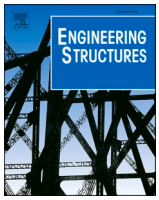
Information theory-guided machine learning to estimate seismic response of non-linear SDOF structures

Engineering Structures 336 (2025) 120448

Contents lists available at ScienceDirect

Engineering Structures

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



Information theory-guided machine learning to estimate seismic response of non-linear SDOF structures

M. De Iuliis *, E. Miceli, P. Castaldo

Department of Structural, Geotechnical and Building Engineering (DISEG), Politecnico di Torino, Turin, Italy

ARTICLE INFO

Keywords:
Machine learning algorithm
Artificial excitation
Non-linear SDOF system
Information theory
Seismic parameters
Peak relative displacement
Hysteretic energy

ABSTRACT

This study presents a parametric numerical analysis for the selection of the best seismic parameters characterising seismic records to model the dynamic response of non-linear single-degree-of-freedom (SDOF) structural systems by using Machine Learning (ML) techniques. This analysis is carried out using appropriate indices within Information Theory (IT), which allow for estimation of the amount of usable information from input data. Specifically, 400 artificial seismic excitations were generated, and, for each one, 23 seismic parameters were evaluated. Subsequently, step-by-step numerical analyses were conducted to study the seismic responses of 1000 equivalent elastic perfectly-plastic SDOF systems with different mechanical properties. The "conditional information" index was thus evaluated for both peak relative displacement and hysteretic energy response, given the input values of specific seismic parameters. The same data were treated using supervised learning techniques with 20 ML algorithms: linear regression, decision trees, support vector machine (SVM), boosted trees, bagged trees and artificial neural networks (ANN). Each analysis considered the identical set of seismic parameters, used for the conditional information index, to verify whether a higher theoretical amount of information, obtainable from the input parameters, can lead to a more efficient ML modelling. Finally, the most effective model estimation, derived from a single ML algorithm with the best combination of the input parameters, have been compared with the results of the parametric step-by-step analyses performed for some natural ground motions. The results validate the proposals and show that a higher amount of information, gained from the input parameters, generally corresponds to a better performance estimation of the ML models. This allows for the identification of which and how many seismic parameters should be considered as the best-performing combination of the input parameters for the modelling algorithm. Furthermore, when the training phase is suitably calibrated, considering the specific site hazard and the best seismic parameters, the ML model can effectively estimate the seismic performance. This highlights considerable potentials of integrating ML techniques within the performance-based seismic design approach.

1. Introduction

The accurate evaluation of the seismic response of structural systems poses a significant challenge in earthquake engineering. Specifically, single-degree-of-freedom (SDOF) systems are commonly employed to represent the dynamic response of structures subjected to seismic actions. Conventional numerical approaches, like incremental parametric evaluations, have been widely employed to assess the response of these systems. Nonetheless, as large datasets and computational capabilities become more accessible, Machine Learning (ML) methods provide an alternative for predicting structural responses. The capacity of ML to analyze large quantities of data and identify significant patterns makes it especially appropriate for modelling seismic responses. Information Theory (IT) and ML are two closely related branches of the theory underlying data analysis. The former, born with the pioneering contribution by Shannon [1], has actually a consolidated theoretical basis and has been applied in various scientific contexts. On the contrary, the latter [2] has recently received significant impulse for the following two reasons: the vast amount of data available at low costs, related to the expansion of the web network, and the increased computational capacity of devices due to technological innovation.

IT can be defined as the mathematical treatment of the concepts, parameters and rules governing the transmission of information through communication systems [3]. A meaning of the term “information” is

* Corresponding author.
E-mail addresses: massimiliano.deiuliis@polito.it (M. De Iuliis), elena.miceli@polito.it (E. Miceli), paolo.castaldo@polito.it (P. Castaldo).

https://doi.org/10.1016/j.engstruct.2025.120448
Received 29 April 2024; Received in revised form 5 April 2025; Accepted 23 April 2025

referred to as what removes uncertainty regarding the realization of a random variable [4]. From this perspective, a seismic signal, characterized by synthetic seismic parameters variable over time, represents a vehicle of information for different random variables. The main contributions related to the use of IT within earthquake engineering depend on the variable adopted to evaluate the informational contribution. This variable can denote the seismic response of a structural system [5,6], characteristics of the rocks crossed by the seismic wave [7], occurrence of a seismic event in a given region and its propagation [8,9], or measurements of both seismic intensity and features characterizing a seismic time-history [10,11]. In all these cases, through the concept of entropy, IT provides a formal mathematical tool capable to compare the amount of information contained in a dataset or quantify how much knowledge of one variable can contribute to reduce uncertainty about a different correlated variable.

ML is a set of algorithms that gives computers the ability to achieve a mathematically formalised goal without being explicitly programmed, based solely on the available data and capacity to derive information from the data. ML algorithms can be divided into three main categories: supervised learning, unsupervised learning and reinforcement learning. Supervised learning category uses prior information contained in a labelled dataset to determine a model that as best as possible approximates the relationship between the input and labelled output data. Differently, unsupervised learning aims to infer the natural structure present in a no labelled dataset, while reinforcement learning is based on an autonomous agent's searching for the best strategy to maximise its gain through interaction with an environmental system. The use of these procedures within civil engineering research dates to the early 2000s [12] with the idea of integrating soft-computing techniques, such as artificial neural networks, genetic algorithms, fuzzy logic and wavelet analysis, into the solution of numerical engineering problems. An extensive analysis on evolutionary computation in the context of structural design, such as topological optimal design, was conducted by Kicinger et al. [13], while various mathematical modelling schemes, aimed at the optimal design of steel frames, were examined by Saka and Geem [14]. Applications to earthquake engineering have considerably increased in recent years due to the possibility of high-performance computing machines, at ever lower costs, capable of handling large databases in an acceptable timeframe. Xie et al. [15] provided an exhaustive overview of the numerous contributions distinguishing four sectors: seismic hazard analysis, system identification and damage detection, seismic fragility assessment, structural control for earthquake mitigation. ML techniques therefore offer a new paradigm in the modelling and prediction of the dynamic response of a structural system, based exclusively on the processing of databases. Several studies confirm the applicability and effectiveness of ML methods for dynamic response analysis of different structural types [16–19], both in the linear and non-linear field [20–23] as well as in the case of soil-structure interaction [24].

With reference to the prediction of SDOF systems non-linear response, recent contributions analyzed different structural types [25–27] and seismic protection systems [28]. Abdellatif et al. [29] showed how an Artificial Neural Network (ANN) model emerges as a fast tool for precisely predicting the dynamic response of SDOF systems, particularly those characterized by vibration periods exceeding 0.5 s. Hammal et al. [30] presented a model to predict inelastic response spectra of 5 % damped SDOF systems using an ANN with a Back-Propagation (BP) learning algorithm. Furthermore, Dwairi & Tarawneh [31] investigated the inelastic displacement demand of structures by means of an ANN and Bayesian regularization algorithm, considering four types of hysteretic models. Besides neural networks, other ML algorithms have also been used to predict structural response. In detail, Yinfeng et al. [32] demonstrated that the two-stage method based on support vector machines (SVM) provides excellent performance in generalization and accuracy, so, it can be a powerful tool for non-linear system simulation and prediction. Meanwhile, Gharehbaghi et al. [33] showed that an explicit multi-gene genetic programming-based mathematical model is both practical and effective in quantifying the potential seismic damage of structures. Moreover, decision tree methods have been used by Demir et al. [34], namely Random Forest, eXtreme Gradient Boosting and Stochastic Gradient Boosting, to predict the seismic response of structures. Additionally, Gaussian process regression was used by Gentile & Galasso [35] to predict the probabilistic seismic demand modeling of SDOF systems.

Regarding the use of ML techniques aimed to effectively characterise a seismic signal, it is possible to select appropriate data from the available databases of recorded events [36,37] for modelling future events probabilistically and studying their impact on the built environment. Within this topic, there are many studies dealing with neural networks [38,39], genetic programming techniques [40], modelling through tree structures [41] and mixed methods [42].

As for the modelling of seismic response of a structural system, the characterisation of seismic signals represents a crucial point in vulnerability assessment, particularly when the dynamic response is strongly non-linear [43,44]. The mathematical model of the structural dynamic response is generally obtained, through supervised or unsupervised ML techniques, without a preliminary study aimed at understanding which parameters as best as possible characterise the properties of the seismic excitation and, therefore, allow for a more reliable modelling of the seismic response. Similar sensitivity studies have been conducted in [45], however, within a traditional numerical approach. In this context, the use of IT to analyse the information contained in a seismic signal can be very useful to simulate the dynamic response of a structure by using ML techniques with very important applicative advantages: it allows for the definition of more reliable structural models, it limits the amount of information to be stored and processed for a single seismic signal and, consequently, the computation time, making it possible to perform seismic response analysis within classical design approach without high performance computer machines. An explicit and synergic approach between IT and ML, where IT is used to analyse the usable information in the characteristics of a seismic input to model the seismic response through ML, can lead to relevant results, according to [46,47].

This research focuses on creating predictive ML models guided by IT to estimate the non-linear seismic response of SDOF systems and examines the choice of seismic parameters to improve the model precision. It is described a parametric analysis for the selection of the best seismic parameters characterising seismic records to model the dynamic response of non-linear structural systems by using ML techniques. This analysis is carried out by using appropriate indices defined within the IT, which allow to estimate the amount of information from the input data. Specifically, 400 artificial seismic excitations were generated, and, for each one, 23 seismic parameters were evaluated. Subsequently, step-by-step numerical analyses were conducted to study the seismic responses of 1000 equivalent elastic perfectly-plastic SDOF systems with different mechanical properties. The obtained data were initially analysed using IT elements. Specifically, the "conditional information" index was evaluated for each seismic response, expressed in terms of both peak relative displacement and hysteretic energy, based on the input values of specific seismic parameters. The same data were treated using supervised learning techniques with 20 ML algorithms: linear regression, decision trees, SVM, boosted, bagged trees and ANN. Each analysis considered the identical set of input parameters, analysed by using the conditional information index, to verify whether a higher theoretical amount of information, obtainable from the input parameters, corresponds to a more efficient ML modelling. A set of High Predictive Seismic Parameters (HPSP) are thus identified as those with the best estimation levels in accordance with two significant metrics, showing how they correspond to the parameters with the highest informational content. Finally, with the aim to validate the obtained results, the most effective model predictions, deriving from a single ML algorithm together with the best-performing combination of the seismic parameters, have been compared with the results of the parametric step-by-step

analyses performed for a new group of natural ground motions.

2. Generation of synthetic seismic records

The ML techniques are based on algorithms that analyse data to numerically infer the relationship between input and output variables: the higher the amount of available data, the more information can be utilized by the algorithm to effectively reproduce a reliable structural response. This consideration naturally highlights the need to have enough data defining seismic demand by selecting seismic input signals with respect to the seismic hazard of the site where the structure is located. Therefore, according to performance-based design and regulations [48], the need to select a set of seismic events as realistic as possible becomes crucial. ML algorithms create numerical models by analysing available data in "training phase" based on the comparison between the simulated response, obtained from the numerical algorithm, and actual response of the structural system to several accelerograms. Given any specific site conditions, a very large number of natural seismic records for training a ML model is not available, therefore, the alternative is to synthetically generate such accelerograms. Adopting the "specific barrier model" by Aki [49,50], variables such as magnitude, Joyner-Boore distance [51], soil type and tectonic regime (i.e., intra-plate, inter-plate, active extensional tectonic regimes) are considered. Such model has been implemented through the Seismoartif software [52] and calibrated by means of a large database of response spectral amplitudes related to seismic events in intraplate areas, interplate regions and regions of tectonic extension [53]. Within this approach, artificial accelerograms are defined starting from a single synthetic signal, compatible with the target spectrum, and adapting its frequency content using the Fourier transform method. The correction of the random process is carried out at each iteration using Eq.(1) [54]:

$$\Phi_{i+1}(f) = \Phi_i(f) \bullet \frac{S_T(f)}{S_i(f)} \tag{1}$$

where $S_T(f)$ is the value of the target spectrum at frequency f, $S_i(f)$ is the value of the response spectrum at frequency f corresponding to the i-th iteration, $\Phi_{i+1}(f)$ and $\Phi_i(f)$ are the values of the accelerogram in the frequency domain at step i and $i+1$, respectively. Eq.(1) is defined in the frequency domain, where the signal correction process occurs. However, this correction requires calculating the response spectra using the signal representation in the time domain. This entails the need to continuously move between the two domains by means of direct and inverse Fourier transforms. Fig. 1 shows a schematic representation of this process.

Note that this process starts with a single synthetic acceleration obtained through a completely random Gaussian signal (i.e., white noise), which is multiplied by a modulating function [55] and, then, adapted to a fixed target spectrum. Once this first synthetic signal is generated, in the interactive correction process, there is no need of additional modulating functions as the process does not start from a random process but from the synthetic signal preliminarily obtained. The total duration ([s]) of seismic signals is taken constant and evaluated by using the magnitude (M), as follows [56]:

$$T_t = 10^{(0.31 \bullet M - 0.774)} \tag{2}$$

In the present study, the following parameters have been selected to describe site seismic hazard: far field earthquake with moment magnitude equal to 8 leading to $T_t = 50.82s$, Eurocode 8 part C ground type A design spectrum [57], importance factor equal to 1, target PGA= 0.24 g, linear site effect, $V_{S30} = 940m/s$, distance between epicentre and site randomly selected according to a uniform distribution within 15 and 25 km. In this way, 400 artificial seismic excitations have been randomly generated. In Fig. 2(a), the time-history of one generated signal is represented, whereas the elastic pseudo-acceleration spectra $S_a[g]$ for all the artificial seismic records together with the reference spectrum, corresponding to Eurocode 8 – ground type A [57], are plotted in Fig. 2(b).

The present investigation is focused on studying soft computing prediction of structural response to seismic records by evaluating 23 seismic parameters (S_p) for every single synthetic seismic signal [58]. These seismic parameters are listed in Table 1.

