prediction of steel frame structures. In: Structural engineering research. Frontiers; 2007. p. 1–12.
[50] Gupta A, Krawinkler H. Seismic demands for the performance evaluation of steel moment resisting frame structures. Berkeley, CA: The John A Blume Earthquake Engineering Center, Stanford University; 1998.
[51] Rou WAkbas, B, Shen J. Practical moment–rotation relations of steel shear tab Connections. J Constr Steel Res 2013;88:296–308.
[52] Fayaz J, Zareian F. Reliability analysis of steel SMRF and SCBF structures considering the vertical component of near-fault ground motions. J Struct Eng 2019;145(7):04019061.
[53] Ellingwood B, Galambos TV, MacGregor JG, Cornell CA. Development of a probability based load criterion for American National Standard A58. Washington, DC: US Dept of Commerce; 1980.
[54] FEMA P-2012 Assessing seismic performance of buildings with configuration irregularities. Redwood City, CA: Applied Technology Council; 2018.
[55] Barbato M, Gu Q, Conte JP. Probabilistic pushover analysis of structural and soil–structure systems. J Struct Eng 2010;136(11):1330–41.
[56] Kim J, Park J-H, Lee T-H. Sensitivity analysis of steel buildings subjected to column loss. Eng Struct 2011;33:421–32. doi:10.1016/j.engstruct.2010.10.025.
[57] Zhang D, Lin X, Dong Y, Yu X. Machine-Learning-based uncertainty and sensitivity analysis of reinforced-concrete slabs subjected to fire. Structures 2023;53:581–94. doi:10.1016/j.istruc.2023.04.030.
[58] Trong Ha N, Xuan Hung D. Sensitivity analysis of the design portal frames of steel industrial buildings. MATEC Web Conf 2018;193:04025. doi:10.1051/matecconf/201819304025.
[59] Javidian MM, Kim J. Fuzzy Sensitivity Analysis of Structural Performance. Sustainability 2022;14:11974. doi:10.3390/su141911974.
[60] Rodríguez D, Brunesi E, Nascimbene R. Fragility and sensitivity analysis of steel frames with bolted-angle connections under progressive collapse. Eng Struct 2021;228:111508. doi:10.1016/j.engstruct.2020.111508.
[61] Tipu Kumar, R Panchal VR, Pandya KS. An ensemble approach to improve BPNN model precision for predicting compressive strength of high-performance concrete. Structures 2022;45:500–8. doi:10.1016/j.istruc.2022.09.046.
[62] Somala SN, Chanda S, Karthikeyan K, Mangalathu S. Explainable machine learning on New Zealand strong motion for PGV and PGA. Structures 2021;34:4977–85. doi:10.1016/j.istruc.2021.10.085.
[63] Le HA, Le DA, Le TT, Le HP, Le TH, Hoang HGT, et al. An extreme gradient boosting approach to estimate the shear strength of FRP reinforced concrete beams. Structures 2022;45:1307–21. doi:10.1016/j.istruc.2022.09.112.
[64] Truong GT, Choi KK, Kim CS. Implementation of boosting algorithms for prediction of punching shear strength of RC column footings. Structures 2022;46:521–38. doi:10.1016/j.istruc.2022.10.085.
[65] Rajbahadur GK, Wang S, Oliva GA, Kamei Y, Hassan AE. The impact of feature importance methods on the interpretation of defect classifiers. IEEE Trans Softw Eng 2021;48(7):2245–61.
[66] Saarela M, Jauhiainen S. Comparison of feature importance measures as explanations for classification models. SN Appl Sci 2021;3(272).
[67] Covert I, Lundberg S, Lee SI. Understanding global feature contributions with additive importance measures. Adv Neural Inf Process Syst 2020;33:17212–23.
[68] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceeding of the advances neural information processing system; 2017. p. 4765–74.
[69] Hooker G, Mentch L, Zhou S. Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. Stat Comput 2021;31:1–16.
[70] Breiman L. Random forests. Mach Learn 2001;45:5–32. doi:10.1023/A:1010933404324.
[71] Gromping U. Variable importance assessment in regression: linear regression versus random forest. Am Statistician 2009;63:308–19. doi:10.1198/tast.2009.08199.
[72] Parhi SK, Patro SK. Prediction of compressive strength of geopolymer concrete using a hybrid ensemble of grey wolf optimised machine learning estimators. J Build Eng 2023;71:106521. doi:10.1016/j.jobe.2023.106521.
[73] Mangalathu S, Hwang SH, Jeon JS. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. Eng Struct 2020;219:110927. doi:10.1016/j.engstruct.2020.110927.
[74] Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems; 2017. p. 4768–77.
[75] Cover T, Hart P. Nearest neighbour pattern classification. IEEE Trans Inf Theory 1967;13(1):21–7. doi:10.1109/TIT.1967.1053964.