Next-generation non-linear and collapse prediction models for short- to long-period systems via machine learning methods

Davit Shahnazaryan [1], Gerard J. O'Reilly [*],[2]

Scuola Universitaria Superiore IUSS di Pavia, Italy

ARTICLE INFO

ABSTRACT

Since the 1960 s, a cornerstone of earthquake engineering has been estimating the non-linear response of structures based just on lateral strength, modal properties, and the anticipated seismic demand. Over the years, several studies have quantified this empirical relationship and integrated it within seismic design and assessment methodologies. These have been widely accepted for practical application and adopted in building codes worldwide. While these models work reasonably well, there are still areas in which improvements can be made, especially concerning their robust quantification of uncertainty. This is mainly due to the amount of data used to quantify these empirical relationships, the choice of functional forms during the fitting, the model fitting and testing process, and how the ground motion shaking intensity is characterised. This study tackles these issues via the non-linear analysis of single-degree-of-freedom oscillators to train several machine learning (ML) models. This was to examine the accuracy and applicability of such models within a seismic engineering context and explore potential gains in quantifying the non-linear response of structures via next-generation intensity measures, namely average spectral acceleration, Sa_{avg}. The results show that the Decision Tree and XGBoost models worked well across a broad range of periods, accurately predicting collapse and non-collapse responses. Appraising these with existing models showed a notable improvement all around. It indicates that the models based on data-driven ML approaches represent a positive step and can be seamlessly integrated with seismic analysis methodologies utilised worldwide. Subsequently, a Python-based library of these XGBoost ML models was developed and made available online to foster practical application.

1. Introduction

One of the biggest challenges in earthquake engineering was accurately quantifying the non-linear dynamic response of ductile systems using a structure's linear elastic and static characteristics. This was an essential building block for many seismic design codes [1–4] and assessment guidelines [5–7], as it is the basis of several analysis methods. When the non-linear response is directly considered, methods like non-linear static pushover [8] scale the lateral force pattern to get a complete picture of structural behaviour from elasticity to yielding and towards collapse. Building on that, Bertero [9] mentioned the concept of using incremental time-history analysis, which was formalised through the advent of incremental dynamic analysis [10].

Much work has been conducted in the decades since then, and numerous researchers have proposed and calibrated models to estimate

the non-linear response via linear static methods. Some notable examples of design methods include the equivalent lateral force and response spectrum methods utilised in several building codes. In terms of assessment, research in the 1970 s spawned two different approaches to assessing the non-linear seismic demand in a structure. The first centred around an equivalent linear system or substitute structure [11]. This idealised the structure as a linearly responding oscillator with a reduced or secant stiffness in tandem with an overdamped elastic response spectrum to account for the energy dissipation during the non-linear behaviour of the actual structure. The level of damping typically depends on the level of non-linear demand and the hysteretic energy dissipation [12]. Methods utilising this approach include the capacity spectrum method [13] in addition to the SLaMa (Simple Lateral Mechanism Analysis) method [14]. The second pertains to directly estimating non-linear demand through inelastic response spectra. Methods that

---

* Corresponding author.
  E-mail address: gerard.oreilly@iusspavia.it (G.J. O'Reilly).
  [1] ORCID: 0000–0002-0529–5763
  [2] ORCID: 0000–0001-5497–030X

adopt this approach include the displacement coefficient method included in ATC-40, FEMA-273, FEMA-274, and ASCE-41 [8,15–17] and the N2 method [18] incorporated in Eurocode 8 [19].

Several researchers in the 1960 s developed inelastic spectra, typically as a function of the ratio to the elastic spectral demand, R, ductility demand, μ, and period, T, to form what are commonly referred to as R-μ-T relationships. Of note were Veletsos and Newmark [20], who noted that for medium to long-period structures, the inelastic spectral demand of a non-linear system tended to match the elastic spectral demand of the corresponding linear structure (i.e., $\mu \approx R$), giving rise to the so-called 'equal displacements' rule. However, for short-period structures, there tended to be an amplification of the non-linear demand (i.e., $\mu > R$). Work by Newmark and Hall [21] also noted this and developed means with which an inelastic spectrum could be derived from an elastic spectrum through simple analytical R-μ-T functions. Several research groups around this time developed similar functions from a few non-linear dynamic analyses on single degree of freedom (SDOF) oscillators [22]. Interested readers are referred to Miranda and Bertero [23] for a more in-depth review. In addition, Miranda [24] utilised elasto-plastic SDOF systems to derive constant-ductility spectra, where various earthquake source conditions were considered to account for their possible influence in developing these functions. Riddell et al. [25] derived models for response modification factors for short-period structures of 0.1 to 0.4 s. While formulated differently, the SPO2IDA tool [26,27] estimates percentiles of incremental dynamic analysis curves of SDOF systems from their piecewise static pushover curve, but the driving force behind the scenes of this tool is essentially R-μ-T relationships.

Among these various studies, the developed functions were often based on preselected functional forms. How appropriate these functional forms were was typically a user choice and not scrutinised extensively. Another limitation was the computational power availability in past decades, which resulted in few ground motion records being used to carry out non-linear dynamic analysis. For example, the observations of Newmark and Hall [21] were made mainly using the 1940 El Centro NS ground motion recording along with reportedly similar observations from other records. In the calibration of their model, Vidic et al. [28] utilised 20 ground motions, whereas Vamvatsikos and Cornell [26] utilised a suite of 30 ground motions, for example. Considering that the number of ground motion records used nowadays is a source of debate, with too few ground motions at times possibly being problematic [29]. With the immense increase in computational power and available ground motion recordings [30], some modest improvements are to be expected, especially intending to quantify not just average response but also its associated dispersion. Another limitation of these models is that the SDOF systems utilised generally employed non-degrading models meaning the impact of strength degradation and collapse was not considered; however, there have been some developments in recent years [26,31] whereby non-linear response up to complete collapse can be estimated. A further issue relates to the definition of the intensity measure (IM) used. Using a strength ratio (or reduction factor) R means that the spectral acceleration at the fundamental period, $Sa(T_1)$, is inherently assumed as the IM. Much research [e.g., [32]] has recently been conducted to develop more efficient IMs for quantifying non-linear response and collapse capacity. A common conclusion in these studies is that $Sa(T_1)$ tends not to be the best predictor of non-linear response or collapse, with a more robust average spectral acceleration, Sa_{avg}, showing itself to be more sufficient and efficient in this respect. Additionally, past studies used simple linear regression methods to fit the data. They typically involve training and validation of models on the same dataset resulting in possible bias and overfitting, as the introduction of new datasets may alter model accuracies. A recent study by Gentile and Galasso [33] used Gaussian Process regression on a large dataset of SDOF systems, where cross-validation (CV) was adopted to avoid these issues. However, the physical meaningfulness of such relationships based on engineering principles and knowledge of structural

dynamics should not be discounted as they allow engineers to appreciate the response generally without any need for large amounts of data as in the case of ML.

To alleviate the limitations of past studies described above, this study uses state-of-the-art machine learning (ML) methods by incorporating a vast suite of ground motions recorded on different soil conditions. Different ML methods were explored to examine their applicability and relative effectiveness in this context. The non-linear SDOF models were created using hardening and degrading branches to predict both non-linear and collapse capacity. The ML methods were trained and tested on different datasets of a similar type to avoid possible overfitting and reduction of quality if newer datasets are introduced. Additionally, the study aimed to develop models based both on $Sa(T_1)$ and Sa_{avg} to highlight the ML capabilities and ease of adaptability for various applications. In essence, the models trained and tested required no functional form, used a large suite of SDOF systems and ground motion records, while considering the degrading characteristics of non-linear model relationships using either $Sa(T_1)$ or Sa_{avg}. Another advantage of ML-based models allows for scalability, allowing for accommodating increased data and retraining the models without being constrained by a single functional form. Models were generated to predict both non-collapse and collapse response of systems assuming a trilinear backbone curve to account for post-peak degrading characteristics. Finally, the best predictive model was adapted within a Python-based tool and made available to the public for fast predictions.

2. Overview of the study

Following the brief review and motivation for further work on this topic using more robust and extensive analyses in the previous section, the development of R-μ-T and ρ-μ-T relationships followed the workflow presented in Fig. 1, where ρ represents dynamic strength ratio, which inherently assumes Sa_{avg} as the IM. The steps are described as follows:

Step 1: A large set of SDOF systems with varying backbone parameters was generated, including viscous damping ratio, ξ. The static response parameters (i.e., lateral yield force, F_y, and yield displacement, Δ_y) were used to generate non-linear models. Then, the linear elastic response parameters were normalised by the yield point to obtain the strength ratio, R, and ductility demand, μ, for a given ground motion record.

Step 2: A ground motion database was identified to cover a wide range of past earthquakes. Cloud analysis was performed on SDOF systems to obtain a peak dynamic response parameter. Over 26,000,000 scenarios were investigated, where the relationship between peak response in terms of peak displacement and IM in terms of $Sa(T_1)$ or Sa_{avg} was obtained. Two distinct Sa_{avg} values were utilised within this study (i.e., Sa_{avg3} for collapse capacity estimation [32] and Sa_{avg2} for moderate non-linearity [34] more suited to loss estimation) and are computed using Eq. (1):

$$Sa_{avg} = \left(\prod_{i=1}^{N} Sa(c_i T) \right)^{1/N} \tag{1}$$

where $N = 10$ and c_i is a linearly spaced coefficient ranging from 0.2 to 2.0 for Sa_{avg2} and 0.2 to 3.0 for Sa_{avg3}.

Step 3: The dataset generated via cloud analysis in Step 2 was split into training and testing sets. The data was analysed to understand specific relationships between the independent (i.e., SDOF backbone parameters) and dependent parameters (i.e., dynamic response parameters), where the independent variables were used to predict the dependent variables via several ML methodologies.

Step 4: Finally, the study concluded with comparative applications alongside existing models in the literature and considerations on how these results may be integrated within modern performance-based earthquake engineering (PBEE).
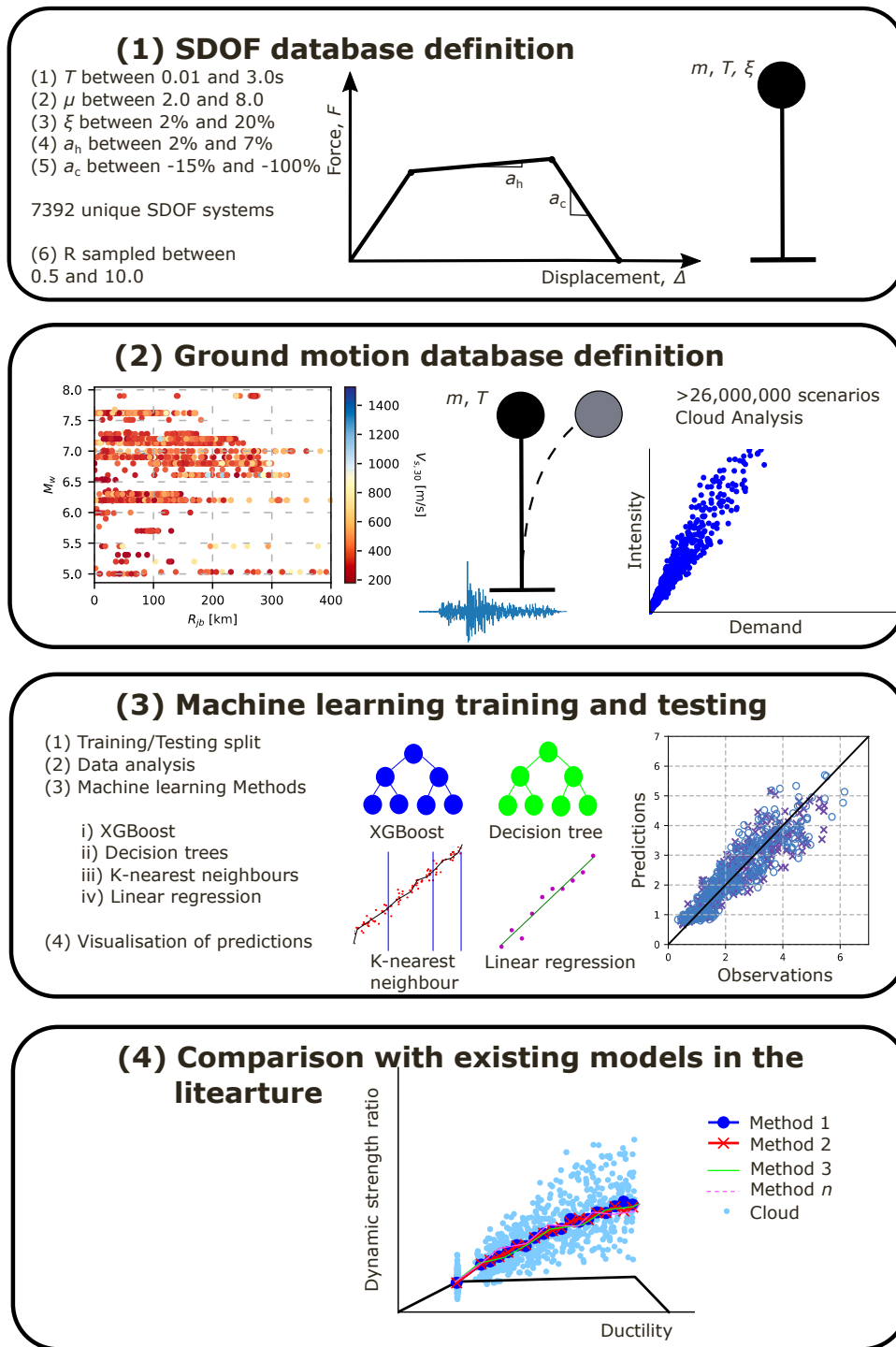
Fig. 1. Workflow followed within this study.