3. Dynamic response of non-linear SDOF systems

The assessment of seismic performance of non-linear structural systems with respect to the characteristics of seismic inputs represents one of the primary objectives within earthquake engineering. This issue requires an accurate modelling of the non-linear behavior of the system with a suitable representation of expected collapse modes and an effective probabilistic modelling of both material properties and seismic inputs. In this context, the adoption of ML models can represent a turning point in the numerical analysis of the non-linear structural response. However, the use of such algorithms should never neglect the physical understanding of the obtained results. This suggests the need to classify the seismic parameters, describing the input motions, to identify the most significant ones with respect to the seismic response.

The elastic perfectly-plastic model for SDOF system has been adopted to evaluate the seismic response in terms of both relative displace-

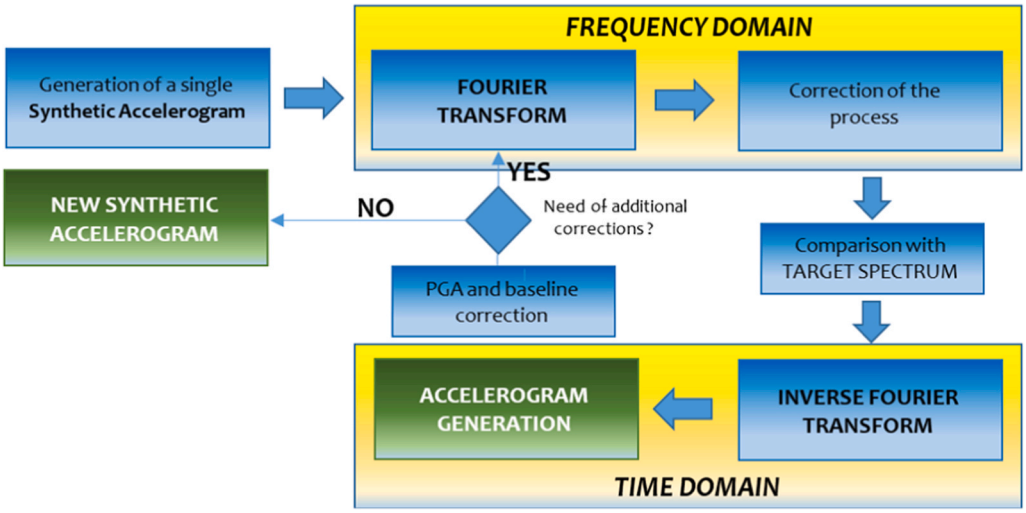


Fig. 1. Scheme of the synthetic accelerogram generation (modified from [53]).

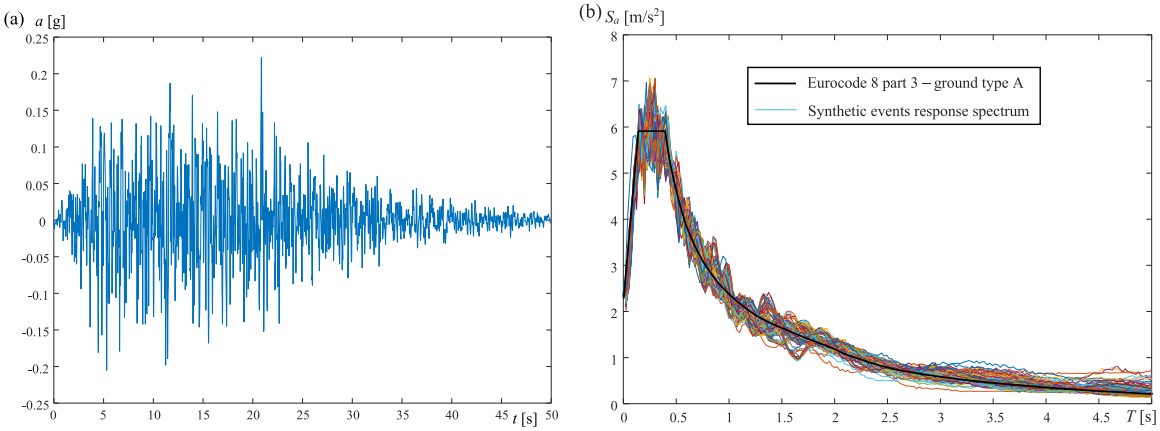


Fig. 2. (a) artificial seismic signal; (b) elastic pseudo-acceleration spectra of the 400 artificial seismic records.

ment and hysteretic energy. This structural model, introduced by Shibata and Sozen [70] and widely used in both scientific literature and design provisions [71–73], is characterized by three parameters: the elastic stiffness (k), yield strength (f_y) and relative displacement capacity (x_m) (Fig. 3). The dynamic equation governing the seismic response of a non-linear SDOF system with mass m, inherent viscous damping coefficient c and resisting force $f_S(x)$ can be written as follows:

$$\ddot{x} + 2\xi\omega_0\dot{x} + \omega_0^2 x_y \bar{f}_s(x) = -\ddot{u}_g \tag{3}$$

where the circular frequency $\omega_0 = \sqrt{\frac{k}{m}}$, damping factor $\xi = \frac{c}{2m\omega_0}$ and $\bar{f}_s(x) = \frac{f_S(x)}{f_y}$. In the following, the value of the yield strength has been defined in each case as the product between the peak value of the earthquake-induced resisting force in the corresponding linear system f_0 and the yield strength factor R_y, as follows: $f_y = R_y \bullet f_0$. During the elastic response, $\bar{f}_s(x)$ represents the restoring force and depends on k. The corresponding elastic vibration period is denoted as T_0.

The dynamic response of a non-linear SDOF system can also be expressed through the energy balance equation, which can be written as follows [74–75]:

$$E_I(t) = E_K(t) + E_D(t) + E_S(t) + E_H(t) \tag{4}$$

where: $E_I(t) = -\int_0^x m\ddot{u}_g(t)dx$ is the input energy, $E_K(t) = \frac{m\dot{x}^2}{2}$ is the kinetic energy of the mass associated with its motion relative to the ground, $E_D(t) = \int_0^x c\dot{x}(t)dx$ is the energy dissipated by viscous damping, $E_S(t) = \frac{kx^2}{2}$ is the recoverable strain energy of the system and $E_H(t) = \int_0^x f_S(x)dx - E_S(t)$ is the hysteretic energy (e.g., dissipated by yielding).

Fig. 4 shows a dynamic response in terms of relative displacement time history, force-relative displacement and hysteretic dissipated energy for a non-linear SDOF system ($T_0 = 0.5$ s, $R_y = 3$). The yield strength is the same in both directions of the response and strongly influences the non-linear response.

A comprehensive parametric analysis, aimed at determining the seismic demand and, thus, the maximum ductility and hysteretic energy for the 400 generated seismic excitations, has been carried out. Specifically, for each seismic excitation, 1000 different elastic perfectly-plastic SDOF systems have been considered. The initial stiffness k has been modified such that T_0 varies in the range [0.05 s, 3.00 s]. Similarly, the yield strength has been varied so that R_y ranges from a minimum value of 0.05 to a maximum value of 1, corresponding to a linear dynamic response of the system. The results of the parametric analysis are represented, respectively, in Figs. 5–8 in terms of peak relative displacement, x_{peak}, peak relative velocity, $\dot{x}_{peak}$, peak absolute acceleration, $\ddot{u}_{peak}$, and hysteretic energy, E_H, for a single record together with their mean values over all the 400 records on varying elastic and inelastic mechanical parameters of the SDOF systems.

In the following, the peak relative displacement and hysteretic energy are the response variables of the non-linear SDOF systems to estimate through ML models combined with the IT suggestions.

4. Information theory and time-histories conditional entropy

In the field of earthquake engineering, application of concepts from IT can serve as an important tool to support modelling with soft computing techniques. The main idea of this theory is based on the concept of entropy [1], which applies as follows for a random variable Y:

$$H(y) = -\int_{y\epsilon Y} p(y) \bullet \log_2(p(y))dy \tag{5}$$

where $p(y)$ is the probability density function (PDF) for the variable Y. This quantity (Shannon entropy) measures the level of unpredictability of a random variable and reaches its maximum value when all possible outcomes are equally probable, corresponding to maximum uncertainty regarding the realization of the variable. As evident, the higher Shannon entropy, the higher the amount of information provided by realizations of the variable. To use Eq.(5), it is necessary to know the PDF associated to the variable Y. If enough data are available, it is possible to adopt a non-parametric estimator of $H(y)$, called "empirical entropy", whose expression derives from Eq.(5) by considering an appropriate number b of bins and substituting the PDF with the occurrence frequency, f_j:

$$\widehat{H}(y) = -\sum_{j=1}^{b} f_j \bullet \log_2\left(f_j\right) \tag{6}$$

This is a correct estimator of entropy, efficient in the case of a large number of observations which are representative of the underlying distribution of the variable. The number b of bins, in this work, is selected with the Freedman-Diaconis' Rule [76], which estimates the optimal width h of each bin through the InterQuartile Range (IQR) and sample size S, according to Eq.(7):

$$h = \frac{2 \bullet IQR}{\sqrt[3]{S}} \tag{7}$$

Fig. 9 represents the empirical entropy calculated by using the 400 synthetic accelerograms for all the considered seismic parameters (Table 1). This figure highlights the significant differences, in terms of Shannon information, between the seismic parameters for the 400 selected excitations. Particularly, parameters such as the peak acceleration, peak velocity, peak displacement, bracketed duration and A95 parameter show higher entropy levels, while parameters like the acceleration RMS and predominant period exhibit more uniform values.

The issue now is to explore how much information from the seismic

Table 1
Seismic parameters characterizing the signals adopted in numerical analyses.

S_p	Seismic parameter	Description
No. 1	Peak Acceleration	Peak acceleration of the seismic input time-history: $\ddot{u}_{g,peak}$
No. 2	Peak Velocity	Peak velocity of the seismic input time-history: $\dot{u}_{g,peak}$
No. 3	Peak Displacement	Peak displacement of the seismic input time-history: $u_{g,peak}$
No. 4	AV ratio	Peak ground acceleration-to-velocity ratio: $AV = \frac{\ddot{u}_{g,peak}}{\dot{u}_{g,peak}}$
No. 5	Acceleration RMS	Square root of the mean of acceleration squares: $a_{RMS} = \sqrt{\frac{\int_0^{T_t} \ddot{u}_g^2 dt}{T_t}}$, where T_t is the time duration of the signal
No. 6	Velocity RMS	Square root of the mean of velocity squares: $v_{RMS} = \sqrt{\frac{\int_0^{T_t} \dot{u}_g^2 dt}{T_t}}$, where T_t is the time duration of the signal
No. 7	Displacement RMS	Square root of the mean of displacement squares: $x_{RMS} = \sqrt{\frac{\int_0^{T_t} u_g^2 dt}{T_t}}$, where T_t is the time duration of the signal
No. 8	Arias Intensity	Time-integral of the square of the acceleration: $I_A = \frac{\pi}{2g}\int_0^{T_d} \ddot{u}_g^2(t)dt$, where T_d is the time duration of the signal above a threshold equal to $0.05g$ [59]
No. 9	Characteristic Intensity	$I_C = a_{RMS}^{1.5} \bullet t_s^{0.5}$, where t_S is the significant duration [60]
No. 10	Specific Energy Density	Energy of the velocity-time history: $SED = \int_0^{T_d} \ddot{u}_g^2(t)dt$, where T_d is the time duration of the signal above a threshold equal to $0.05g$ [61]
No. 11	Cumulative Absolute Velocity	Area under the accelerogram over a significant duration: $CAV = \int_0^{T_t} \left\|\ddot{u}_g^2(t)\right\| dt$, where T_t is the time duration of the signal [62]
No. 12	Acceleration Spectrum Intensity	Area under the acceleration response spectrum between periods of 0.1 s and 0.5 s: $ASI = \int_{0.1}^{0.5} PSA(T,\xi)dT$, where PSA is the pseudo spectral acceleration; T is the natural period and $\xi = 0.05$ is the damping coefficient [63,64]
No. 13	Velocity Spectrum Intensity	Area under the velocity response spectrum between periods of 0.1 s and 0.5 s: $VSI = \int_{0.1}^{0.5} PSV(T,\xi)dT$, where PSV is the pseudo spectral velocity; T is the natural period and $\xi = 0.05$ is the damping coefficient [63,64]
No. 14	Housner Intensity	$SI_H = \int_{0.1}^{2.5} S_v(T,\xi)dT$, where S_v is the pseudo-velocity response spectrum; T is the natural period and $\xi = 0.05$ is the damping coefficient [62]
No. 15	Sustained Maximum Acceleration	Third-highest acceleration absolute value largest peak in the time history record: SMA [65]
No. 16	Sustained Maximum Velocity	Third largest velocity peak in the time history record: SMV [65]
No. 17	Effective Design Acceleration	Peak acceleration that remains after filtering out seismic signals above 9 Hz: EDA [66]
No. 18	A95 parameter	Peak value of acceleration corresponding to 95 % of Arias Intensity value: $A95$ [62]
No. 19	Predominant Period	Period corresponding to the maximum spectral acceleration occurs in a 5 % damped acceleration response spectrum: T_p [67]
No. 20	Mean Period	Average frequency content: $T_m = \sum_i \left(\frac{C_i^2}{f_i}\right) / \sum_i C_i^2$, where C_i are the Fourier amplitudes, and f_i represent the discrete Fourier transform frequencies between 0.25 and 20 Hz [68]
No. 21	Uniform Duration	Total time of the earthquake motion in which the amplitudes exceed a specified threshold (0.025 g): D_u [69]
No. 22	Bracketed Duration	Elapsed time between the first and last acceleration excursions greater than a specified threshold (0.025 g): D_b [69]
No. 23	Significant Duration	Time required for the release of accumulated strain energy by rupture along the fault. It is herein defined as the duration between 5 % and 95 % of the Arias intensity: D_S [69]

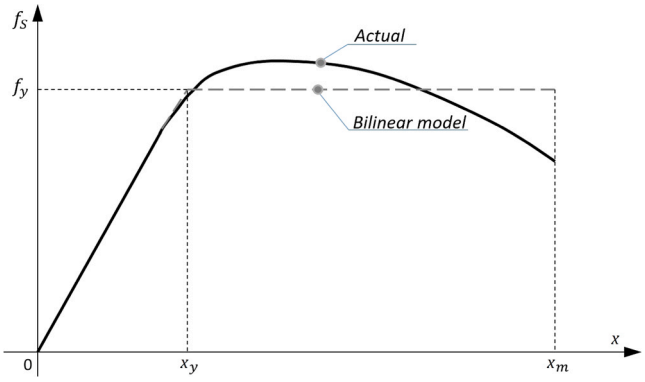


Fig. 3. Force-relative displacement curve: actual response compared with the elastic perfectly-plastic model.

parameters, which represent also input parameters, is usable in defining the seismic response. It is possible, in fact, that the variability of a seismic parameter does not lead to a similar variability in output quantities, meaning that the input information provided by one or more seismic parameters might be "compressed" by the process into a narrow range of seismic response values. To assess this, a specific index, originally introduced by [77], called the "conditional information" is examined. Let Y be the output variable and $\vec{V}$ a vector of input parameters with size n, the conditional information can be computed as follows:

$$I\left(Y|\vec{V}\right) = H(Y) + H(V_1, V_2, V_3, \ldots, V_n) - H(Y, V_1, V_2, V_3, \ldots, V_n) \tag{8}$$

where $H(Y)$ is the marginal entropy of Y, $H(V_1, V_2, V_3, \ldots, V_n)$ is the joint entropy between the components of input parameters vector $\vec{V}$ and $H(Y, V_1, V_2, V_3, \ldots, V_n)$ is the joint entropy including both the output Y and input $\vec{V}$ parameters. The joint entropy is formally defined as Eq.(6), with the only difference in the number of bins, equal to the product of the bins of each individual variable. The conditional information was initially calculated considering a single input parameter V_i by selecting the peak relative displacement or the hysteretic energy as the output variable Y:

$$I(Y|V_i) = H(Y) + H(V_i) - H(Y, V_i) \tag{9}$$

In Fig. 10, for all 23 input seismic parameters (S_p), the minimum, maximum and average values evaluated across all the responses of 1000 SDOF systems to the 400 records are shown. Specifically, the results in Fig. 10(a) represent the case $Y = x_{peak}$, while those in Fig. 10(b) refer to the case $Y = E_H$.

For the same cases, Fig. 11 plots the quantity defined in Eq. (10), which measures the ratio of the information carried by the input parameter V_i and transferred by the dynamic process to the output variable (i.e., seismic response).

$$r_{H,i} = \frac{I(Y|V_i)}{H(V_i)} \tag{10}$$

From Fig. 10, it is clear that a set of seismic parameters appear to be more "effective" in terms of contributing information within the dynamic process for the seismic response: Peak Velocity (No. 2), AV ratio

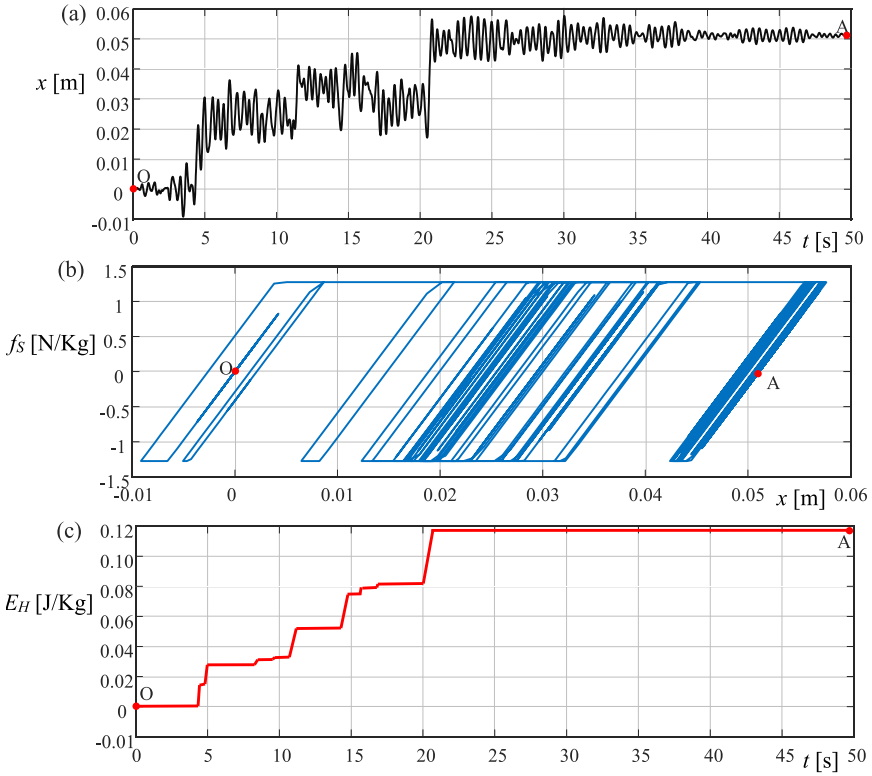


Fig. 4. Relative displacement (a), force-relative displacement (b), hysteretic energy (c) time-histories for an elastic perfectly-plastic SDOF model.

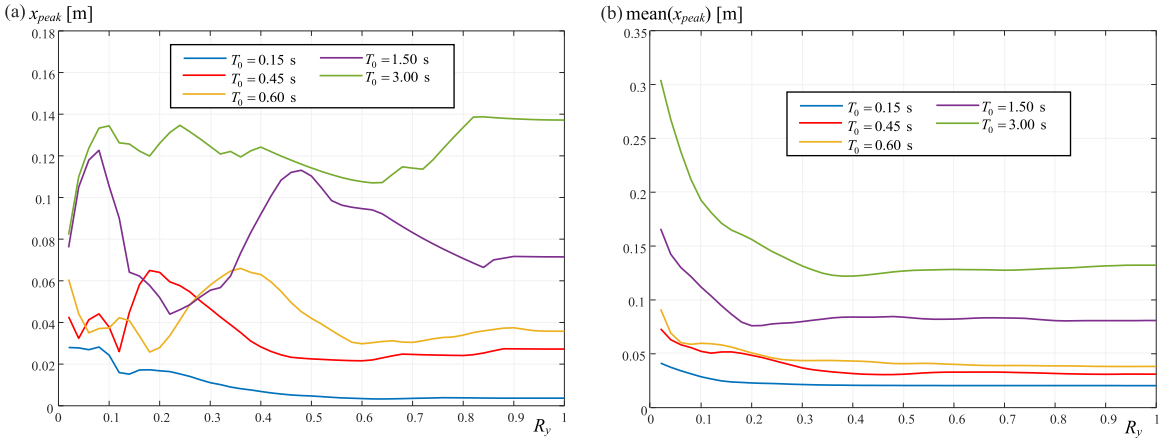


Fig. 5. Peak relative displacements of the SDOF systems: (a) single artificial record, (b) average over the 400 records.

(No. 4), Velocity RMS (No. 6), Arias Intensity (No. 8), Specific Energy Density (No. 10), Cumulative Absolute Intensity (No. 11), Sustained Maximum Velocity (No. 16) and Significant Duration (No. 23).

Furthermore, Fig. 11 shows that the amount of information from each individual seismic parameter translated into information about the output quantity ranges between 60 % and 90 % when the peak relative displacement is considered as the output variable, whereas the variability is significantly greater, ranging from 40 % to 90 %, when the output variable is the hysteretic energy.

Additionally, note that a single seismic parameter (i.e., No. 5, Acceleration RMS) exhibits significantly lower values of both empirical entropy and Shannon conditional information compared to all the other ones. These results concern the case where a single seismic parameter is considered as input.

The subsequent issue to address is how the knowledge of additional seismic parameters may increase the amount of information usefully employed by the process to get the output. In fact, the possibility of considering either 2 or 3 input parameters has been investigated. Specifically, the conditional information $I(Y|V_1, V_2)$, by considering 2 seismic parameters as input, was evaluated in average terms over all the considered 1000 non-linear SDOF systems subjected to the 400 records. For the sake of readability, only the values of the conditional information corresponding to the pairs including the best and worst single predictive parameter, respectively, Specific Energy Density - No. 10, and Acceleration RMS - No. 5, are represented in Fig. 12. The results show the minimum information is obtained when the two selected seismic parameters coincide: $I(Y|V_1, V_1) = I(Y|V_1)$, as expected. Meaningfully, the maximum usable information defining the output seismic response is obtained, in most cases, when one of the two seismic parameters has an energetic physical insight, such as Specific Energy Density (No. 10) or

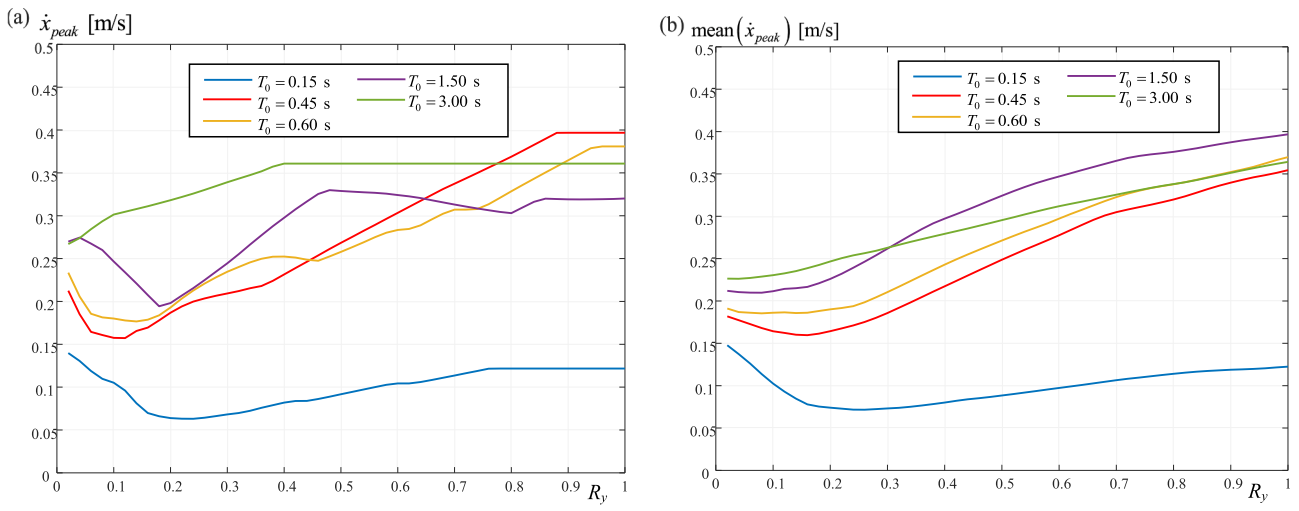


Fig. 6. Peak relative velocities of the SDOF systems: (a) single artificial record, (b) average over the 400 records.

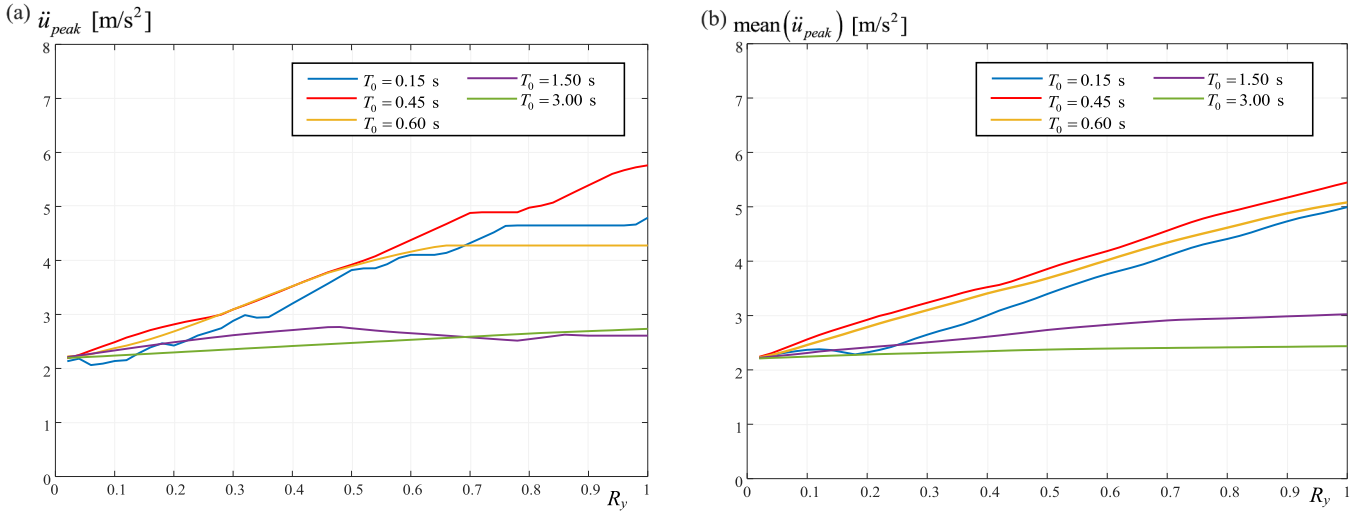


Fig. 7. Peak absolute accelerations of the SDOF systems: (a) single artificial record, (b) average over the 400 records.