3. Generation of non-linear dynamic analysis datasets

3.1. Definition of SDOF oscillators

The first task in generating the non-linear dynamic analyses to be fed to the different ML algorithms was the definition of the structural systems to examine. For this purpose, simple SDOF oscillators were employed, whose backbone parameters are illustrated in Fig. 2. Since the study's goal was to establish a structural system's dynamic properties until lateral sideways collapse is reached, both post-yield hardening and post-peak degrading branches were considered to result in the trilinear backbone illustrated. This was instead of using typical elastic-perfectly-plastic or elastic-hardening backbones with infinite ductility capacity used in past studies [24,28,33,35]. It was deemed more representative of the typical systems encountered in seismic design and assessment, whose ductility capacity depends on the structural configuration and the typology used. Therefore, not only were the elastic properties necessary, but also the plastic or hardening branch describing the post-yield ductility of the system and a negative stiffness branch, characterising brittle or ductile collapse towards gradual strength loss. Countless experimental test campaigns [36–38] and numerical modelling simulations [39,40] have repeatedly shown this to be a typical
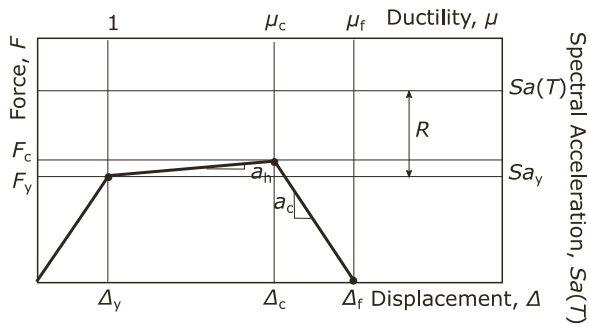
Fig. 2. Trilinear backbone curve of the SDOFs utilised for non-linear dynamic analyses.

backbone behaviour for various structural typologies, indicating the general applicability of the developed model and hence was utilised herein.

Fig. 2 illustrates the shape of the hysteretic backbone in terms of lateral force, *F*, and displacement, *Δ*. It consists of a linear elastic branch until the nominal yield. The force and displacement quantities are also normalised by these yield point values to result in the *R* and μ. It is followed by a hardening branch characterised through a slope a_h up to the peak capacity at a ductility μ_c. This corresponds to the load-deformation curve's peak strength and the structural degradation initiation. This softening branch is defined by the post-capping stiffness a_c, which leads to zero strength corresponding to the failure ductility, μ_f. Fig. 3 presents the cyclic pushover curves of several hysteretic models.

3.2. Backbone parameters

With a general definition of the SDOF oscillator, the next step was to define the range of parameters to be analysed. First, the period of

vibration, *T*, of the systems was varied between 0.01 s, 0.05 s, 0.1 s, 0.25 s, 0.5 s, 0.75 s, 1 s, 1.5 s, 2 s, 2.5 s and 3 s, while *ξ* of the system was varied between 2%, 5%, 7%, 10%, 15%, and 20%. This was to ensure that the developed model could apply to both very short and also moderately long-period structures. It is noted that oscillators with extreme values of short and long period correspond to very stiff and very flexible structures, whose non-linear cyclic behaviour may not be entirely representative of the hysteretic rule adopted here; therefore, these cases are provided more with the objective of highlighting trends in such zones and their use in practice should be handled with care. Guerrini et al. [35], for example, have highlighted the particularities of non-linear response for very short period structures; hence, the lower bound value of *T* = 0.01 s was explored here, with a limitation that practically may be deemed unsuitable for *T* lower than 0.05 s. Likewise, periods of up to 3 s can be anticipated for some structures depending on their typology and overall height. However, it is likely that the first mode dominance will not be so strong in these structures, and SDOF representations may begin to lose some of their predictive power.

The next step was to define the range of values for μ_c. Looking at the anticipated ductility capacity of some structural systems [41] and subsequently in the recommended building code prescriptions [3,19,42], a range of μ_c values between 2 and 8 was adopted with an increment of 1. To limit the number of analyses, although further extension is possible and planned for the continuous update of the proposed ML model, the hardening and post-capping stiffnesses were initially set as $a_h = 2\%$, 5%, 7%, 10% and $a_c = -100\%, -50\%, -30\%, -15\%$.

To determine the strength of the SDOF system expected to undergo non-linear response, an *R* was defined with respect to each ground motion record employed. That is, an *R* value in the range of 0.5 to 10 was randomly sampled, meaning that the lateral strength of each SDOF is given by Eq. (2), where the *Sa(T)* is the spectral acceleration of each individual record and Sa_y is the yield spectral acceleration of the SDOF system, illustrated in Fig. 2. With the period *T* and lateral yield strength
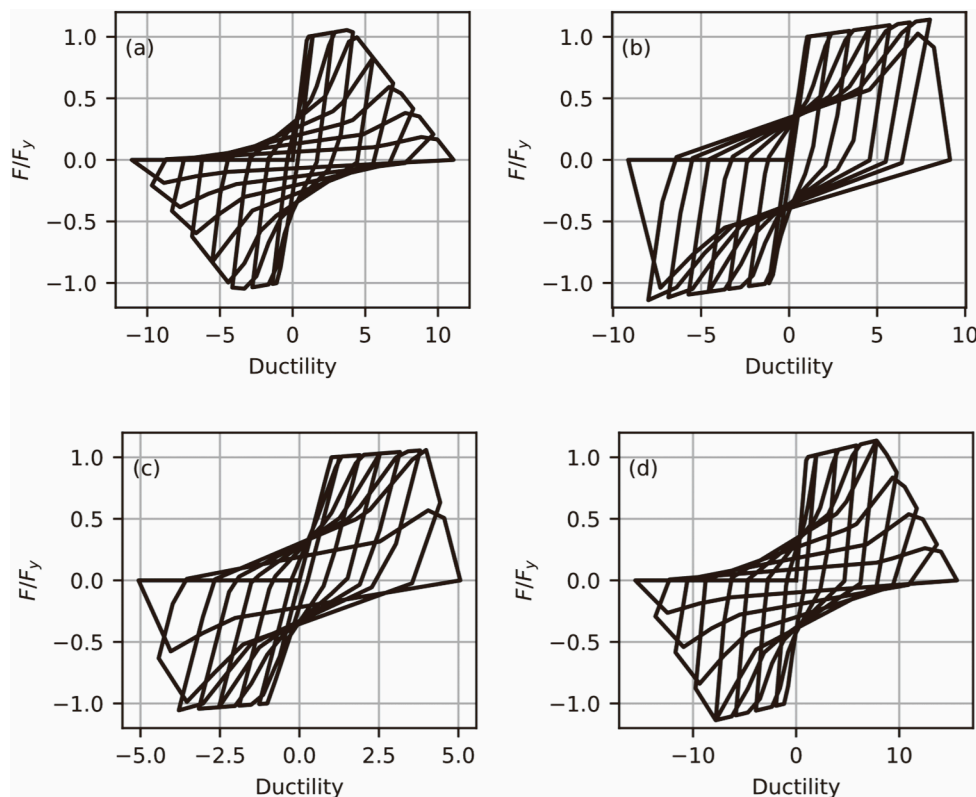

Fig. 3. Examples of reference hysteretic models, where $a_h = 2\%$: (a) *T* = 0.1 s, $a_c = -15\%$, $\mu_c = 4$; (b) *T* = 0.5 s, $a_c = -100\%$, $\mu_c = 8$; (c) *T* = 1.0 s, $a_c = -100\%$, $\mu_c = 4$; (d) *T* = 2.0 s, $a_c = -15\%$, $\mu_c = 8$.

F_y known, the mass, m, of the system was set at a nominal value of 100 tonnes and the yield displacement was computed according to Eq. (3). The considered variations resulted in 11 T values, 6 values of ξ, 7 values of μ_c, 4 values of a_h, and 4 values of a_c, amounting to 7292 unique SDOF systems.

$$F_y = mSa_y = m\frac{Sa(T)}{R} \tag{2}$$

$$\Delta_y = \left(\frac{T}{2\pi}\right)^2 \frac{F_y}{m} \tag{3}$$

Note that these parameters were independently sampled for the stated ranges to construct the SDOF oscillators to analyse. Possible correlations between some of these parameters (e.g., period, strength ratio, ductility capacity) as implied by some common design methods are not investigated here, but the discussions later in Section 4 may provide some insights in this regard.

3.3. Numerical modelling

With the definition of the SDOF system backbone parameters outlined, the numerical implementation to conduct the non-linear dynamic analysis is described. Numerical models were built using the OpenSeesPy [43] framework. The backbone was defined as per Fig. 2 using the uniaxial Hysteretic model. No cyclic stiffness or strength degradation were considered, and the pinching characteristics during reloading corresponded to 0.8 and 0.5 for deformation and force, respectively, which have been observed to work well when compared to the pinching observed in fibre-based models [44]. This differs from recent work by Gentile and Galasso [33], for example, who opted to make the hysteretic behaviour an input parameter in their predictive model. A viscous damping ratio was modelled as tangent stiffness proportional.

3.4. Ground motion database

To conduct the non-linear dynamic analysis, an ample and suitable ground motion database was required. As noted previously, many past models tend to use limited sets of ground motions to develop such R-μ-T models, or the ground motion databases available at that time. More recently, the NGA West2 database [30] has been made available and contains over 21,000 ground motions from active crustal regions from several events worldwide. This entire database was adopted for this study, but several criteria were applied to exclude ground motion records that may be deemed unsuitable for general use. As such, the following filtering criteria based on the recommendations of Davalos and Miranda [45] were implemented:

- Only ground motions recorded during earthquakes with magnitudes $M_w \geq 5$ were used. The lower magnitude events were neglected as they would not be expected to be sufficiently intense to produce large non-linear deformations or collapse in engineered structures without scaling.
- Near-fault pulse-like ground motions were excluded.
- As mentioned in Section 2, the average spectral acceleration IM will be adopted. The computation of Sa_{avg} is based on spectral ordinates at periods much larger than the anchoring period of vibration, T; therefore, only records whose minimum usable frequency of both components was smaller than $1/(2\,T)$ or $1/(3\,T)$ were considered. This limit varies depending on which definition of Sa_{avg} (i.e., Sa_{avg3} for collapse prediction and Sa_{avg2} for moderate non-linearity) is used.
- Only ground motions recorded on sites with mean shear wave velocities in the upper 30 m, $V_{s,30}$, larger than 180 m/s, corresponding to NEHRP [46] site class A to D or Eurocode 8 [19] site class A to C, were considered.
- Only recordings from strike-slip, reverse, and reverse-oblique events from active shallow crustal tectonic environments were included.