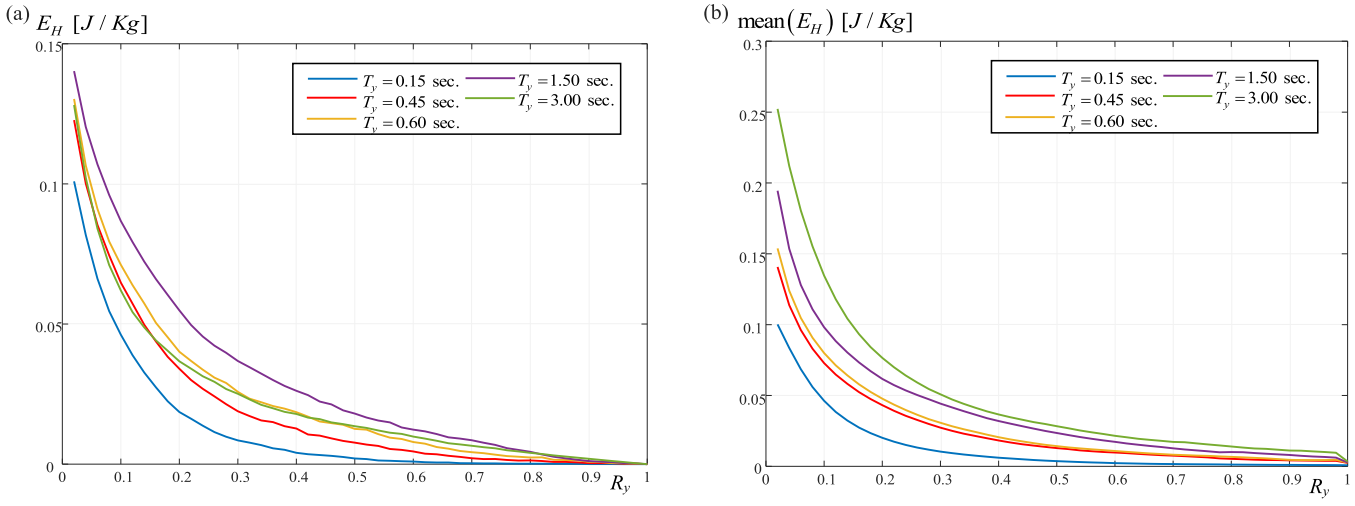


Fig. 8. Hysteretic energy of the SDOF systems: (a) single artificial record, (b) average over the 400 records.

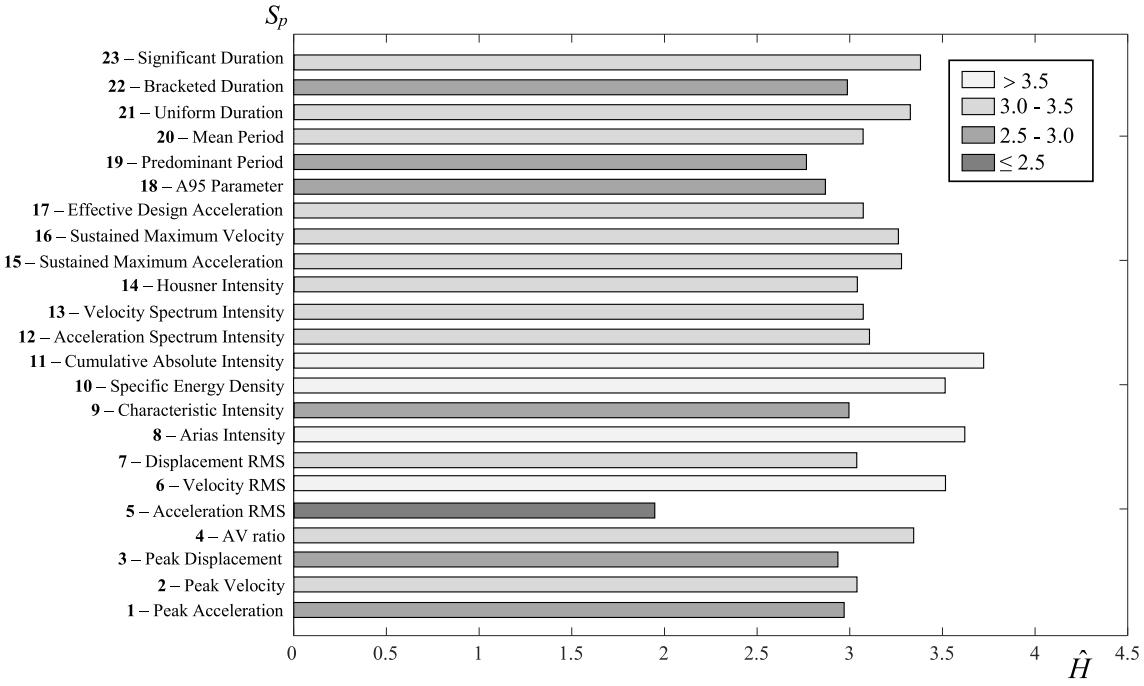


Fig. 9. Shannon empirical entropy for the considered seismic parameters.

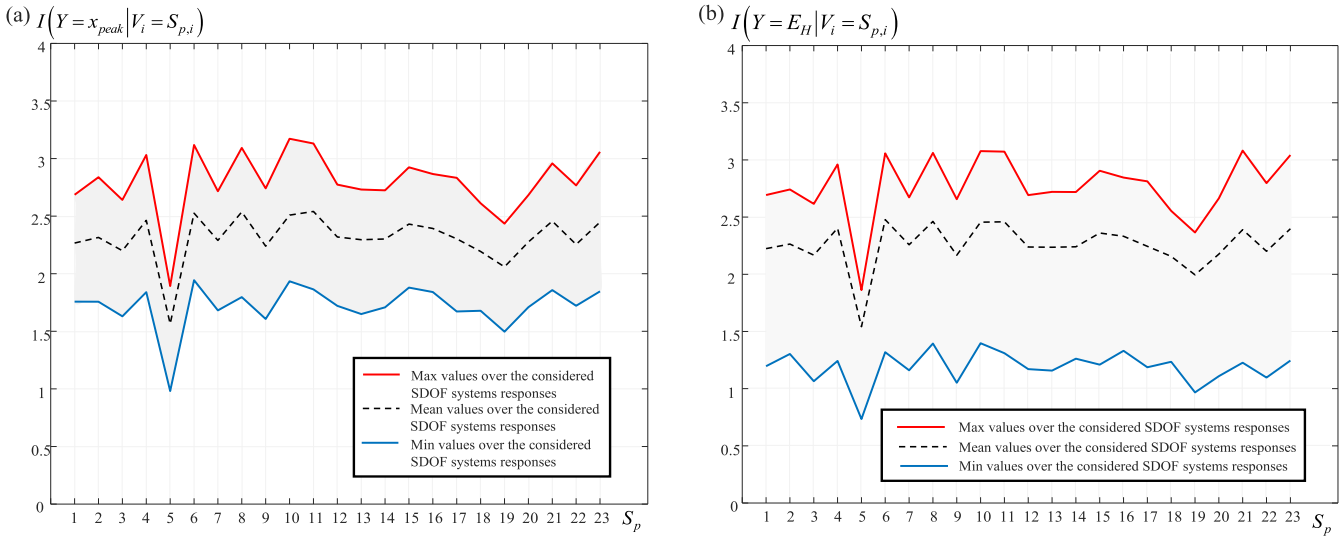


Fig. 10. Shannon conditional information for the considered seismic parameters.

Cumulative Absolute Intensity (No. 11), which together represent the most informative pair of seismic parameters (Fig. 12). Fig. 12 also shows that there are negligible differences in average Shannon's conditional information depending on the selected output. This trend was observed across all the analyses performed for the two response variables, with the difference that the average conditional information values, obtained using hysteretic energy as the output, are slightly lower than those obtained when peak relative displacement is considered.

Fig. 13 highlights the cases where the conditional information value, considering peak relative displacement as output variable, is significant even in the absence of an energy-based parameter. Specifically, this is evident for the pair Velocity RMS (No. 6) and Arias Intensity (No. 8), or for the parameters AV ratio (No. 4) and Displacement RMS (No. 7), whose information is more effectively integrated with a duration-based parameter, Significant Duration (No. 23), rather than an energy-based one.

From all the responses of the 1000 non-linear SDOF systems to the 400 records, Fig. 14 summarizes the values of average conditional information for the case of single parameters, pairs and triplets of seismic parameters. Specifically, for each seismic parameter considered, the figure shows the average conditional information corresponding to that parameter, as well as the best pair and triplet containing that parameter. The figure explicitly indicates the best pairs, while for triplets, only the configuration corresponding to the maximum conditional information is indicated. The results highlight the increase in conditional information associated with the inclusion of a second and third seismic parameter. Adding a second seismic parameter generally leads to a 10–30 % increase in conditional information relative to the output variable, while adding a third parameter a smaller, though still significant, increase is observable. Furthermore, a reduction in differences in the information

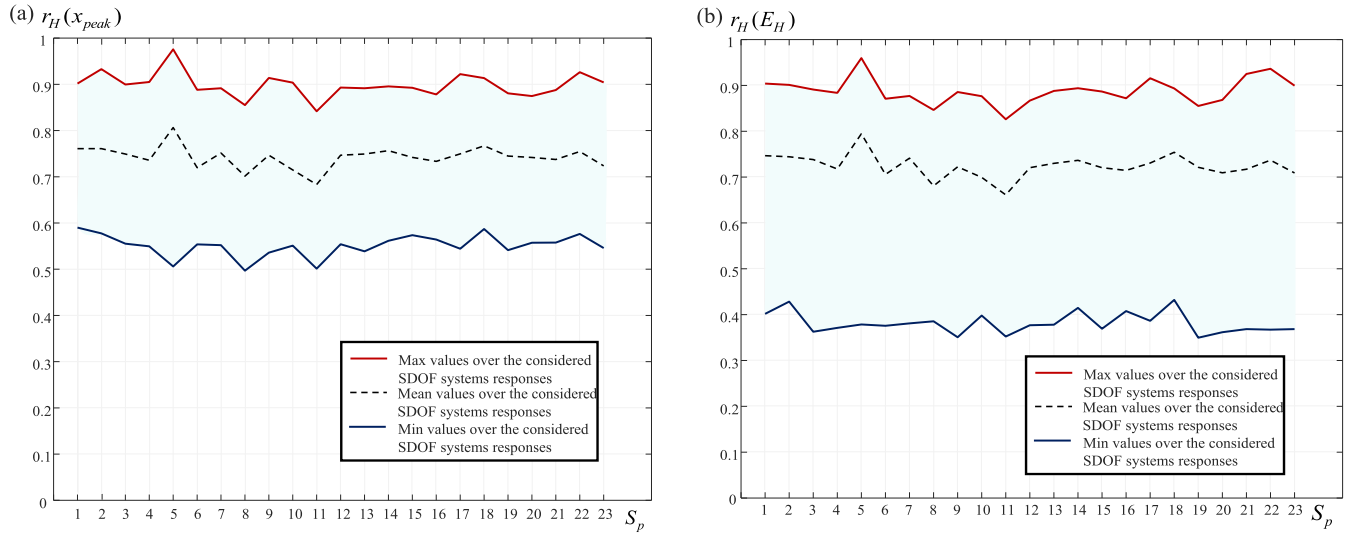


Fig. 11. Ratio of the input parameter information transferred to the output seismic response.

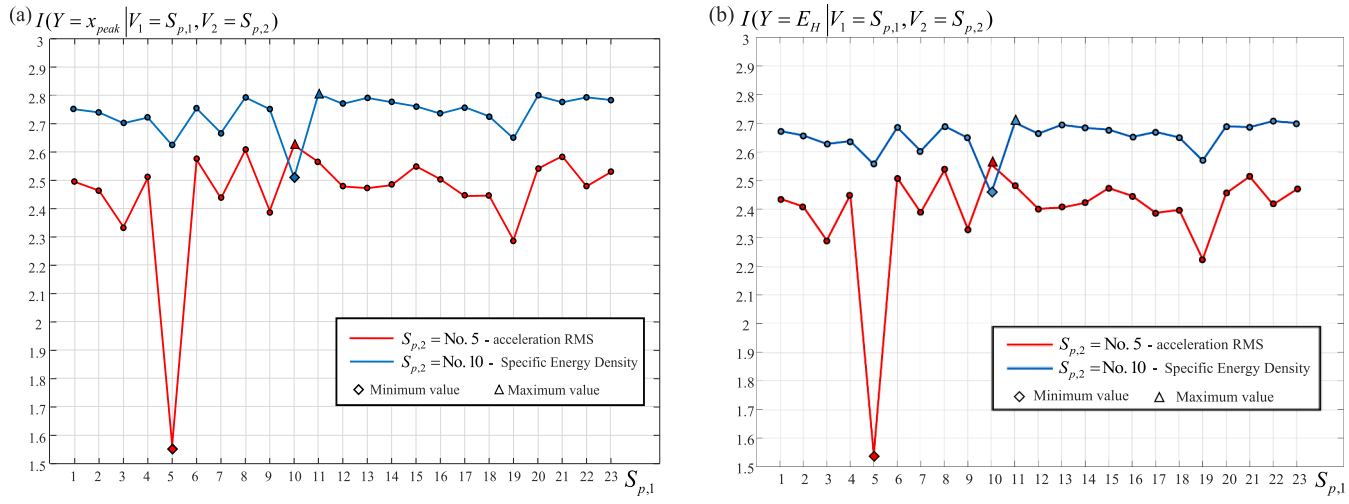


Fig. 12. Average conditional information in case of 2 input parameters - best and worst cases.

conveyed by the best-performing triplets is observed. This is due to evidence that even the presence of a parameter with limited informational content can be compensated by the inclusion of other input parameters selected to maximize conditional information. It is also noted that the results show minimum values corresponding to the seismic parameters Acceleration RMS (No. 5) and Predominant Period (No. 19). Among the best-performing triplets, the one characterized by the highest average conditional Shannon information is composed of the parameters Specific Energy Density (No. 10), Cumulative Absolute Intensity (No. 11) and Housner Intensity (No. 14), which are integral-based parameters related to the energy of the signal. The results also indicate that the seismic parameters most frequently appearing in such triplets are Specific Energy Density (No. 10) and Cumulative Absolute Intensity (No. 11). The presence of Arias Intensity (No. 8) is also recurrent, while it is observed that temporal parameters (No. 19 – No. 23) appear in best-performing triplets in combination with peak or RMS values of acceleration, velocity and displacement. These considerations hold true for both the output variables considered.

5. Supervised machine learning models

The previous results representative of the non-linear response of the considered SDOF systems have been utilized as input and output data in an extensive numerical analysis involving 20 ML algorithms. The aim is to study the relationship between the mechanical properties characterizing SDOF systems, the characteristics of the seismic excitations and the seismic responses in terms of both peak relative displacement and hysteretic energy. Specifically, the two properties characterizing the system, namely k and R_y, together with the 23 seismic parameters, representing the seismic signal, have been provided as input data for the specific ML algorithm defining the corresponding ML model. The aim of the numerical analysis is to identify the best set of seismic parameters (Table 1) to model the non-linear dynamic response of several structural systems using supervised ML techniques. This analysis also involves a comparison of the results using the indications deriving from the conditional information index.

The supervised ML algorithms, adopted in the numerical analysis, belong to five macro-categories:

- Linear Regression: this category represents the classical approach to model the relationship between input parameters and output variables through linear functions [78]. Four different versions of algorithms have been considered:

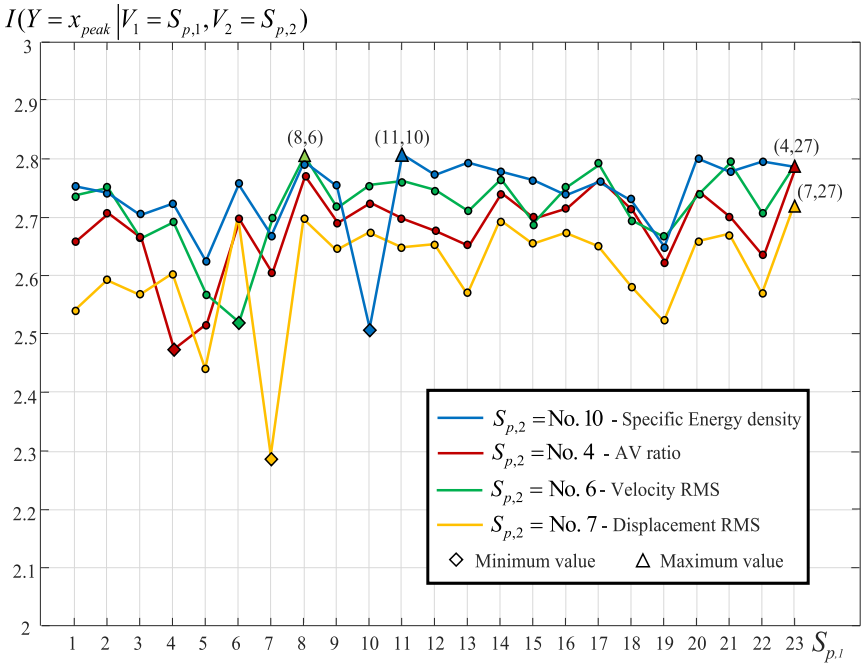


Fig. 13. Average conditional information in case of 2 input parameters – relevant pairs.

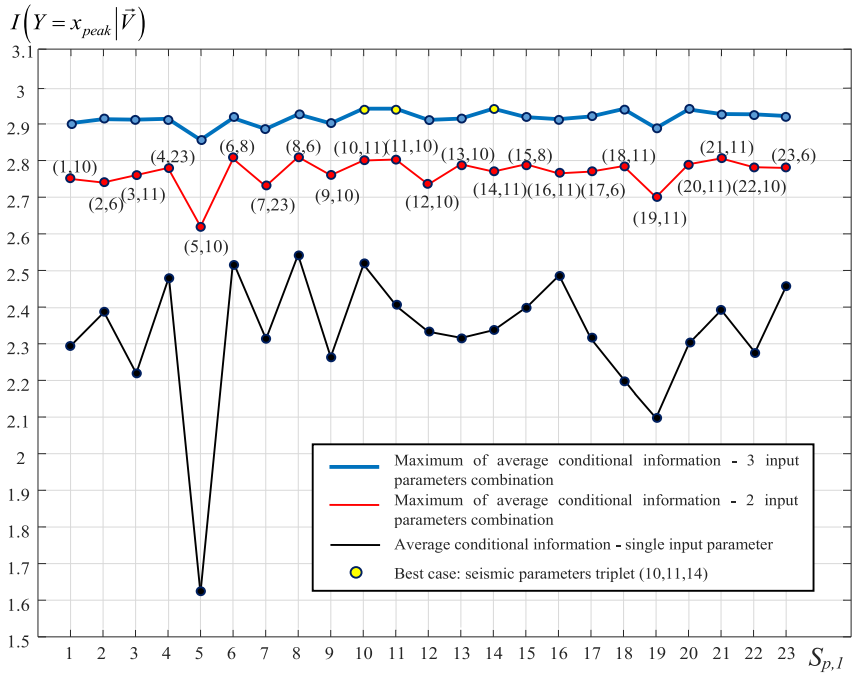


Fig. 14. Conditional information comparison between different input parameters cases.

- linear: the algorithm minimizes the sum of squared differences between predicted and observed values;
- interaction linear: the algorithm introduces interactive terms capturing non-linear relationships or conditional dependencies between independent parameters;
- robust: the algorithm introduces a loss function to mitigate the impact of outliers during training, ensuring reduced sensitivities to errors from anomalous values;
- stepwise: the algorithm sequentially adds or removes parameters basing on statistical criteria, this simplifies the model by considering only the most significant parameters.

- Tree: this category is based on decision trees, which iteratively split data into homogeneous subsets based on their most significant features. Three versions of algorithms have been utilized: fine, medium and coarse, based on the number of terminal nodes in the tree structure. Finer structures capture more details from input data but are more susceptible to overfitting, creating a model specific to the training set rather than generalizing to the entire dataset [79].
- Support Vector Machine (SVM): the goal of this category is to find a hyperplane in an N-dimensional space that optimally classifies input data. Six versions of algorithms have been tested: three versions named as linear, quadratic and cubic SVM, utilizing polynomial

kernel functions of the first, second and third degree, respectively. These kernels transform data from the original space to a higher-dimensional space to define more complex decision boundaries. The other three versions, named as SVM with fine or medium or coarse Gaussian kernel, present a kernel function between two vectors defined by a Gaussian function, with its width depending on parameter γ. This parameter varies in the three cases, influencing the algorithm's ability to adapt to complex patterns [80].

- Ensemble: this category combines different approaches to enhance overall performance. Two distinct versions of algorithms are considered:
  - Boosted Trees: this algorithm combines several simpler approaches, often shallow decision trees or weak learners, to create a more robust predictive ML model. The iterative training process focuses on correcting errors made by preceding ML models;
  - Bagged Trees: this algorithm combines multiple instances of decision trees using bootstrap sampling with replacement. Each tree is trained on a different subset of the data, introducing diversity to the ensemble and improving overall robustness [81].
- Neural Networks: this category is inspired by the functioning of the human brain, consisting of layers of interconnected artificial neurons (i.e., nodes). Each node receives input, undergoes a weighted transformation and produces an output. Training involves optimizing connection weights, enabling the network to learn complex patterns in the data. The various algorithms derive from the number of nodes considered and number of hidden layers. Specifically, five versions of algorithms are considered: three with a single hidden layer of different sizes: narrow (i.e., 10 nodes), medium (i.e., 25 nodes) and wide (i.e., 100 nodes); two with two and three hidden layers, respectively [82].

Therefore, 20 different algorithms have been analyzed with respect to the 23 seismic parameters to model the responses of the 1000 non-linear SDOF systems to the 400 artificial records.

As for the ML model validation, the method called "*N*-fold cross-validation" has been employed. This implies to divide the original dataset, composed of the 400000 cases, into *N* equal parts, denoted as "validation folds". The training and validation process is repeated *N* times, each time using a fold as validation set and the remaining $N-1$ folds as the training set. This way, each observation in the dataset has the opportunity to be in both the training and validation sets. During each iteration, the model is trained on the $N-1$ training folds and evaluated on the basis of the single validation fold. In this study, *N* was set to 5, dividing the dataset into 5 subsets for each analyzed case (i.e., each algorithm with each seismic parameter).

Two evaluation metrics are then computed: specifically, the root mean squared error (RMSE) and the Mean Absolute Error (MAE). The RMSE provides a measure of the dispersion between the model-predicted values and actual values in the validation dataset, as expressed by Eq.(11):

$$RMSE = \sqrt{\frac{1}{N}\sum_{s=1}^{N}\left(x_{s,true} - x_{s,pred}\right)^2} \tag{11}$$

in which *N* represents the total number of observations in the validation dataset (i.e., *N*=400000), $x_{s,true}$ is the target parameter value for the non-linear SDOF system, represented by the peak relative displacement and hysteretic energy obtained from the parametric analysis (Section 3), and $x_{s,pred}$ is the predicted value by using the considered ML model. The MAE is a measure of the accuracy of the prediction model, it is the average of the absolute errors, which is the absolute value of the difference between the model's predictions and the observed actual values, Eq.(12):

$$MAE = \frac{1}{N}\sum_{s=1}^{N}\left|x_{s,true} - x_{s,pred}\right| \tag{12}$$

The combined use of both metrics allows for a more comprehensive evaluation of the model's performance. In fact, while the RMSE reveals the presence of outliers and level of dispersion in the predictions, the MAE provides a simple and straightforward measure of the average error, being less sensitive to high-intensity prediction errors. This approach is particularly useful when dealing with limited datasets, aiming to obtain reliable performance estimates. It also helps to identify the variability in model performance across different partitions of the dataset, providing a more robust assessment of its ability to capture the relationship between input and output data.

To show some results of the validation process, Figs. 15–16 report both the RMSE and MAE values for the 20 algorithms considered, assuming the peak velocity as the seismic parameter (Table 1) and the two output variables (i.e., peak relative displacement and hysteretic energy) related to all the 1000 non-linear SDOF systems subjected to the 400 signals. Figs. 15 and 16 also highlight that the Bagged Trees and Fine Tree algorithms respectively represent the best predictive algorithms for peak relative displacement and hysteretic energy outputs.

Similar analyses were also carried out by considering the other input parameters (Table 1) to define the corresponding most predictive algorithms, as deducible in the following.

In addition, for such best-performing algorithm, Fig. 17 shows the partial dependent plot where the structural response in terms of peak relative displacement averaged over the 1000 SDOF systems predicted by the ML model is illustrated with respect to the peak velocity of the 400 signals. This emphasizes the type of relationship between the input parameters and output variables within the ML model obtained during the training phase.

6. Comparison between the ml models

The performance of all the 20 considered ML algorithms has been assessed with respect to the 23 seismic parameters through numerical analyses, where a set of seismic parameters, denoted as $S_{P,i}$ (Table 1), is provided as input data together with the 400 signals and dynamic characteristics of the 1000 elastic perfectly-plastic SDOF systems. The main scope is to numerically test and compare the significance of the seismic parameters within ML models and verify if the conditional information, previously computed, can be considered an effective index to select the best-performing combination of the input parameters in estimating the non-linear dynamic response of the structural systems.

To achieve this goal, 960 ML models have been implemented, as follows:

- a single seismic parameter S_P as input parameter and 20 ML algorithms (460 cases);
- 15 significant pairs of seismic parameters $\left(S_{P,1},\ S_{P,2}\right)$ as input parameters and 20 different ML algorithms (300 cases);
- 10 significant triplets for seismic parameters $\left(S_{P,1},\ S_{P,2},\ S_{P,3}\right)$ as input parameters and 20 different ML algorithms (200 cases).

For all these models, the effectiveness of the model prediction has been analyzed by using both the RMSE and MAE indices with respect to both peak relative displacement and hysteretic energy outputs, as commented in the following subsections.

6.1. A single seismic parameter as input parameter

The main results related to a single input parameter are summarized in Figs. 18–20. In Fig. 18, for each seismic parameter S_p the ML algorithm that achieved the best performance is highlighted. Specifically, Fig. 18(a),(b) refer to the RMSE index for both peak relative displacement and hysteretic energy, respectively, while Fig. 18(c),(d) present the results related to the MAE index.

The results in Fig. 18 clearly highlight that the most effective

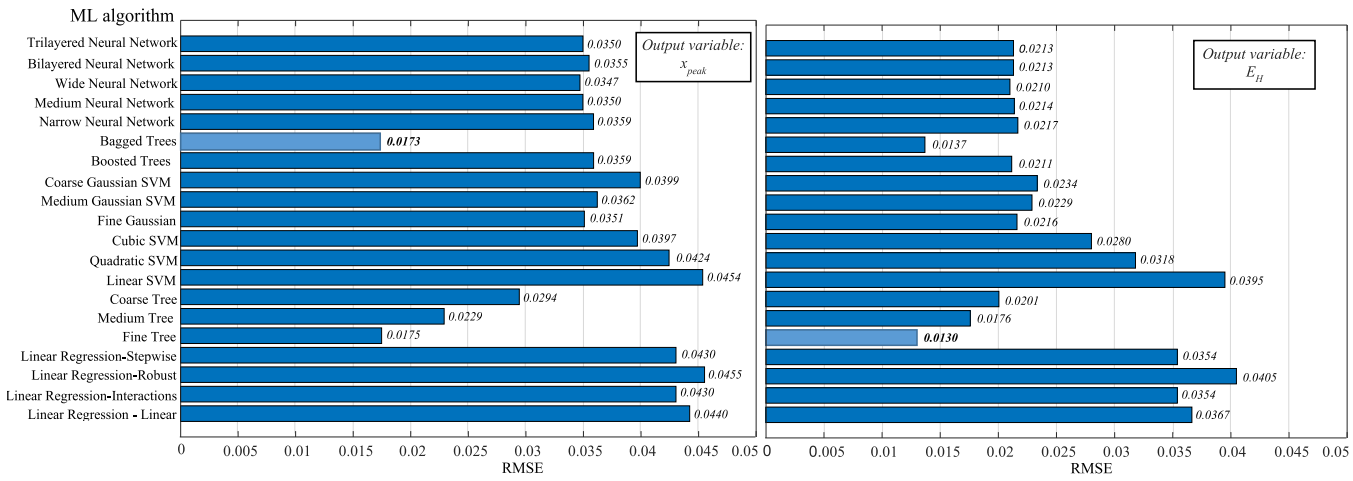


Fig. 15. RMSE for each considered algorithm assuming the peak velocity as seismic parameter.