- Ground motion recordings were constrained based on instrument location. This was based on the Geomatrix 1st letter code of the NGA-West2 flat-file. Recordings were included if they came from instruments located on free field or either below the surface or in the first storey of low-rise structures (less than four storeys).
- Earthquakes that were recorded at a single station were excluded.
- Due to the arbitrary orientations of ground motion recording instruments with respect to active faults, only one arbitrary direction of a record pair was considered. Therefore, the IM was assumed to be an arbitrary component.
- The selected records were not conditioned on the 1999 Chi-Chi earthquake despite the significant amount of recordings associated with the earthquake. As stated by Davalos and Miranda [45], if some of the records were to be removed, the dataset will be smaller and can lead to undersampling, while keeping the entire record set corresponding to the 1999 Chi-Chi earthquake can influence the predictions undesirably. However, Boore and Atkinson [47] have investigated the influence of Chi-Chi earthquake in detail and have concluded that the predictions were largely influenced only at longer periods of vibration (i.e., $T > 5$ s) and when the percentage of the recordings exceed 40% of the total selected set.

The final period-dependent ground motion database ranged from 2424 to 3527 records corresponding to periods of 3 s to 0.01 s of Sa_{avg3}. It is worth noting that given T, the ground motions selected for Sa_{avg3} were automatically eligible for Sa_{avg2} as the lower usable frequency is higher for the latter. Fig. 4 demonstrates the moment magnitude, M_w, versus Joyner-Boore distance, R_{jb}, distributions of records for periods of $T = 0.25$ s and 3.0 s, and IM of Sa_{avg2} and Sa_{avg3}, where the $V_{s,30}$ is also indicated. Below $T = 0.25$ s, the number of records did not change for both IMs and amounted to 4222. This reduction is related to eliminating records with the lower usable frequency associated with higher T. Additionally, with the decrease in lower usable frequency, the number of records at higher values of R_{jb} is diminished. With the number of unique SDOF systems outlined in Section 3.2, over 26,000,000 non-linear dynamic analyses were conducted.

3.5. Analysis procedure

With the SDOF backbone, numerical modelling strategy and the ground motion database all described, the last remaining step was conducting the analysis. The steps below outline the process followed to define each SDOF, conduct the non-linear dynamic analysis and the output data that was collected and stored for each case. It is noted that all records were as-recorded and no amplitude scaling was used.

Step 1: For each ground motion and randomly sampled R, the spectral acceleration $Sa(T)$ was computed and used to identify F_y of the SDOF system using Eq. (2).

Step 2: The Δ_y corresponding to F_y was computed using Eq. (3).

Step 3: For a given value of peak ductility μ_c, the displacement and strength at the end of the hardening branch were computed following Eqs. (4) and (5).

$$\Delta_c = \mu_c \Delta_y \tag{4}$$

$$F_c = F_y(1 + (\mu_c - 1)a_h) \tag{5}$$

Step 4: The displacement at the failure ductility corresponding to lateral collapse was calculated by Eq. (6).

$$\Delta_f = \Delta_y \left(\mu_c + \frac{\Delta_c}{\Delta_y a_c}\right) \tag{6}$$

Each hysteresis model realisation (T, ξ, R, μ_c, μ_f) was analysed through non-linear dynamic analysis using the ground motion database. Following each analysis, the maximum absolute displacement was recorded and used to compute the dynamic ductility demand, μ_{dyn}. If μ_{dyn} exceeded μ_f, then the model was considered to have collapsed and
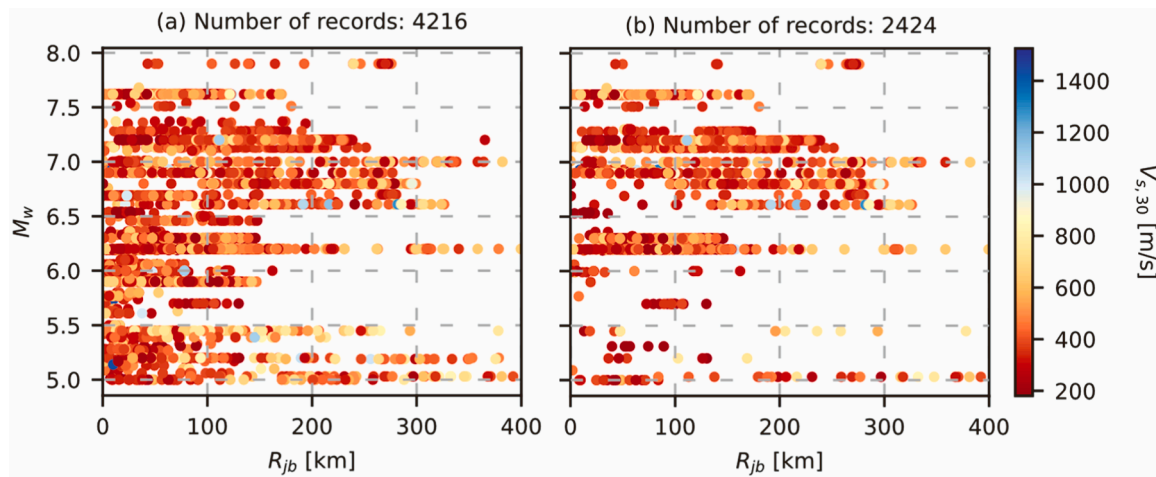
Fig. 4. M_w and R_{jb} distribution of records used for (a) $T = 0.25$ s and Sa_{avg2}; and (b) $T = 3.0$ s and Sa_{avg3}.

only its intensity was noted; otherwise, the model was considered to have not collapsed and both its intensity and ductility demand were stored, discussed further in Section 4.1. Additionally, the average spectral accelerations, Sa_{avg2} and Sa_{avg3}, were also stored for each case and their dynamic strength ratios [48] were computed as:

$$\rho_2 = Sa_{avg2}/Sa_y \tag{7}$$

$$\rho_3 = Sa_{avg3}/Sa_y \tag{8}$$

4. Algorithmic predictions of dynamic strength ratio

4.1. Methodology

Following extensive numerical analysis and creation of the database, the ML framework for the derivation of R-μ-T and ρ-μ-T relationships consists of the following steps: (1) identification of independent and dependent variables; (2) splitting of the dataset into training and testing sets; (3) definition of feature engineering pipelines; (4) description of ML modelling approaches; (5) application of k-fold CV to ensure that the score of the model does not depend on how the training and testing subsets were selected; (6) selection of the best-performing model based on validation metrics; and (7) interpretation of results via feature importance assessment and finally a demonstration of the developed R-μ-T and ρ-μ-T relationships. Fig. 5 illustrates each of these steps in the

procedure, which are described further below. A Python-based tool [49] was developed and available on PyPI (https://pypi.org/project/xgb-rhomut/) and for simple pip installation in Python. Any updates and future extensions to these models' applicability will be adequately documented within the tool.

Step 1 identifies the independent and dependent variables. The independent variables included the characteristics of the SDOF system, T; ξ, a_h; μ_c and μ_f; and the observed peak ductility demand during the non-linear dynamic analysis, μ_{dyn}. The dependent variables were ρ_2 and R, which were developed to predict the non-linear demand in non-collapsing structures, and ρ_3, which was developed solely for predicting the collapsing capacity of structures. The dependent variables were selected to estimate the intensity distribution needed to surpass a specific ductility demand in the structure. This aligns with risk-oriented design and assessment methods mentioned in Vamvatsikos et al. [50]. To do this, non-collapsing and collapsing analysis cases were separated to train distinct algorithms. As a result, the dataset containing only non-collapsing cases was reduced to 9509,777 and the dataset containing only collapsed cases included 17,984,829 data points. The statistical distributions of all independent and dependent variables are provided in Table 1.

Step 2 involved splitting the dataset into training and testing subsets. In past studies and generally in ML applications, good results are obtained when the dataset is split between 70–80% training and 30–20% testing sets. However, given the large dataset, even a lower percentage
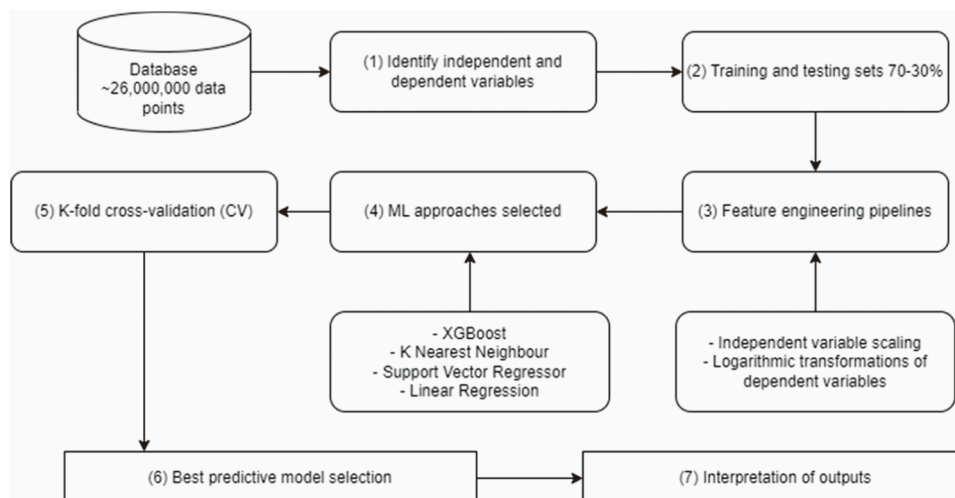

Fig. 5. Illustration of the steps followed in the application of ML.

Table 1
Statistical information of numerical independent and dependent variables.

Variable	Minimum	Maximum	Mean	Standard Deviation	Type
T	0.01	3.00	0.91	0.94	independent
ξ	2.00	20.00	9.81	6.09	independent
a_h	2.00	10.00	6.00	2.92	independent
μ_c	2.00	8.00	5.02	2.00	independent
μ_f	3.02	19.33	9.06	3.62	independent
μ_{dyn}	0.26	19.33	3.23	2.33	independent
ρ_2	0.10	24.88	3.31	2.46	dependent
ρ_3	0.65	26.08	6.65	3.11	dependent
R	0.50	10.00	5.22	2.74	dependent

of testing set may be considered. To obtain optimal models, the models were trained using multiple scenarios of splits and indicated similar results considering testing sets from 5% to 30%. The splits were further validated by changing the split's random seed, where the consistent results indicated that the splitting methodology was stable and a value between the selected range could be considered for training the models. As a result, a 70–30% ratio of training and testing sets was adopted.

In Step 3, feature engineering pipelines were identified to prepare the dataset for training. Pipelines help avoid data leakage during the pre-processing steps so that the trained model may be safely used for unseen test and validation datasets. This feature engineering generally involves cleaning of data, scaling, imputations of missing data points, and transformations. However, given the nature of the dataset, no data was missing, and the only cleaning involved setting aside the collapsing and non-collapsing scenarios depending on the dependent variables. In ML, the model maps the data from independent to dependent variables. The larger differences between data points of independent variables increases the uncertainty of the results in the model, therefore scaling is applied to reduce those differences. To have similar variance for all independent variables, the independent dataset was scaled between 0 and 1 using Eq. (9), where x is the original value of the independent variable and x' is the normalised value, indicating significant improvement in trained models. Subsequently, the dataset had smaller standard deviations. Additionally, scaling helps the ML algorithms converge using fewer iterations. Pertaining to the dependent dataset, a logarithmic transformation was utilised (Fig. 6) using Eq. (10). It is typically used to make highly skewed distributions into a more normal one. Additionally, incorporating the natural logarithm of one plus helps prevent negative values in the model predictions.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \qquad (9)$$

$$y = \ln(1 + y) \qquad (10)$$

To avoid multicollinearity, which describes the occurrence of high correlations among two or more independent variables and recognising

possible relationships among independent and dependent variables, a correlation matrix is displayed in Fig. 7. As seen, no strong correlations can be observed between any of the independent variables, except for μ_c and μ_f, which is expected since they describe similar features of the SDOF system, as in ductility; however, μ_c is important up to peak capacity, μ_f is important for collapse prediction. Therefore, μ_f was removed from the independent variable set. Additionally, there is a strong relationship between μ_{dyn} and dependent variables for non-collapsing scenarios and it is expected to impact significantly the trained models. For what concerns the collapse scenarios, as already mentioned μ_{dyn} is not considered. Interestingly, the relationship between T and dependent variables is not very pronounced. All things considered, the correlation matrices shown in Fig. 7 are an exploratory step to visualise the strength of the relationships among the variables. While it might seem that there is no strong correlation between T and some of the dependent variables, it is important to note that a weak correlation does not necessarily imply the absence of a relationship but could be an implication of no directivity between an independent and dependent variable.