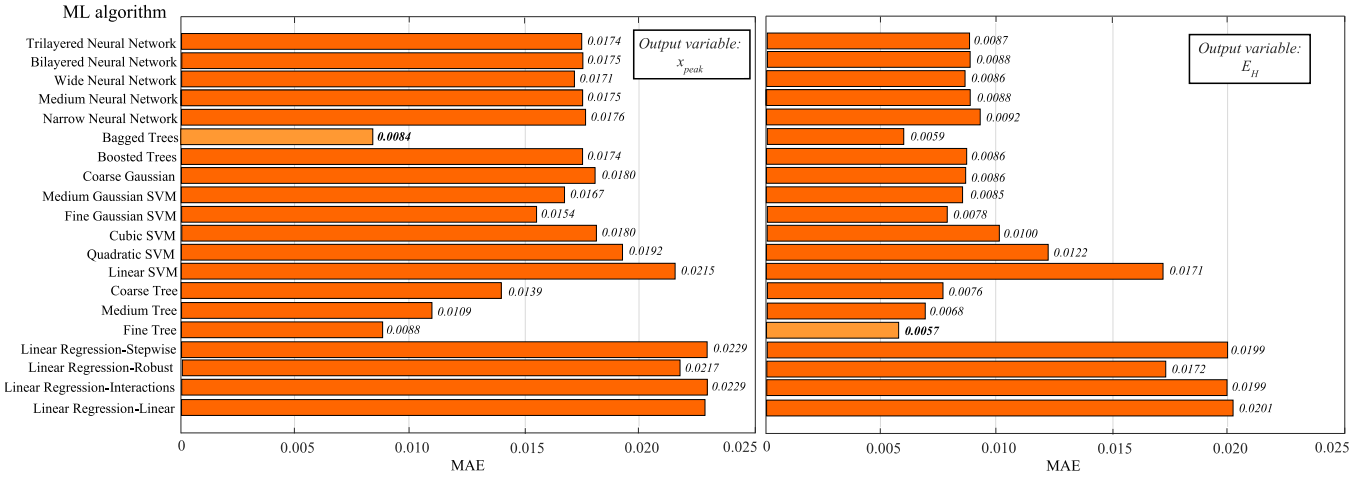


Fig. 16. MAE for each considered algorithm assuming the peak velocity as seismic parameter.

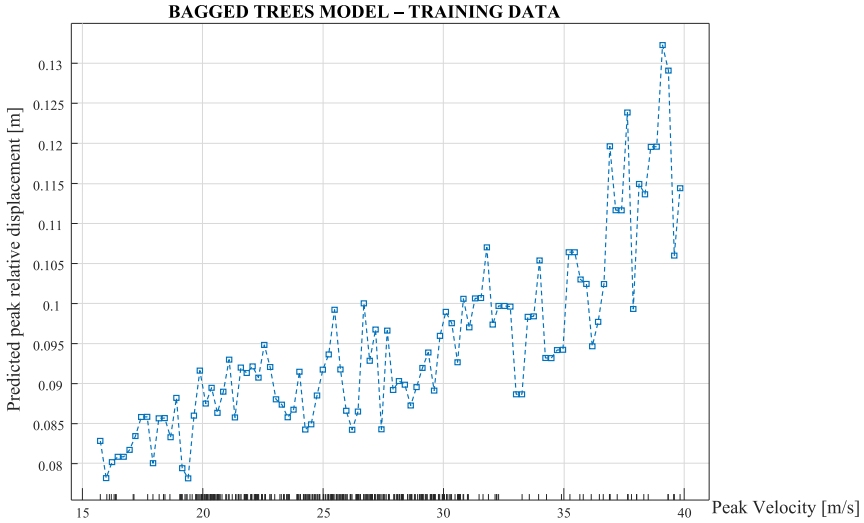


Fig. 17. Partial dependent plot for bagged trees model in terms of peak relative displacement.

algorithms for achieving the prediction objective are Bagged Trees, Fine Tree, Coarse Tree, Wide Neural Network and Fine Gaussian SVM. Among these, Bagged Trees and Fine Tree stand out as the best choices, particularly, in cases involving seismic parameters which lead to better predictive performances. Generally, algorithms based on decision trees exhibit higher average predictive performance in this case, as shown in

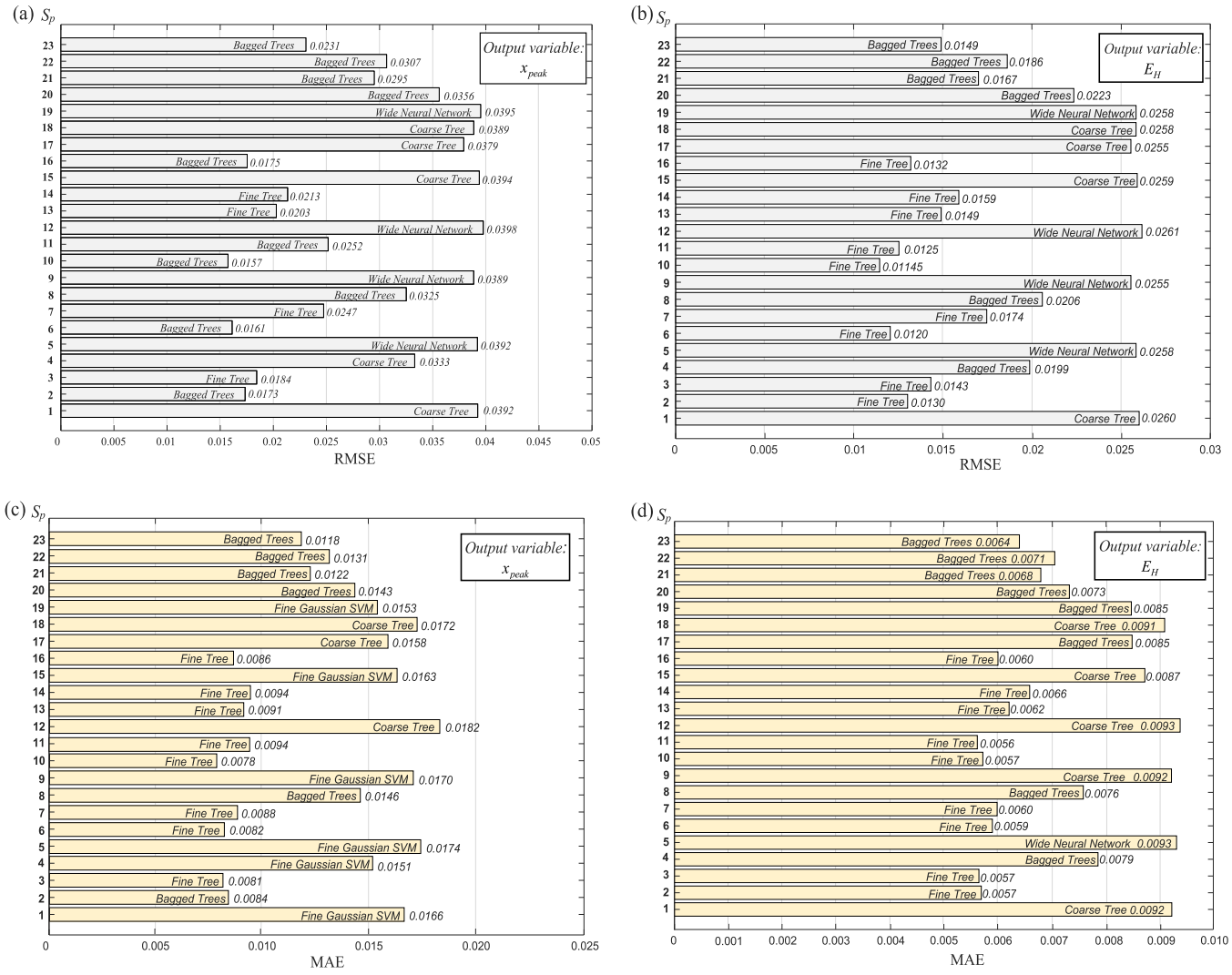


Fig. 18. Prediction performance obtained by the most effective ML algorithm.

Fig. 19, for both the metrics and the two output variables considered. This figure shows the performance achievable with some relevant algorithms as the seismic parameters vary, for both the metrics and the two output variables. It highlights how the choice of the seismic parameter S_p plays a significant role. Based on the performance predictions, the set of 23 seismic parameters can be divided into two groups: (i) a group of five parameters (denoted as High Predictive Seismic Parameter (HPSP) set) where the RMSE index ranges between 1.5 % and 2.0 % for the peak relative displacement and 1.0 %-1.5 % for the hysteretic energy as well as the MAE index ranges between 0.75 % and 1 % for the peak relative displacement and around 0.75 % for the hysteretic energy; (ii) a second group where these indices are significantly higher.

The HPSP set includes the following seismic parameters: Peak Velocity, Peak Displacement, Velocity RMS, Specific Energy Density and Sustained Maximum Velocity. It is important to underline that the seismic parameters in the HPSP set achieve high predictive performance regardless of the metric used. The seismic parameters in the HPSP set involve velocity and displacement, as well as integral-based energetic factors. In contrast, parameters directly related to acceleration or time are less effective in this context.

In Fig. 20, the actual response in terms of peak relative displacement and the response obtained using the Bagged Trees model (i.e., x_{pred}), for all the signals and non-linear SDOF systems, are compared in two cases: one where the seismic parameter is the Cumulative Absolute Intensity (S_p = No. 11 – Figure20(a)) and another one where the seismic parameter is represented by a HPSP, namely the Specific Energy Density (S_p = No. 10 – Figure20(b)). The difference in the predictions obtained with the two ML models is evident, highlighting how the use of Specific Energy Density values allows for a more efficient modeling.

Figs. 21–24 show partial dependence plots to highlight the relationships between the structural properties (i.e., the system elastic period T_0 and the yield strength factor R_y), the response predicted by the ML model and the two seismic parameters (i.e., Specific Energy Density (**SED** – S_p= No. 10) and Velocity RMS ($\boldsymbol{v_{RMS}}$ – S_p= No. 6)), for which the ML model leads to the best predictions.

Specifically, Figs. 21 and 22, respectively, show how the elastic period, T_0, of the structural system and the yield strength factor, R_y, influence the peak relative displacement (Figs. 21(a) and 22(a)) and hysteretic energy (Figs. 21(b) and 22(b)), predicted by the ML model averaging over the 400 signals and other structural characteristics. These figures are related to both the Specific Energy density and Velocity RMS and present identical trend for any seismic parameter. It is highlighted how the Bagged Trees model, with the knowledge of a single seismic parameter belonging to the HPSP set, effectively manages to capture the dependence of the non-linear response on the mechanical characteristics of the SDOF systems. In detail, the relative displacement demand is greater in the case of lower stiffness levels (Fig. 21), while the

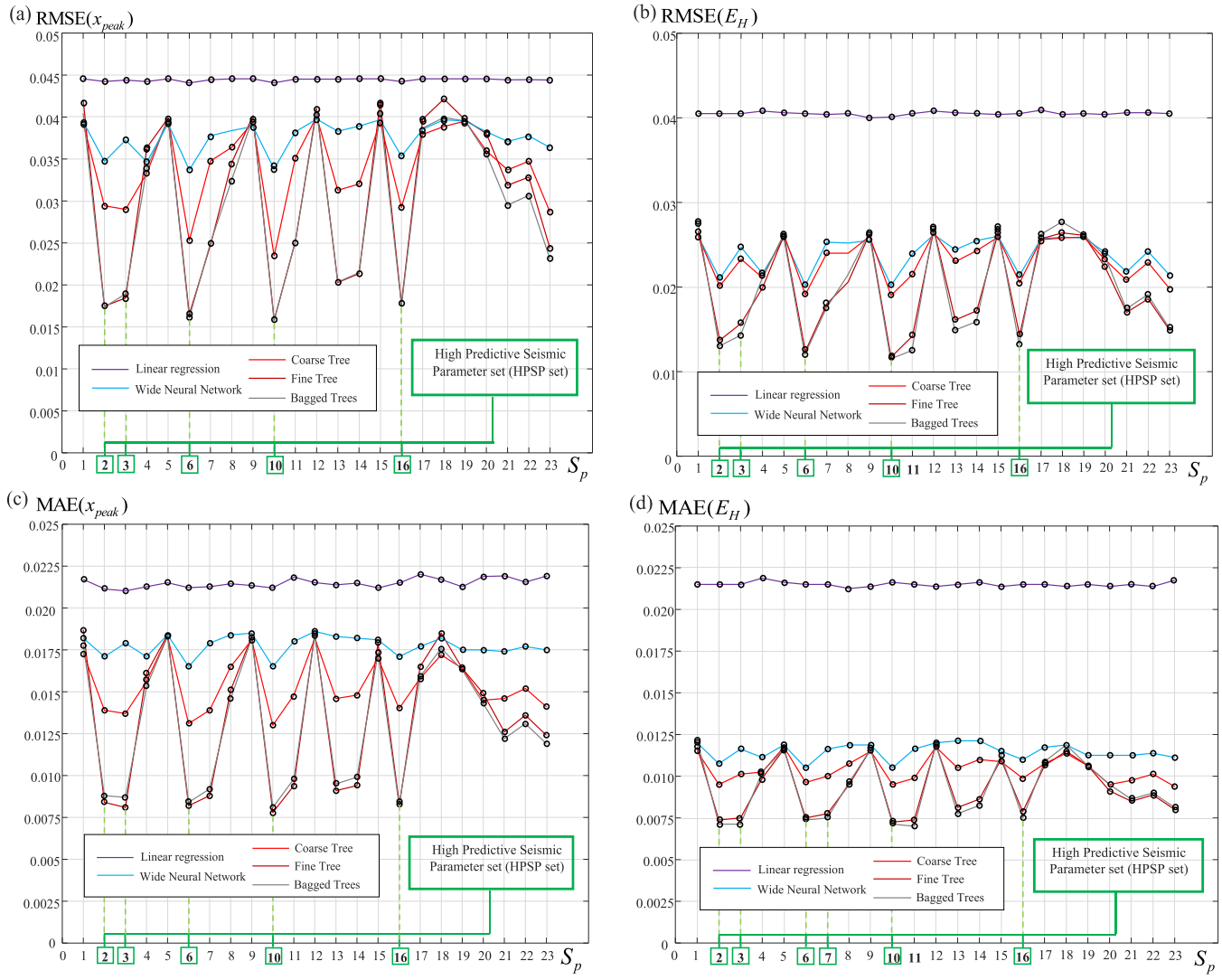


Fig. 19. Performance comparison between the different seismic parameters for some ML algorithms.