Step 4 was conducted by training several parametric and non-parametric ML (assuming non-functional form) models to perform the multiple regression to predict multiple continuous dependent variables. The employed ML models included: Linear regression (LR); K-nearest neighbour (KNN); Decision trees (DT); Extreme gradient boosting (XGBoost). LR is part of the parametric family approach, while the rest could be categorised as non-parametric. Parametric approaches are suitable but not limited to data with fairly well-known distribution, while the non-parametric approaches could be more powerful for relationships where knowledge is limited and underlying distributions do not adhere to any specific distribution. Looking at the distributions of dependent variables with respect to μ_{dyn} in Fig. 8, a trend is observed, whereby by increasing μ_{dyn}, the values of dependent variables increase, as well as their dispersion.

Multiple LR involves two or more independent variables to predict a single dependent variable. A linear relationship is assumed between the dependent and independent variables, which while not observed from Fig. 8, is still not far off. The prediction is based on Eq. (11), where $\widehat{y}$ is the predicted variable, b_0 is the y-intercept, x_i are the independent variables, b_i are the slope coefficients for each independent variable, and ε is the model's residuals.

$$\widehat{y} = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_n x_{in} + \varepsilon \qquad (11)$$

KNN is a non-parametric model that approximates the relationship between independent variables and the dependent variable by averaging the observations within the same neighbourhood. The size of the neighbourhood can be selected through CV to minimise the error. KNN uses feature similarity to predict the values of new data points, meaning that the new point is predicted based on how closely it resembles the points in the training set. Essentially, the closest data points are selected, the average of which is assigned to the predicted point. There are several
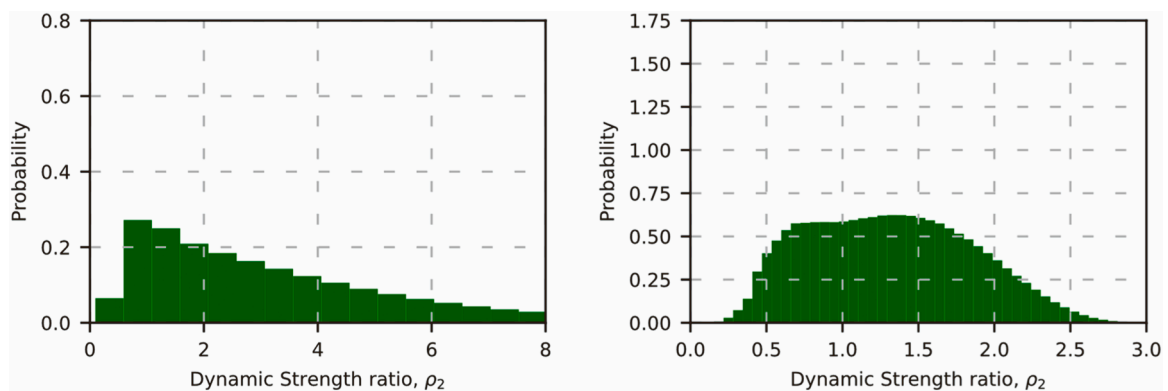

Fig. 6. Distribution of ρ_2 (left) before and (right) after logarithmic transformation.
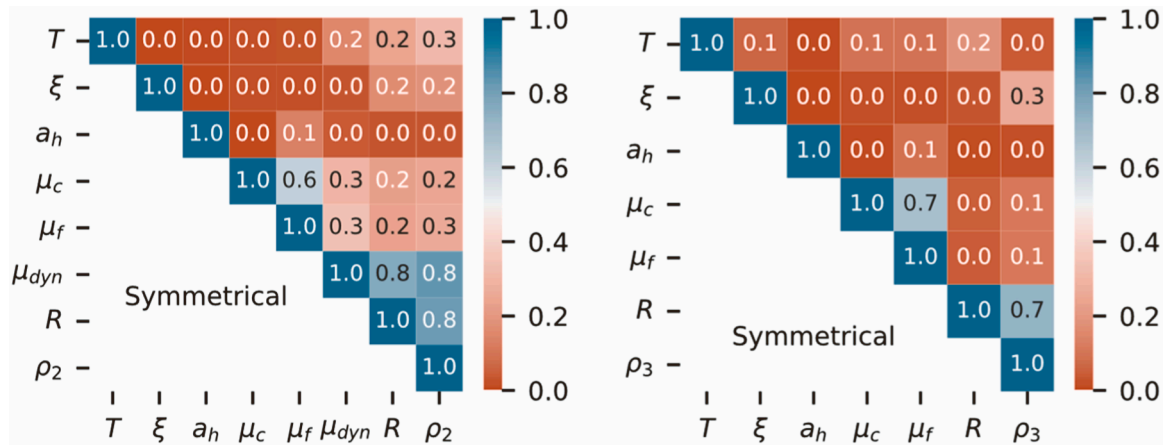
Fig. 7. Correlation between independent and dependent variables for (left) non-collapse and (right) collapse cases.
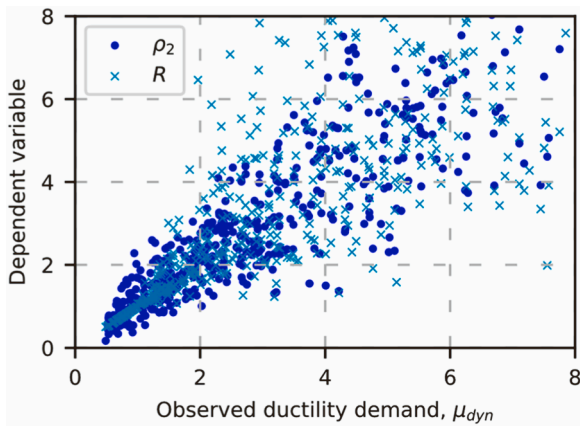

Fig. 8. Relationship between independent and dependent variables for non-collapse.

methods in calculating the distance between the points for numerical variables, which include: Euclidean distance, which is calculated based on the square-root-sum-of-the-square differences between the new and an existing point; Manhattan distance, which is the distance between real vectors using the sum of their absolute difference. The number of points inside the neighbourhood, *k*, is a hyperparameter that needs to be tuned for optimised outputs and minimised error.

Similar to KNN, DT is a non-parametric approach that performs regression using a tree-like structure. It splits the dataset into small subsets starting from the root node, which represents the entire dataset. It is then followed by a branched tree, each of which breaks down to two decision nodes, and eventually leads to a leaf node, where the dataset is no longer split and contains the prediction of the algorithm. The method is based on bootstrap aggregation or bagging that uses sub-samples to create decision trees and make the final prediction as the average of predictions of all trees.

For what concerns XGBoost [51], it is a scalable ML model for tree boosting. Boosting is an ensemble technique of adding new models to correct the residuals observed in past models. The models are added sequentially until no significant improvement is made. At the core of the XGBoost lies a gradient descent algorithm that minimises the error when new models are added. Another advantage of XGBoost is its scalability, meaning that it is not computationally expensive to carry out multiple fold CVs on large numbers of datasets and large boosting rounds on a single machine. The scalability is attributed to several important systems and algorithmic optimisations available within XGBoost. Such features include its parallel and distributed computing capabilities, as well as

out-of-core computations that make the learning process significantly faster in comparison with other available learning approaches. This can potentially allow for more model exploration, analysis, and subsequent accuracy increment. The superior performance of XGBoost has already been shown in many past studies in the field of earthquake engineering [52–54].

To optimise and improve the quality of the trained models, hyperparameter tuning was performed on the training set through a grid search algorithm employing a repeated *k*-fold CV (Step 4 of Fig. 5). Generally, five to ten folds are recommended in the literature [55] but given the extensive dataset employed here, just one or two folds may even be employed to achieve good results. As a compromise, five folds were adopted within this study aiming to improve the accuracy of the models moderately with no great extra cost. The repeated CV is a procedure where the mean results are reported across all folds and runs. In the basis of CV lies the division of the training subset into *k* non-overlapping folds. All folds bar one were used to train the model, where one was used for validation. Essentially, *k* models were fitted, which were then evaluated on *k* hold-out validation sets based on which the mean performance was computed. Several hyperparameters were selected to perform the tuning, including the number of boosting rounds or trees (weak learners) to build, early stopping rounds (training is stopped once the performance is no longer improving for certain rounds), maximum depth of a tree, minimum weight required to create a new node in the tree, and the learning rate. Through a grid search, all possible combinations were identified and iterated to train the model and apply the *k*-fold CV to identify a combination resulting in the best performance in terms of minimised error.

For what concerns the error, a goodness-of-fit test was implemented using mean absolute error (MAE) (Eq. (12)) and the coefficient of determination, R^2 (Eq. (13)) for ranking the training models (Step 5 of Fig. 5), where y_i is the ith observed value, $\widehat{y}_i$ is the ith predicted value, and $\overline{y}$ is the mean value of n data points. MAE is an indicator of the relative error between the predicted and observed values, with smaller values indicating a better result. In contrast, a higher R^2 value is desired, since it is a statistical measure of the proportion of the dependent variable's variance that is explained by the independent variable. Finally, with the metrics identified, the best-performing models were selected as the last step of the ML methodology adopted (Step 6), leading towards interpreting results (Step 7), as per Fig. 5.

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \widehat{y}_i|}{n} \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{13}$$

4.2. Results and discussions

To select the best-performing models, the dependent variables (Table 1) were predicted using the ML models described in Section 4.1 and appraised via MAE and R^2 values in Fig. 9. The MAE is computed on dependent variables, that were transformed to log scale. In general, each approach, except from LR, performed well in terms of both metrics indicating low bias. This meant that there was no underfitting, in addition to the proximity of the metrics for testing and training sets indicating low variance (no overfitting) towards the training set. Even though one could make a case of using a simpler analytical or ML model, such as LR, relatively lower scores highlight the opposite. Additionally, one advantage of the developed models over existing analytical approaches is the use of a testing (i.e., unseen) set for the validation of results. It goes without saying that analytical methods could employ unseen data sets for validation as well, but this has generally not been adopted in past approaches [24,28,35]. Another key aspect is related to the DT approach, where significantly high scores are obtained for training sets, highlighting some level of overfitting (variance); however, given the slight overfitting, the model performed slightly better in terms of both scores with respect to the other methods. There are several methods to avoid overfitting in DTs. If no bounds are set on DTs, the model will memorise the noise of the training set and fail to capture important patterns resulting in almost 100% accuracy on the training set because the tree will grow to its full depth. Therefore, the predictions on the testing set could result in subpar performance. Pruning and use of random forest could help mitigate the overfitting. The former involves trimming off branches of the tree starting from the leaf node so that the overall accuracy is not disturbed. Within this study, a maximum depth of 10 was used, but further hyperparameter tuning may be employed to increase the quality of the models. The latter is an Ensembling learning approach and is more akin to XGBoost. Given the overall high scores in terms of both metrics, it was decided not to perform further pruning or use a second Ensembling method within this study, as the overall trend of well-performing ML methods can be observed from Fig. 9. It is noted that the errors are based on predictions of scaled dependent variables.