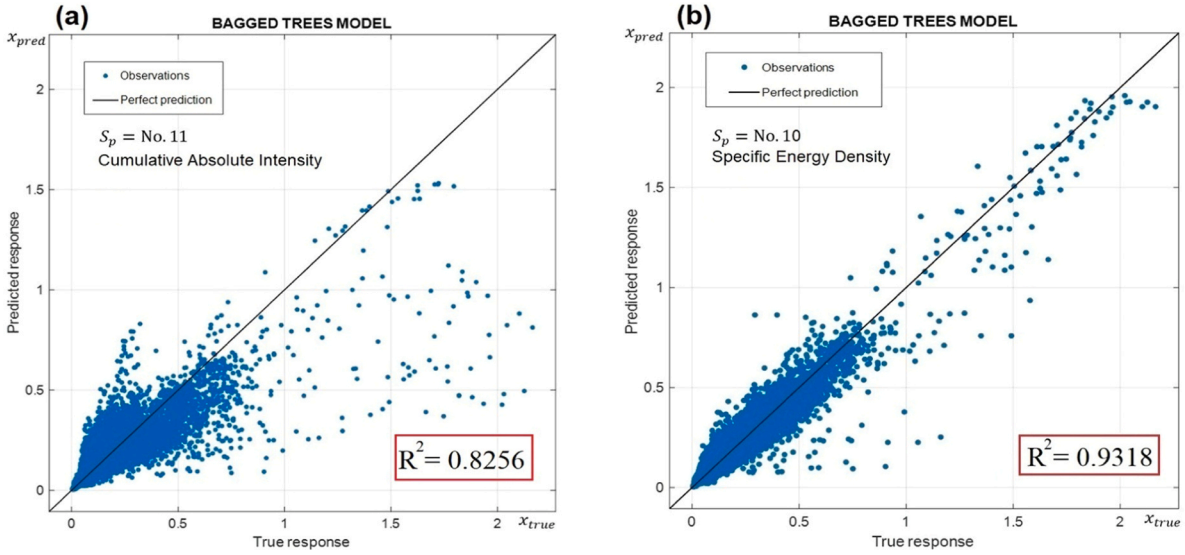


Fig. 20. Comparison between predicted response and true response in Bagged Trees models: (a) S_p=No. 11, (b) S_p= No. 10.

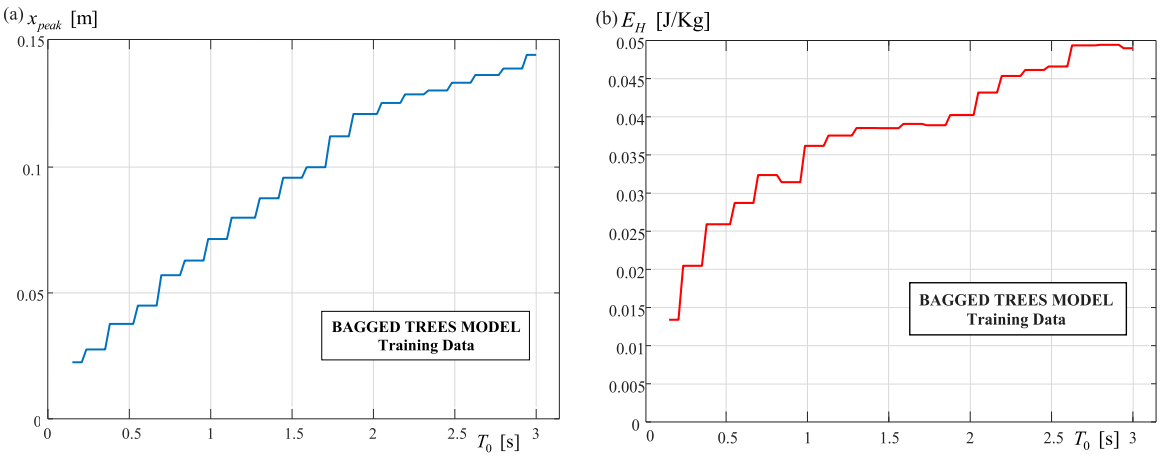


Fig. 21. Partial dependence plot of model prediction depending on SDOF system elastic period: (a) peak relative displacement, (b) hysteretic energy.

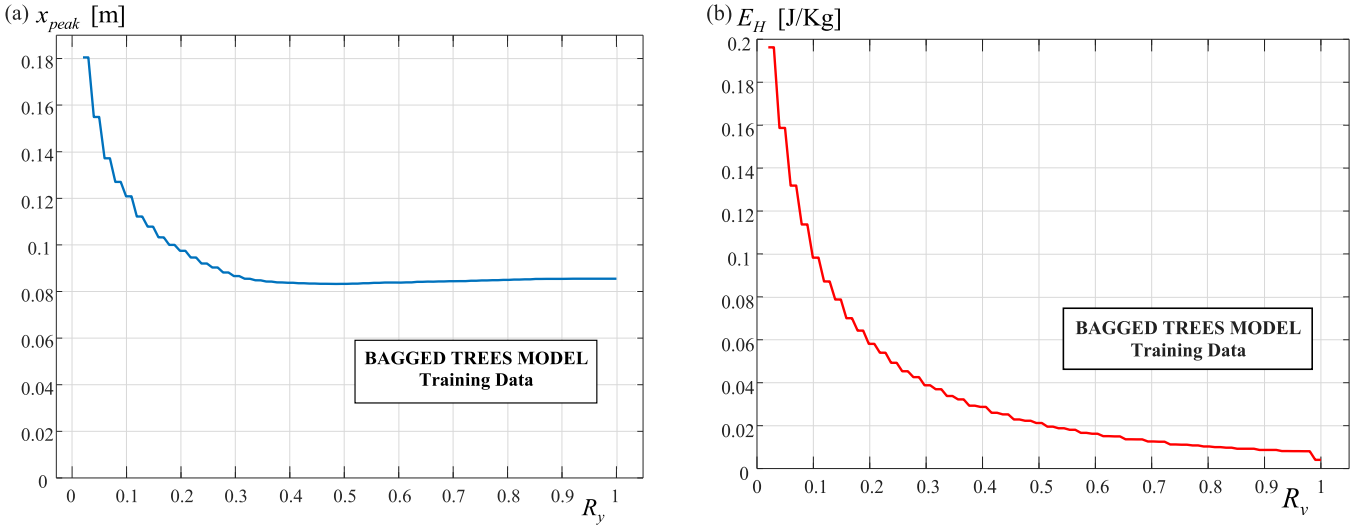


Fig. 22. Partial dependence plot of model prediction depending on SDOF system yield strength factor: (a) peak relative displacement, (b) hysteretic energy.

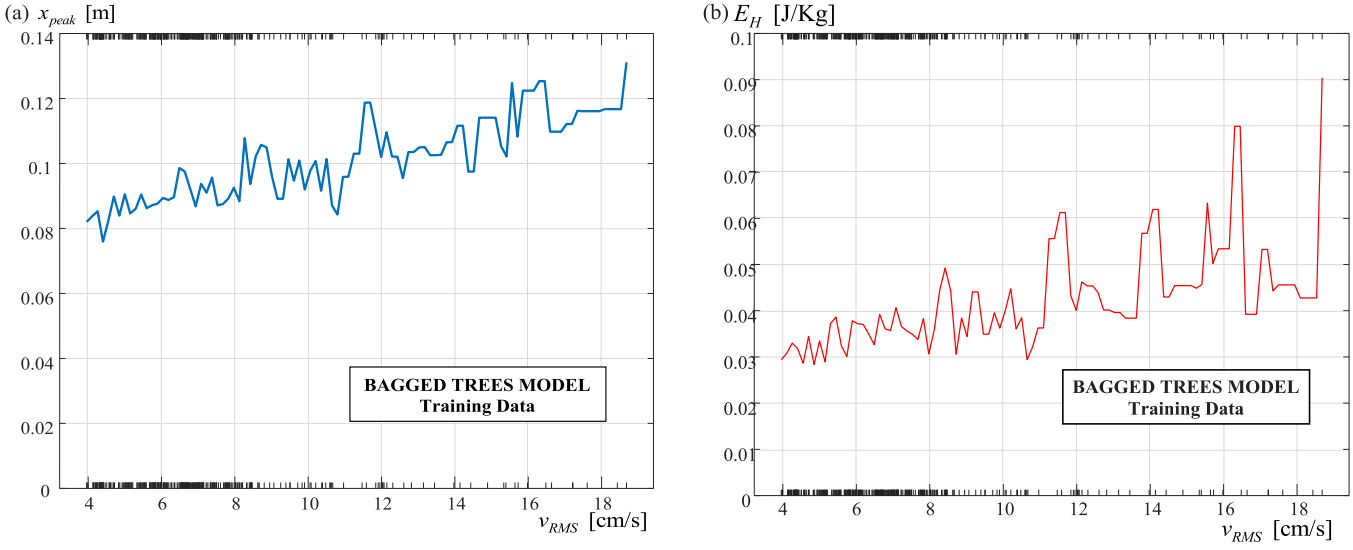


Fig. 23. Partial dependence plot of model prediction depending on seismic parameter (v_{RMS}): (a) peak relative displacement, (b) hysteretic energy.

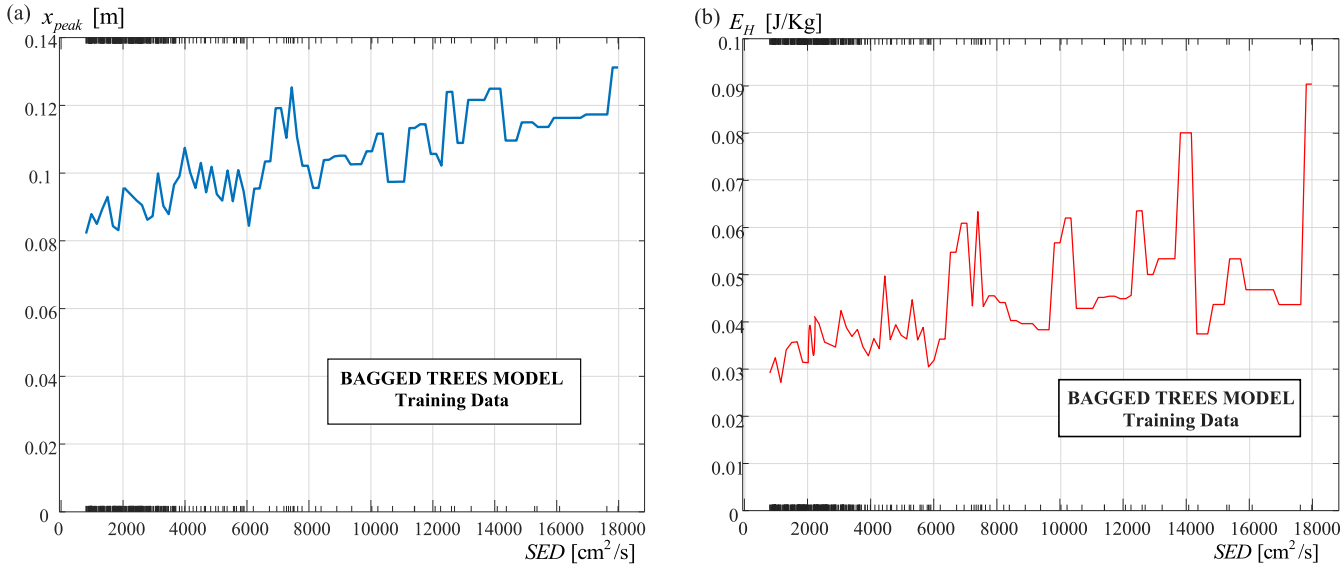


Fig. 24. Partial dependence plot of model prediction depending on seismic parameter (***SED***): (a) peak relative displacement, (b) hysteretic energy.

trend of the peak relative displacement as a function of the elastic resistance of the equivalent SDOF system is consistent with the formulations defining yield strength factors within the displacement-based design [71]. As for the hysteretic energy, the trend shown in Fig. 21 is perfectly consistent with the expected prediction for elastic perfectly-plastic SDOF systems having different initial stiffness: after an initial phase where the hysteretic energy in non-linear field increases rapidly, then the increase is less pronounced [83].

Figs. 23 and 24 show how these seismic parameters influence, respectively, the structural response in terms of peak relative displacement and hysteretic energy predicted by the ML model averaging over both the 400 signals and 1000 SDOF systems. The analysis of this relationship is particularly useful for understanding the physical role that each seismic parameter plays in the non-linear response of the system. In detail, the Bagged Trees model predicts a quite regular increasing trend, highlighting a significant linear correlation between both the input data and the seismic responses, especially in the range of feature values where the largest number of examples, used during the model's training phase, is available.