As seen in Fig. 9, XGBoost, KNN and DT performed better compared to the LR, while KNN performed with some overfitting compared to DT and XGBoost as the R^2 and MAE scores were further apart for training and testing sets; therefore, the latter methods were utilised to visualise the predictions in further detail. Since the dependent variables were transformed using logarithmic transformations during training, they were inversed using Eq. (14). Fig. 10 plots the predictions and accuracy metrics for dependent variables ρ_2, ρ_3 and R. For visualisation purpose, only 500 data points are shown for both training and testing sets, which were randomly sampled from the entire dataset to avoid excessive data points being plotted and making the plots difficult to interpret. Nonetheless, the metrics listed still refer to the entire dataset. Fig. 10 shows that relatively good predictions are observed for both the DT and XGBoost methods when predicting both dependent variables. There is some dispersion observed for both methods with regard to the testing

set. Since there is some scatter of predictions with respect to the testing, the R^2 score cannot be used as the only reliable metric. This is because, despite the methods managing to predict the dependent variables' variance with relative ease, the plots highlight that many predictions are not exactly on the diagonal line. However, because each testing data point is quite close to the projected diagonal line, which is also indicated via small MAE values, it may be concluded that both models achieved their goals in predicting the dependent variables with good accuracy.

$$y = e^y - 1 \tag{14}$$

Another interesting output of the models could be validated with respect to the non-linear dynamic analysis results for a specific SDOF scenario. Fig. 11 plots the actual observations and the predicted values from the XGBoost model using only the testing dataset. As observed, the predictions of the XGBoost models match the median response of the dynamic analysis very well. The median values of the dynamic analysis were computed by binning the clouds of data at ductility bin widths of 0.25 and computing the median of the data. This confirms the accuracy of the fitted models with respect to the dataset on a more physical and engineering-based representation. Additionally, Fig. 11 plots the collapse capacity prediction based on ρ_3 using the XGBoost model. Here, the XGBoost model was trained to predict the median, η, capacity of the SDOF, while the dispersion was inferred from non-linear analysis results.

4.3. Comparison with existing R-μ-T models

As mentioned previously in Section 1, one of the biggest motivations for this study was to implement ML-based techniques and illustrate how they can improve the empirical models used in seismic design and assessment worldwide. In many instances, this improvement can be achieved without any radical shift in analysis method or design philosophy, but by simply substituting the proposed R-μ-T models. Several comparisons are provided here to motivate this to illustrate how the proposed XGBoost model performs against the empirical data for a select number of scenarios along with several existing R-μ-T analytical models. Only the XGBoost model was utilised herein as it has been shown previously in Fig. 9 and Fig. 10 to be among the top performers. This comparison is only carried out for R-μ-T models since no other ρ-μ-T model exist (at least for bilinear systems [48]) with which to compare. Section 5 will discuss ρ-μ-T models further and their motivation for adoption in practice. Nevertheless, the comparison of R-μ-T models here is useful to illustrate the general benefit of the ML-based approach to non-linear demand quantification. Fig. 12 presents comparative plots of the predictions of XGBoost model trained within this study and several analytical models available in the literature [1,21,23,28,35,56]. XGBoost model predictions generally lie within similar bounds compared to the existing analytical models for short and medium periods as well as low and high ductility systems. A distinguishable difference of the XGBoost model comes in non-smooth predictions, which is a direct consequence of using a non-parametric approach with no analytical formulation and is based on minimising the error between the
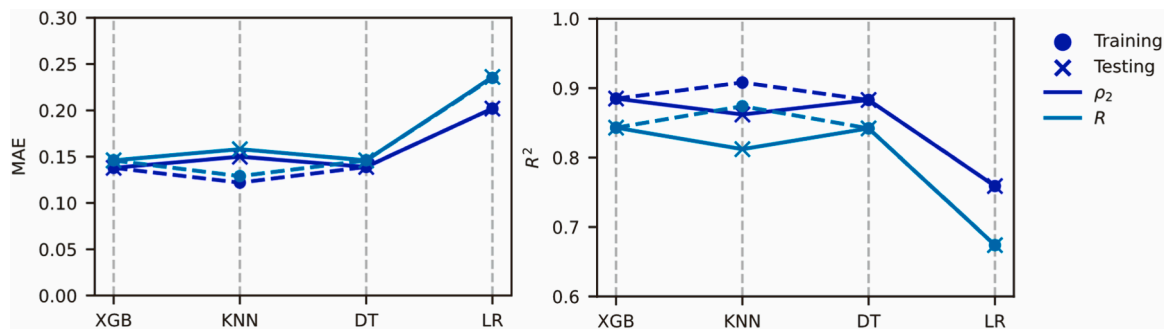

Fig. 9. Training and testing (left) MAE and (right) R^2 values of the applied ML models on the non-collapse scenario.
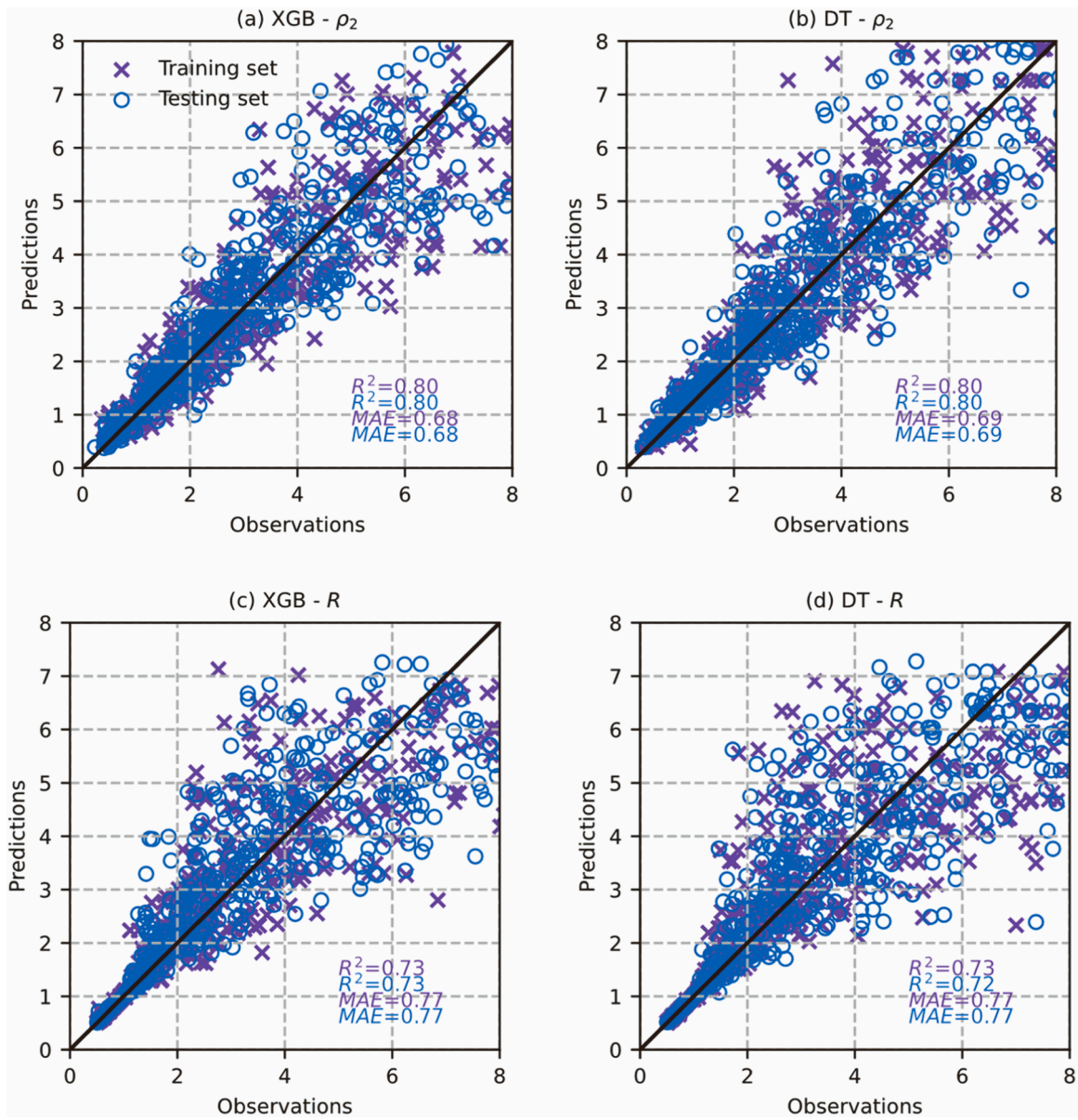
Fig. 10. Observed and predicted values of dependent variables: (a) XGBoost – ρ_2; (b) DT – ρ_2; (c) XGBoost –R; (d) DT – R.
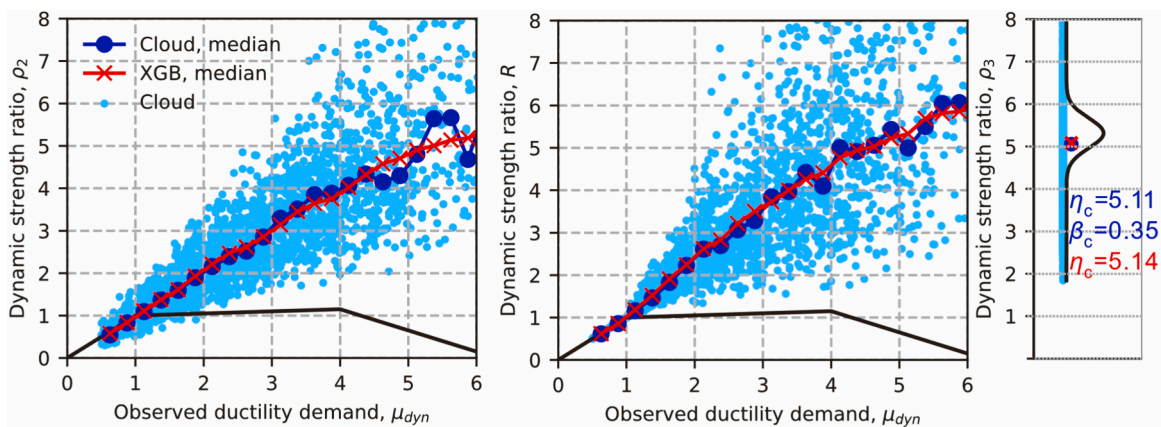

Fig. 11. Observed demands for (a) ρ_2; (b) R; and (c) η of collapse; given a specific SDOF scenario ($T = 1.0$ s, $\xi = 5\%$, $a_h = 5\%$, $\mu_c = 4.0$) using XGBoost.
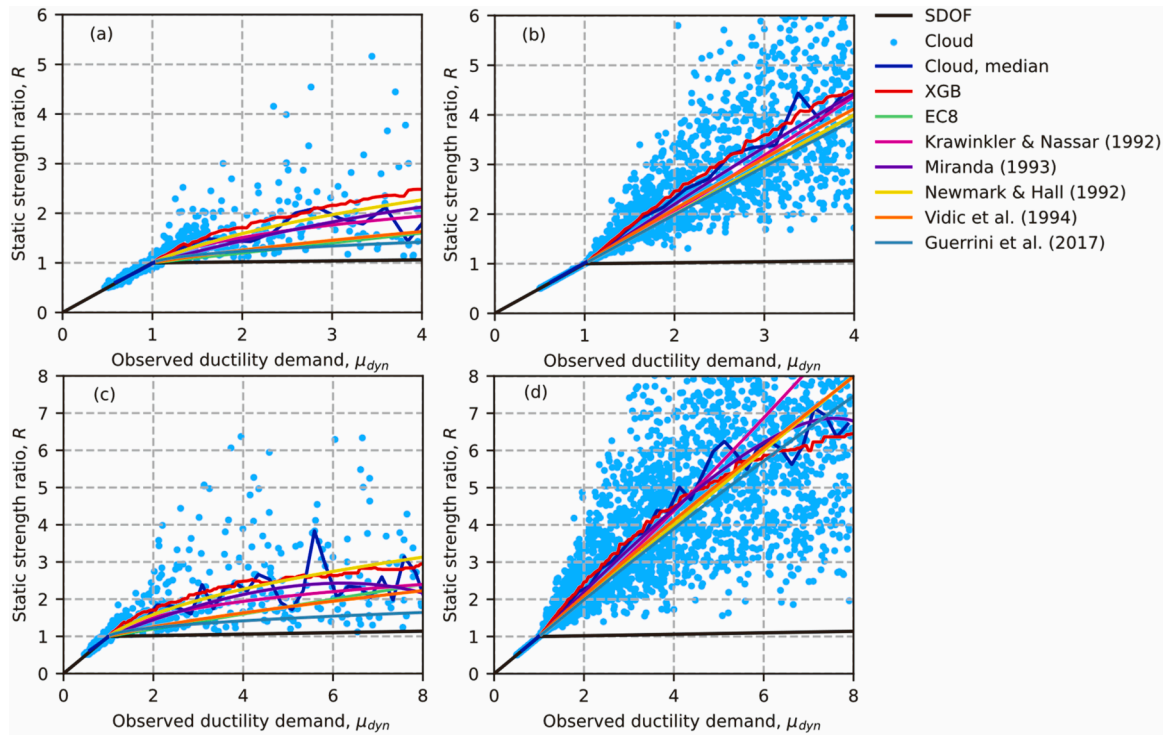
Fig. 12. Predictions of R-μ-T relationships by XGBoost and analytical methods, where $\xi= 5\%$, $a_h= 2\%$ and (a) $T = 0.1$ s, $\mu_c= 4$; (b) $T = 1.0$ s, $\mu_c= 4$; (c) $T = 0.1$ s, $\mu_c= 8$; and (d) $T = 1.0$ s, $\mu_c= 8$.

median of the cloud data and its prediction, in addition to the discrete amount of data points available. An apparent advantage of such model comes in ease of adaptability and transfer learning when a new dataset with different characteristics is introduced for training, but it should be noted that for widespread practical application of these relationships, further data should be added to smoothen such curves or possible analytical functions could be explored. Nevertheless, the focus of this study remains the utilisation of alternative methods that are easily improved and scalable once new data is available. Performing the same operation with analytical approaches may require significant changes in functional form and become somewhat difficult, although still possible in theory.

To better visualise the predictions of different analytical methods jointly with the XGBoost model's predictions, the median values were plotted in Fig. 13 and Fig. 14, while varying μ_c and T, respectively. Combined with Fig. 12, at different levels of T and μ the median of the cloud and the XGBoost predictions are in proximity to a different analytical model. Interestingly, Newmark and Hall [21] used very few

ground motions to develop the analytical model and it matches quite well with the cloud and XGBoost model's median predictions. This is especially the case for the $T = 1.0$ s cases shown in Fig. 12 although it tends to lose its accuracy with increased ductility demand, which is due to their consideration of infinite ductility capacity. In contrast, at high T and high μ_c the cloud median and XGBoost median predictions start saturating, which is similar to the predictions by Miranda [23] and Guerrini et al. [35]. At shorter periods, it is noted that many models tend to underpredict the value of R with increasing ductility. In practice and as part of a non-linear static assessment procedure, this would translate to an overestimation of the ductility demand, which is contrary to Guerrini et al. [35], who are seen to be one of the biggest under predictors of R in Fig. 12; however, it is recalled that that study was specifically referring to masonry structures, whose hysteretic behaviour differs, hence these are mere observations relative to the data presented and not necessarily conclusions. The models by Vidic et al. [28] and EC8, however, illustrate the same discrepancy with respect to the data but in this case, these models would have been expected to perform
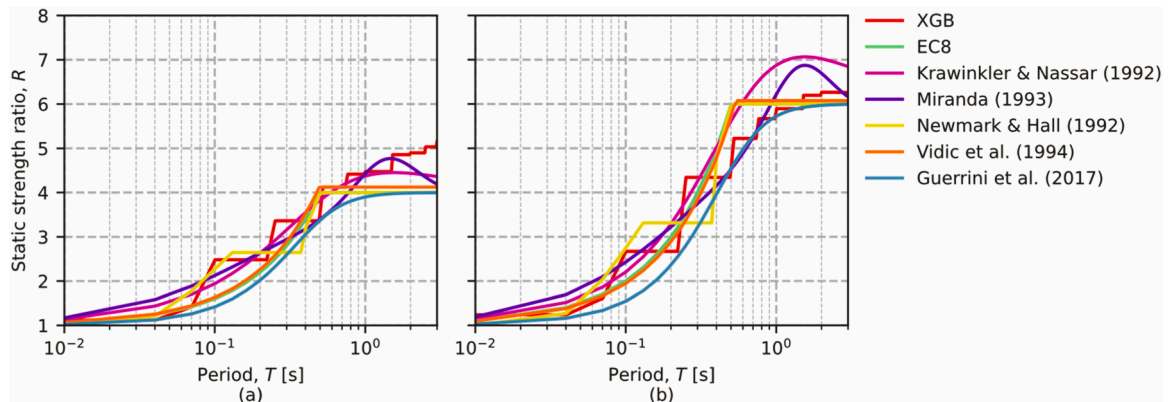

Fig. 13. Predictions of R vs T relationships by XGB and analytical methods, where $\xi= 5\%$, $a_h= 2\%$ and (a) $\mu_c= 4$; and (b) $\mu_c= 6$.
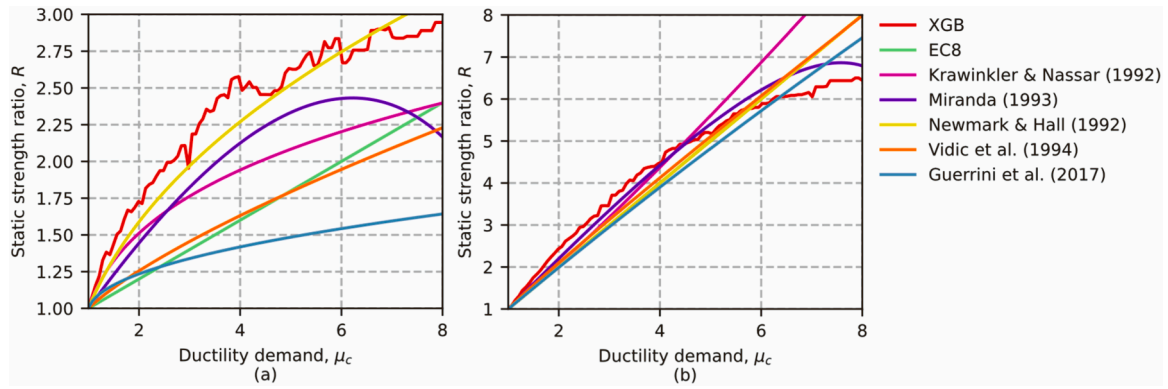
Fig. 14. Predictions of *R* vs μ relationships by XGB and analytical methods, where ξ= 5%, a_h= 2% and (a) *T* = 0.1 s; and (b) *T* = 1.0 s.

better. Fig. 13 further illustrates the comments made previously that at shorter periods, many existing models tend to underpredict the strength ratio. Fig. 14(a) also clearly illustrates that this is the case across all ductility ranges. In general, the existing methods are not bad predictors, but they may possess some limitations towards the bounds for which they were calibrated. The advantage of the ML approach described here is that it allows for a more robust quantification of the dispersion, which is discussed next. They are not constrained by predefined assumptions regarding the underlying distribution or functional form, allowing ML models to capture patterns and dependencies that may be potentially difficult for conventional parametric models. Additionally, the data-driven approach allows capturing complex relationships and provides scalability with more data availability. The implementation within a Python-based tool and availability on PyPI means that further updates to the model (e.g., more hysteretic relationships, more datapoints) can be made and released to provide a more dynamic and up-to-date tool. Nevertheless, these comparisons are in terms of strength ratio *R*, but it is recalled that this study developed models for ρ (i.e., Sa_avg) which is becoming accepted as a more advanced and accurate IM for use in design and assessment.

5. Utility within performance-based earthquake engineering

While the previous sections have highlighted how ML methods can effectively quantify the seismic demands in structural systems, this section offers further remarks on how these may be implemented within a PBEE-based framework. It is important to recall that the ML models focused on predicting the strength ratio for a given ductility demand and a set of SDOF system parameters. These predictions were observed to be quite good in terms of the metrics used to quantify ML-based predictions (Fig. 9), and the comparisons with existing *R*-μ-*T* models in Section 4.3 highlight the benefit to be gained. However, it is well established that when speaking in a more formal risk context in modern PBEE, the uncertainty surrounding the response estimation is almost as important as the estimation itself. Here, the uncertainty is due to record-to-record variability and is seen via the scatter in the cloud data illustrated in Fig. 11. This uncertainty is non-negligible. This section briefly describes how the results could be utilised in a more PBEE-oriented context and provide a solid basis for the plethora of risk and loss-targeted seismic design [44,57–61] and assessment [7,62–65] methodologies currently being developed in the literature.

When referring to modern PBEE, it is generally accepted that this refers to the Pacific Earthquake Engineering Research (PEER) Centre PBEE methodology [7]. This focuses on probabilistically quantifying the mean annual frequency of exceedance (MAFE) of a limit state, λ_LS, by integrating the probability of exceeding a limit state for a chosen IM, *P* [*LS*|*IM*= *s*], with the site hazard curve, *H*(*s*). If the typical lognormal distribution with median, η, and dispersion β, is assumed to represent the limit state fragility function, the MAFE can be computed as follows [66]:

$$\lambda_{LS} = \int_0^{+\infty} P[LS|IM = s]|dH(s)| = \int_0^{+\infty} \Phi\left[\frac{\ln\left(\frac{s}{\eta}\right)}{\beta}\right]|dH(s)| \quad (15)$$

where *s* is the IM, and Φ is the cumulative distribution function of the standard normal distribution. For what concerns the limit state fragility function, two key items are needed: the median intensity associated with a limit state's exceedance and the dispersion. Given that the empirical models quantified and discussed in Section 4 describe the strength ratio for a given level of demand in a structural system, this can be easily related to an intensity via Eqs. (2), (7), and (8). Therefore, it is crucial to determine the full potential of the ML model established in Section 4 to create a PBEE-compliant tool in conjunction with the observed dispersion.

The first task involves looking at the median strength ratios for both non-collapse and collapse cases. While the ML model gives its best prediction, this does not necessarily correspond to a known distribution. It is through the assumption of lognormality in Eq. (15) that this becomes a requirement. The assumption of lognormality stems from how strength ratios are related to ground motion shaking intensity via Eqs. (2), (7), and (8), which are described by lognormal distributions when used as fragility functions Fig. 15, therefore, plots a comparison of the ML predictions and the median values of the observed data. Given all SDOF scenarios, the factors between the η_NC from the cloud analysis and XGB predictions lie around 1.0 for ρ_2, which is encouraging. The comparison of *R* however, indicates that there are several cases where XGBoost tends to overpredict. Overall, Fig. 15 indicates good quality of predictions further highlighting the applicability of ML approaches for engineering purposes.

In addition to the median intensity required to exceed a given limit state, Eq. (15) requires the level of uncertainty, or dispersion, associated with this. To quantify this, the same data was assumed to be lognormally distributed and the dispersion was computed as the standard deviation of the logarithmic transform of the strength ratio data via:

$$\beta = \sqrt{\frac{\sum (x_i - \hat{x})^2}{N}} \quad (16)$$

where x_i is the logarithm of the ρ_2, ρ_3 or *R*, $\hat{x}$ is the corresponding mean, and *N* is the number of data points in the distribution. Fig. 16 presents the variation of β with respect to μ_dyn for the non-collapsing cases. As anticipated, the dispersions are significantly lower for *R* values at a low μ level as it would correspond to the elastic range of performance. The picture then changes at increasing μ levels, where ρ_2 has a consistently lower dispersion with respect to *R*, signifying the greater efficiency of Sa_avg compared to $Sa(T_1)$, which has been observed in several past studies [32]. Additionally, some SDOF scenarios exhibit significant
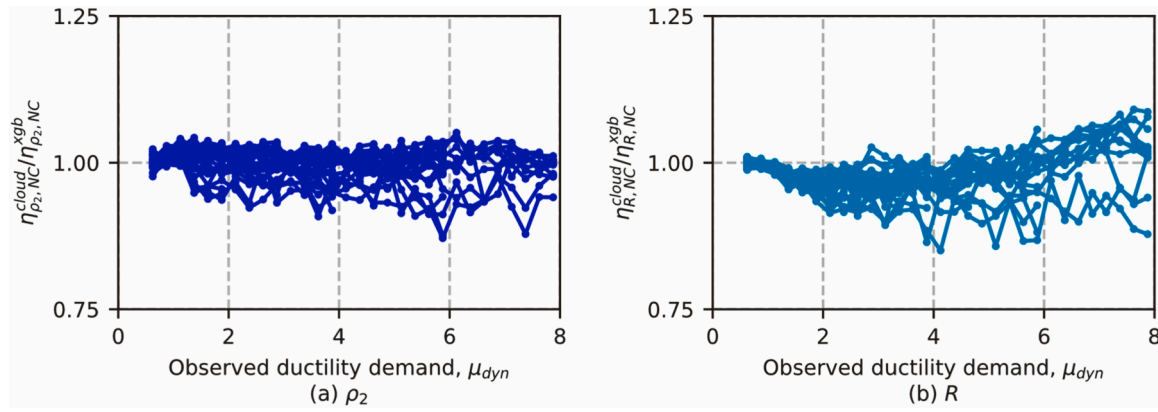
Fig. 15. Comparison of XGBoost predictions with respect to cloud analysis: (a) ρ_2 and (b) R.
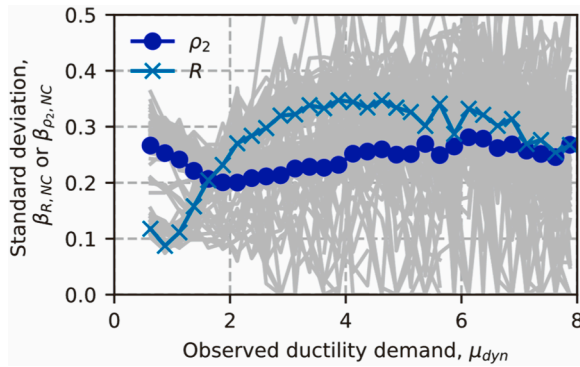

Fig. 16. Dispersions in the intensities for non-collapsing SDOF systems, $\beta_{R,NC}$ and $\beta\rho_{2,NC}$.

drops in β, which is related to insufficient data for the computation as the SDOFs have already been marked as collapsed.

Pertaining to collapsing cases, R and ρ_3 were used to analyse dispersions. Importantly, no further predictions were made using the ML approaches, as the goal of the application was to provide the η_C from ML and the β_C was computed using the observed data corresponding to each bin of μ_{dyn}. Fig. 17 presents the dispersions using both R and ρ_3 metrics for different values of μ_c and T. These observed trends β_C values may be utilised in engineering applications in conjunction with η_C values predicted using the ML model described previously. It should be noted that these dispersion values are also returned via the Python-based tool discussed next.
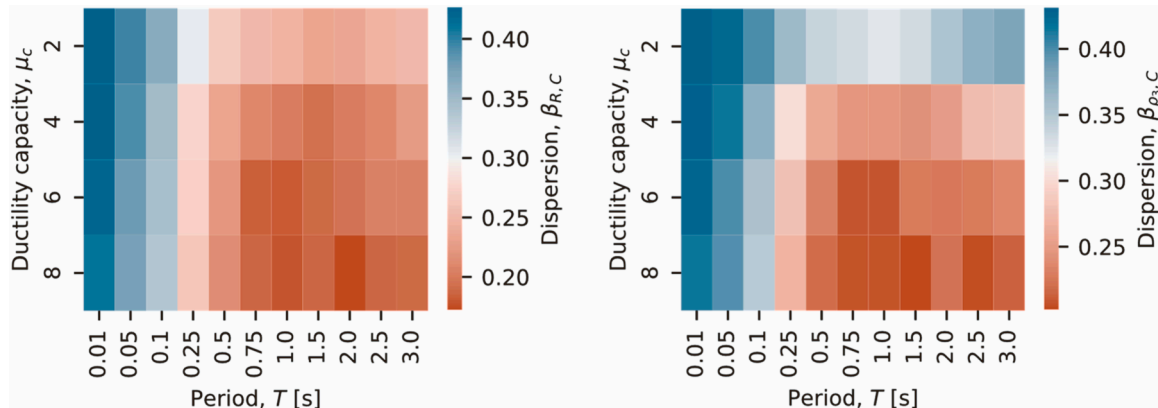
5.1. Python-based tool

A tool utilizing the XGBoost model for carrying out collapse and non-collapse predictions for R, ρ_2 and ρ_3 was implemented in a Python library [49] and is available on PyPI (https://pypi.org/project/xgb-rhomut/) and for simple pip installation. Fig. 18 demonstrates the programming structure of the tool. The library requires two input arguments at its initialisation: IM type and whether collapse or non-collapse prediction is desired. The module has a single method to make predictions based on provided independent variable values. Depending on the choice of IM and scenario, it will derive β and make predictions for R or ρ. Additionally the library contains tools to derive R-μ-T relationships using different methods shown in Section 4.3. This tool was implemented via a command-line interface (CLI) instead of a graphical user interface to allow future development and integration into different software and procedures to conduct a seismic risk assessment.

6. Summary and conclusions

This paper has presented an approach with which machine learning (ML) methods can be applied within earthquake engineering. This was via the development of simplified tools that are the building blocks of many seismic design and assessment methods. These are generally referred to as R-μ-T relationships since they relate the dynamic strength ratio R of a single-degree-of-freedom (SDOF) with a known period T to the ductility demand of a non-linearly responding system. Over the years, multiple studies aimed at developing such functions have been conducted; however, they were typically constrained by available computation power and too few ground motion recordings, which resulted in limited datasets being used. Another issue of such models is the use of strength ratio R, which assumes spectral acceleration at the
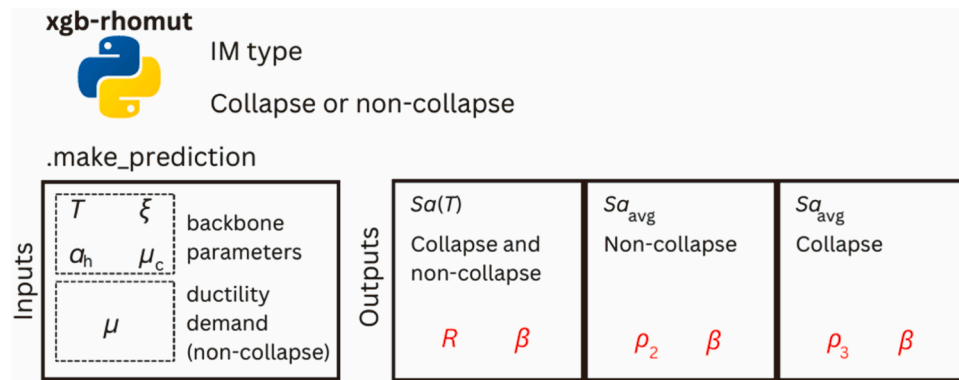

Fig. 17. Dispersions, β_C, for each SDOF scenario: (left) R; and (right) ρ_3.

Fig. 18. Programming structure of the XGBoost-based tool for *R*-μ-*T* and *ρ*-μ-*T* relationship predictions.

fundamental period, $Sa(T_1)$, as the intensity measure (IM). Significant research has been conducted in past years aiming to develop a more efficient and sufficient IM for non-linear response and collapse capacity quantification. The conclusion was that $Sa(T_1)$ is not the best predictor, and a more robust average spectral acceleration, Sa_{avg}, could be used instead, implying that *R*-μ-*T* relationships need updating. Furthermore, the models were generally developed using linear regression methods on a single training dataset, and the functions were not validated on unseen data.

The goal of this study was to traverse the limitations of past studies outlined above with the use of state-of-the-art ML techniques and use training and testing datasets generated through non-linear dynamic analysis and a large suite of ground motions recorded on different soil conditions. From the analysis, fitting and subsequent discussions carried out here, the following remarks can be made:

- Several ML methods were fitted to the training dataset and low testing errors in terms of mean absolute error (MAE) as well as high coefficient of determination (R^2). This indicated the quality of the ML models in capturing the *R*-μ-*T* and *ρ*-μ-*T* relationships and indicated no bias (i.e., no underfitting);
- Additionally, the small difference between the errors on the training and testing datasets highlights the low variance (i.e., no overfitting) of the models. Decision Trees and extreme gradient boosting (XGBoost) performed the best overall, resulting in minor errors. Therefore, the models created using the ML approaches employed here may be used to provide good predictions on the strength ratios required as part of practice-based seismic design and assessment methods;
- The SDOF systems were created using both hardening and degrading branches to predict non-linear and collapse capacity. To that end, the collapsing cloud data points from non-linear dynamic analysis were used to generate ML models for median intensity prediction resulting in the SDOF collapsing. The median predictions of the models matched well with the median response of the SDOFs, highlighting the applicability of the models;
- The predictions of the ML methods were compared to the cloud data obtained from extensive non-linear dynamic analysis as well as other proposals available in the literature. The XGBoost model predictions generally fell within similar bounds compared to existing analytical models for short and medium periods, as well as low and high ductility systems. The main difference of the XGBoost model comes in non-smooth predictions, which is a direct consequence of using a non-parametric approach with no analytical formulation and based on minimising the error between the data points and its predictions. This would be expected to smoothen out upon the addition of more data. Finally, such a model's key advantage is its adaptability and transfer possibilities when newer datasets with different characteristics are introduced. The re-training and re-validation of such

models would come with no additional cost, bar the generation of a new dataset. This is also a key advantage or the availability of these models via Python-based tools available on the widely-used PyPI platform;

- The ML methods utilised herein are anticipated to be effectively used for quantifying seismic demands in structural systems and implemented within performance-based earthquake engineering-compliant frameworks. In conjunction with the median intensity predicted, the uncertainty (i.e., dispersion) associated with it was estimated. It was observed that the dispersion associated with *ρ* values tend to be relatively low and consistent at any ductility demand level. In contrast, the dispersions associated with *R* tend to increase significantly with the increase in ductility demand levels, indicating a poorer efficiency in characterising the non-linear response of structures, which aligns with past observations in the literature.

CRediT authorship contribution statement

The manuscript was elaborated considering the following authors' contributions. **Davit Shahnazaryan:** Conceptualization, Methodology, Resources, Software, Validation, Writing – original draft. **Gerard J. O'Reilly:** Conceptualization, Validation, Writing – review & editing, Supervision.

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 reported in this paper.

Data availability

It is fully described and available in the manuscript.

References

[1] C.E.N. Eurocode 8: Design of Structures for Earthquake Resistance - Part 1: General Rules, Seismic Actions and Rules for Buildings (EN 1998–1:2004). Brussels, Belgium: 2004.

[2] ASCE 7–16. Minimum Design Loads for Buildings and Other Structures. Reston, VA, USA: 2016.

[3] NZS 1170.5:2004. *Structural design actions part 5: Earthquake actions* . Wellington, New Zealand: 2004.

[4] N.T.C. Norme Tecniche Per Le Costruzioni. Rome, Italy: 2018.

[5] FEMA. *FEMA P58*-3. Seismic Performance Assessment of Buildings *Volume 3* - Performance Assessment Calculation Tool (PACT). Washington, D.C.: 2012.

[6] S.E.A.O.C. Vision 2000: Performance-based seismic engineering of buildings 1995.

[7] Cornell CA, Krawinkler H. Progress and challenges in seismic performance assessment. PEER Cent N 2000;3(2):1–2.

[8] ATC-40. *Seismic evaluation and retrofit of concrete buildings.* Redwood City, CA: Applied Technology Council; 1997.

[9] Bertero VV. Strength and deformation capacities of buildings under extreme environments. Eng Struct Mech 1977:211–5.

[10] Vamvatsikos D, Cornell CA. Incremental dynamic analysis. Earthq Eng Struct Dyn 2002;31(3):491–514. https://doi.org/10.1002/eqe.141.

[11] Shibata A, Sozen MA. Substitute-structure method for seismic design in R/C. J Struct Div 1976;102(1):1–18. https://doi.org/10.1061/JSDEAG.0004250.

[12] Blandon CA, Priestley MJN. Equivalent viscous damping equations for direct displacement based design. J Earthq Eng 2005;9(sup2):257–78. https://doi.org/10.1142/S1363246905002390.

[13] Freeman S.A. Prediction of response of concrete buildings to severe earthquake motion. *Douglas McHenry International Symposium on Concrete and Concrete Structures, ACI Special Publication 55* 1978(American Concrete Institute, Detroit, MI): 589–605.

[14] NZSEE. *The seismic assessment of existing buildings, Technical guidelines for engineering assessments, part A: assessment objectives and principles*. Ministry of Business, Innovation and Employment. Earthquake Commission, New Zealand Society for Earthquake Engineering, Structural Engineering Society and New Zealand Geotechnical Society; 2017.

[15] FEMA 273, *NEHRP Guidelines for the Seismic Rehabiliation of Buildings, Report No. FEMA 273*. Washington, D.C.: Federal Emergency Management Agency; 1997.

[16] FEMA 274, *NEHRP Commentary on the Guidelines for the Seismic Rehabiliation of Buildings, Report No. FEMA 274*. Washington, D.C.: Federal Emergency Management Agency; 1997.

[17] *ASCE/SEI 41-17. Seismic rehabilitation and retrofit of existing buildings*. Reston, VA, USA: American Society of Civil Engineers; 2017.

[18] Fajfar P, Gašperšič P. The N2 method for the seismic damage analysis of RC buildings. Earthq Eng Struct Dyn 1996;25(1):31–46. https://doi.org/10.1002/(SICI)1096-9845(199601)25:1<31::AID-EQE534>3.0.CO;2-V.

[19] Eurocode CEN. 8: design of structures for earthquake resistance, Part 1, General Rules, Seismic Actions and Rules for Buildings. Standardization (CEN). Brussels, Belgium: European Committee for,; 2004.

[20] Veletsos A.S., Newmark N.M. Effects of inelastic behavior on the response of simple system to earthquake motions. *Proceedings of the 2nd World Conference on Earthquake Engineering*, Japan: 1960.

[21] Newmark N.M., Hall W.J. *Earthquake Spectra and Design*. Berkeley, CA: Earthquake Engineering Research Institute; 1982.

[22] Riddell R., Newmark N.M. Statistical analysis of the response of nonlinear systems subjected to earthquakes. University of Illinois, Urbana: Studies, Civil Engineering, Structural Research Series No. 468; 1979.

[23] Miranda E, Bertero VV. Evaluation of strength reduction factors for earthquake-resistant design. Earthq Spectra 1994;10(2):357–79. https://doi.org/10.1193/1.1585778.

[24] Miranda E. Inelastic displacement ratios for structures on firm sites. J Struct Eng 2000;126(10):1150–9. https://doi.org/10.1061/(ASCE)0733-9445(2000)126:10(1150).

[25] Riddell R, Hidalgo P, Cruz E. Response modification factors for earthquake resistant design of short period buildings. Earthq Spectra 1989;5(3):571–90. https://doi.org/10.1193/1.1585541.

[26] Vamvatsikos D, Cornell CA. Direct estimation of the seismic demand and capacity of oscillators with multi-linear static pushovers through IDA. Earthq Eng Struct Dyn 2006;35(9):1097–117. https://doi.org/10.1002/eqe.573.

[27] Nafeh AMB, O'Reilly GJ, Monteiro R. Simplified seismic assessment of infilled RC frame structures. Springe Neth 2020;vol. 18. https://doi.org/10.1007/s10518-019-00758-2.

[28] Vidic T, Fajfar P, Fischinger M. Consistent inelastic design spectra: strength and displacement. Earthq Eng Struct Dyn 1994;23(5):507–21. https://doi.org/10.1002/eqe.4290230504.

[29] Sousa L, Silva V, Marques M, Crowley H. On the treatment of uncertainties in the development of fragility functions for earthquake loss estimation of building portfolios. Earthq Eng Struct Dyn 2016;45(12):1955–76. https://doi.org/10.1002/eqe.2734.

[30] Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BSJ, et al. NGA-West2 database. Earthq Spectra 2014;30(3):989–1005. https://doi.org/10.1193/070913EQS197M.

[31] Vamvatsikos D., Akkar S.D., Miranda E. Strength Reduction Factors for the Dynamic Instability of Oscillators With Non-Trivial Backbones 2009(June): 22–24.

[32] Eads L, Miranda E, Lignos DG. Average spectral acceleration as an intensity measure for collapse risk assessment. Earthq Eng Struct Dyn 2015;44(12):2057–73. https://doi.org/10.1002/eqe.2575.

[33] Gentile R, Galasso C. Surrogate probabilistic seismic demand modelling of inelastic single-degree-of-freedom systems for efficient earthquake risk applications. Earthq Eng Struct Dyn 2022;51(2):492–511. https://doi.org/10.1002/eqe.3576.

[34] Kohrangi M., Bazzurro P., Vamvatsikos D., Spillatura A. Conditional spectrum-based ground motion record selection using average spectral acceleration 2017. DOI: 10.1002/eqe.

[35] Guerrini G, Graziotti F, Penna A, Magenes G. Improved evaluation of inelastic displacement demands for short-period masonry structures. Earthq Eng Struct Dyn 2017;46(9):1411–30. https://doi.org/10.1002/eqe.2862.

[36] Berman JW, Bruneau M. Cyclic testing of a buckling restrained braced frame with unconstrained gusset connections. J Struct Eng 2009;135(12):1499–510. https://doi.org/10.1061/(asce)st.1943-541x.0000078.

[37] Verderame GM, Fabbrocino G, Manfredi G. Seismic response of r.c. columns with smooth reinforcement. Part II: Cyclic tests. Eng Struct 2008;30(9):2289–300. https://doi.org/10.1016/j.engstruct.2008.01.024.

[38] Calvi G.M., Magenes G., Pampanin S. Experimental test on a three storey RC frame designed for gravity only. Proceedings of the Twelfth European Conference on Earthquake Engineering 2002; 727(January): Paper Reference 727. DOI: S0376–8716(09)00016–7 [pii] 10.1016/j.drugalcdep.2009.01.003.

[39] O'Reilly GJ, Sullivan TJ. Modeling techniques for the seismic assessment of the existing Italian RC frame structures. J Earthq Eng 2019;23(8):1262–96. https://doi.org/10.1080/13632469.2017.1360224.

[40] Haselton C.B., Liel A.B., Lange S.T. Beam-Column Element Model Calibrated for Predicting Flexural Response Leading to Global Collapse of RC Frame Buildings. *Peer 2007* 2008.

[41] Priestley M.J.N., Calvi G.M., Kowalsky M.J. Displacement based seismic design of structures 2007.

[42] ASCE 7–16. *Minimum design loads for buildings and other structures*. Reston, V.A., USA: 2014. DOI: 10.1126/science.69.1782.217-a.

[43] Zhu M, McKenna F, Scott MH. OpenSeesPy: Python library for the OpenSees finite element framework. SoftwareX 2018;7:6–11. https://doi.org/10.1016/j.softx.2017.10.009.

[44] Shahnazaryan D, O'Reilly GJ, Monteiro R. On the seismic loss estimation of integrated performance-based designed buildings. Earthq Eng Struct Dyn 2022; (January):1–25. https://doi.org/10.1002/eqe.3638.

[45] Dávalos H, Miranda E. A ground motion prediction model for average spectral acceleration. J Earthq Eng 2021;25(2):319–42. https://doi.org/10.1080/13632469.2018.1518278.

[46] FEMA. NEHRP Recommended Seismic Provisions. *Fema P-750* 2009: 406.

[47] Boore D.M. , Atkinson G.M. Boore–atkinson N.G.A. ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. Berkeley, CA: 2007.

[48] Nafeh AMB, O'Reilly GJ. Unbiased simplified seismic fragility estimation of non-ductile infilled RC structures. Soil Dyn Earthq Eng 2022;157(March):107253. https://doi.org/10.1016/j.soildyn.2022.107253.

[49] Shahnazaryan D, O'Reilly GJ. XGB-rhomut 2023. https://doi.org/10.5281/zenodo.7634615.

[50] Vamvatsikos D, Kazantzi AK, Aschheim MA. Performance-based seismic design: avant-garde and code-compatible approaches. ASCE-ASME J Risk Uncertain Eng Syst, Part A: Civ Eng 2016;2(2). https://doi.org/10.1061/AJRUA6.0000853.

[51] Chen T., Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 2016; 13–17-Augu: 785–794. DOI: 10.1145/2939672.2939785.

[52] Feng DC, Wang WJ, Mangalathu S, Hu G, Wu T. Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements. Eng Struct 2021;235(December 2020):111979. https://doi.org/10.1016/j.engstruct.2021.111979.

[53] Hwang SH, Mangalathu S, Shin J, Jeon JS. Estimation of economic seismic loss of steel moment-frame buildings using a machine learning algorithm. Eng Struct 2022;254(January):113877. https://doi.org/10.1016/j.engstruct.2022.113877.

[54] Hwang SH, Mangalathu S, Shin J, Jeon JS. Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames. J Build Eng 2021;34(October 2020):101905. https://doi.org/10.1016/j.jobe.2020.101905.

[55] Kuhn M., Johnson K. *Applied Predictive Modeling*. New York, NY: Springer; 2018.

[56] Krawinkler H., Nassar A. *Seismic design based on ductility and cumulative damage demands and capacities, Nonlinear Seismic Analysis and Design of Reinforced Concrete Buildings*. New York, NY: Elsevier Applied Science; 1992.

[57] Vamvatsikos D, Aschheim MA. Performance-based seismic design via yield frequency spectra. Earthq Eng Struct Dyn 2016;45(11):1759–78. https://doi.org/10.1002/eqe.2727.

[58] Žižmond J, Dolšek M. Formulation of risk-targeted seismic action for the force-based seismic design of structures. Earthq Eng Struct Dyn 2019;48(12):1406–28. https://doi.org/10.1002/eqe.3206.

[59] Fox M, O'Reilly G. Exploring the site dependency of fragility functions in risk-targeted design. Earth Eng Struct Dyn 2023;52(13):4148–63. https://doi.org/10.1002/eqe.3783.

[60] O'Reilly GJ, Yasumoto H, Suzuki Y, Calvi GM, Nakashima M. Risk-based seismic design of base-isolated structures with single surface friction sliders. Earthq Eng Struct Dyn 2022;51(10):2378–98. https://doi.org/10.1002/eqe.3668.

[61] Franchin P, Petrini F, Mollaioli F. Improved risk-targeted performance-based seismic design of reinforced concrete frame structures. Earthq Eng Struct Dyn 2018;47(1):49–67. https://doi.org/10.1002/eqe.2936.

[62] Seismic FEMA. Performance Assessment of Buildings - methodology. Fema P-58-1 2012;1(September):278.

[63] Calvi GM, Sullivan TJ, Welch DP. A seismic performance classification framework to provide increased seismic resilience. Geotech, Geol Earthq Eng 2014. https://doi.org/10.1007/978-3-319-07118-3_11.

[64] Nafeh AMB, O'Reilly GJ. Simplified pushover-based seismic risk assessment methodology for existing infilled frame structures. Bull Earth Eng 2022;21(4): 2337–68. https://doi.org/10.1007/s10518-022-01600-y.

[65] O'Reilly GJ, Calvi GM. A seismic risk classification framework for non-structural elements. Bull Earthq Eng 2021;19(13):5471–94. https://doi.org/10.1007/s10518-021-01177-y.

[66] Cornell CA, Jalayer F, Hamburger RO, Foutch DA. Probabilistic basis for 2000 SAC federal emergency agency steel moment frame guidelines. J Struct Eng 2002;128 (April 2002):526–33. https://doi.org/10.1061/(ASCE)0733-9445(2002)128:4(526).