Analyzing the results obtained, it is worth noting how the performance achievable through ML predictive models, in the case of a single input parameter, is strongly correlated with the quantity of the conditional information calculated for each seismic parameter and relative to the empirical entropy of the seismic response (i.e., output). In Fig. 25, the values of the conditional information relative to each seismic parameter and mean performance values, in terms of prediction (i.e., RMSE), of all ML algorithms considered, are represented. In order to effectively compare these values and highlight that higher information generally corresponds to lower RMSE values, the graph has two scales, with different representations: a classic linear scale for $I(Y|V_1)$, and a

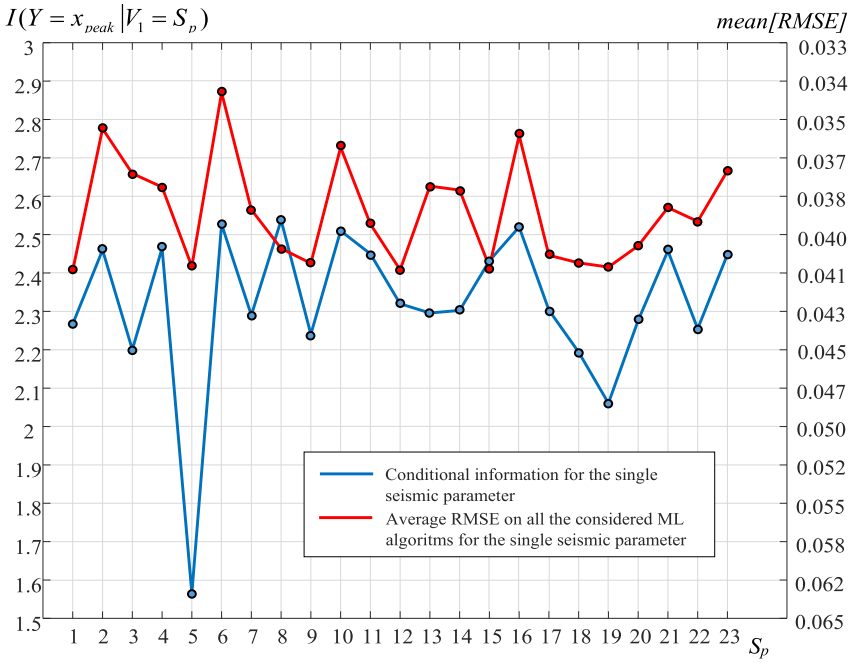


Fig. 25. Comparison between Shannon conditional information and average RMSE for every seismic parameter.

functional scale for $RMSE : (10 \bullet mean[RMSE])^{-1}$. The resulting graph clearly highlights how the parameters that yield higher conditional information are also those for which generally the best predictive performances are obtained by considering the average value over all the algorithms. Exceptions are represented by the Arias Intensity parameter (No. 8), whose choice leads to lower predictive results with respect to the conditioned information, and by the Velocity Spectrum Intensity (No. 13) and Housner Intensity (No. 14) parameters, which lead to better predictions with respect to the conditional information. Qualitatively identical results were obtained in the case where the MAE metric and hysteretic energy were considered. The conditional information therefore appears to be an effective index for guiding the selection of a single seismic parameter to be used as input for the best numerical modeling based on ML algorithms.

6.2. Two seismic parameters as input parameters

The next step in this subsection is to assess how predictive performance can be improved by considering multiple seismic input parameters, selected on the basis of the conditional information. The results, represented in Figs. 11 - 14, allow to identify a series of best-performing combinations in terms of Shannon's conditioned information.

By considering the case of two input parameters, the pairs "Specific Energy Density (No. 10) - Cumulative Absolute Intensity (No. 11)" and "Velocity RMS (No. 6) - Arias Intensity (No. 8)" appear particularly promising, as well as expecting poor predictive results can derive in the case of the pair "Acceleration RMS (No. 5) - Predominant Period (No. 19)". Following these indications and with the aim of verifying whether Shannon's conditional information can also be considered a reliable index, 15 pairs of seismic parameters have been considered as input parameters for the ML algorithms examined. These pairs are listed in Table 2 with their respective alphanumeric labels (Table 1). They were selected with the aim of investigating the performance obtained through input pairs having various informative content, pairs for which higher predictive performance is expected along with cases where lower expected performances are considered.

For each of these pairs, the predictive performance in terms of RMSE and MAE achievable by the different algorithms has been analyzed for both the two model outputs. In detail, in Fig. 26(a)-(b), the RMSE values are shown for the algorithm that proved to be the most efficient in modelling the seismic peak relative displacement and hysteretic energy, whereas in Fig. 26(c)-(d), the same cases are presented using the MAE performance metric.

The results show that, except for cases where pairs with lower conditional information are chosen, the use of two seismic parameters generally allows for an improvement of 10–20 % in predictive performance of the best ML algorithm compared to the scenario where only a single seismic parameter is provided. The ML algorithm providing the best performance is generally the 'Bagged Trees'. It is notable to highlight that the best predictive performances are achieved precisely in the case of pairs with higher conditional information, particularly, combinations denoted as C4, C13 and C14 are the most performing. These combinations consistently include the Specific Energy Density (No. 10) parameter, which individually exhibited higher performance, indicating its effective pairing with other seismic parameters without duplicating the information of any second input parameter.

Table 2
Seismic parameters pairs in numerical analysis.

S_p pair	Seismic parameters
C1	Peak acceleration (No. 1) – Peak velocity (No. 2)
C2	Peak acceleration (No. 1) – AV Ratio (No. 4)
C3	Peak velocity (No. 2) – Velocity RMS (No. 6)
C4	Peak velocity (No. 2) – Specific Energy Density (No. 10)
C5	Peak velocity (No. 2) – Velocity Spectrum Intensity (No. 13)
C6	Peak velocity (No. 2) – Housner Intensity (No. 14)
C7	AV ratio (No. 4) – Significant Duration (No. 23)
C8	Acceleration RMS (No. 5) – Specific Energy Density (No. 10)
C9	Acceleration RMS (No. 5) – Predominant Period (No. 19)
C10	Velocity RMS (No. 6) – Arias Intensity (No. 8)
C11	Velocity RMS (No. 6) – Specific Energy Density (No. 10)
C12	Displacement RMS (No. 7) – Housner Intensity (No. 14)
C13	Specific Energy Density (No. 10) – Cumulative Absolute Intensity (No. 11)
C14	Specific Energy Density (No. 10) – Sustained Maximum Velocity (No. 16)
C15	A95 parameter (No. 18) – Mean Period (No. 20)

6.3. Three seismic parameters as input parameters

In this subsection the combinations of three seismic parameters are considered as input parameters. The triplets with higher values of the conditional information are generally those containing the Specific Energy Density (No. 10) and Cumulative Absolute Intensity (No. 11) in the selected seismic parameters. Particularly, these two parameters combined with the Housner Intensity represent the triplet with the highest informational content with respect to the seismic response of SDOF systems. Other triplets leading to high information include combinations of a peak parameter (e.g., Peak Velocity - No. 2), an energy parameter (e.g., Specific Energy Density - No. 10) and a time parameter (e.g., Bracketed Duration - No. 22). Guided by these considerations, the predictive performances of the ML algorithms were analyzed concerning 10 triplets of input parameters, as indicated in Table 3. These sets were chosen following the same criteria adopted in Subsection 6.2, with the aim to verify a range of different predictive performances: some triplets are between the best-performing ones (i.e., T3, T4, T6, T7, T8), other ones between those expected to have a medium predictive behavior and one with lower informational content (i.e., T5).

In Fig. 27, the predictive performance associated with these choices is illustrated for the best-performing model (i.e., Bagged Trees and Fine Tree model). The results align well with expectations for both the metrics and output variables. They demonstrate a significant improvement in modeling performance, compared to the best-case scenario with two seismic parameters, specifically, for the triplets identified as best-performing with respect to conditional information. Additionally, the analysis underscores how a poor selection of seismic parameters (e.g., T5 triplet) can lead to models with low predictive capacity, although a higher number of parameters. These findings are consistent when the average performance across all the considered ML algorithms is taken into account.

Fig. 28 compares the actual seismic response with the predicted response from the best model for both peak relative displacement and hysteretic energy outputs. The comparison is carried out for: a single seismic parameter, a pair of seismic parameters and a triplet of seismic parameters as input data. In the case of peak relative displacement, a reduced dispersion is evident in the latter cases, primarily attributed to an improvement in estimation for cases where the seismic response is less than 0.8 m. The additional information provided by the second and third input parameter seems not to significantly enhance prediction only in cases where seismic excitations lead to high seismic response values. It is important to note, however, that these cases refer to situations in which the analyzed SDOF system exhibits very low stiffness or resistance values, corresponding to specific practical situations. As for hysteretic energy, the improvement achieved using additional input parameters appears to be more evenly distributed, with a slight underestimation of the actual output value.

Figs. 29–30 respectively illustrate the relationship between the seismic parameter "Specific Energy Density" and the peak relative displacement (Fig. 29) and hysteretic energy (Fig. 30) seismic responses for the single input parameter $S_p = 10$, the input parameter pair C13 and the input parameter triplet T6.

It is important to emphasize how the inclusion of multiple seismic input parameters significantly regularizes the increasing trend of these relationships. The "Specific Energy Density" parameter (SED – S_p= No. 10), therefore, demonstrates a strong correlation with non-linear seismic

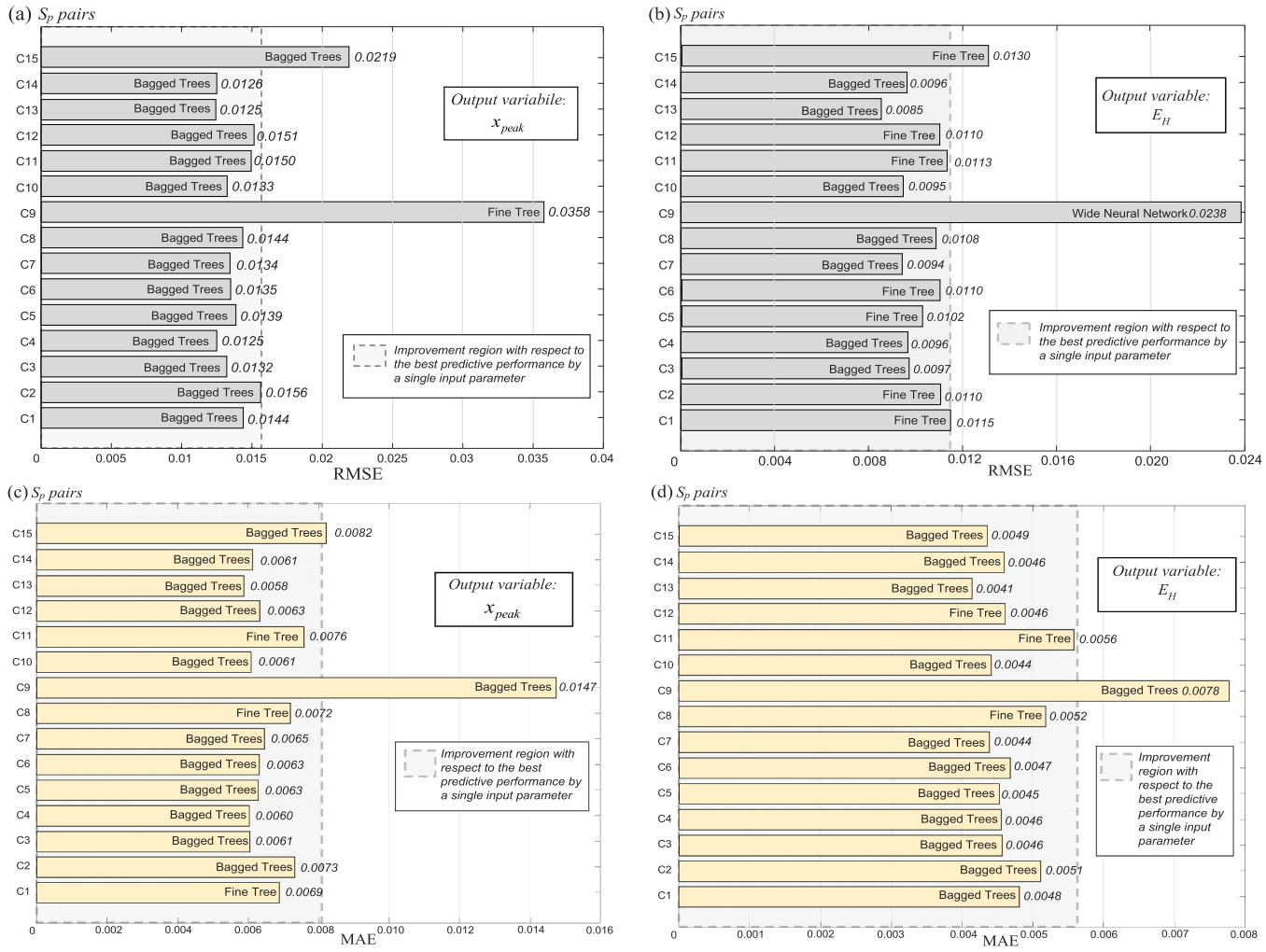


Fig. 26. Minimum values of both RMSE and MAE for each considered seismic parameters pairs.

Table 3
Seismic parameters triplets in numerical analysis.

S_p triplet	Seismic parameters
T1	Peak acceleration (No. 1) – Peak velocity (No. 2) – Peak displacement (No. 3)
T2	Peak velocity (No. 2) – Acceleration RMS (No. 5) – Predominant Period (No. 19)
T3	Peak velocity (No. 2) – Specific Energy Density (No. 10) - Bracketed Duration (No. 22)
T4	AV ratio (No. 4) – Arias Intensity (No. 8) – Significant Duration (No. 23)
T5	Acceleration RMS (No. 5) – A95 parameter (No. 18) – Predominant Period (No. 19)
T6	Specific Energy Density (No. 10) – Cumulative Absolute Velocity (No. 11) – Housner Intensity (No. 14)
T7	Specific Energy Density (No. 10) – Cumulative Abs. Velocity (No. 11) – Sustained Max Velocity (No. 16)
T8	Specific Energy Density (No. 10) – Cumulative Abs. Velocity (No. 11) – A95 parameter (No. 18)
T9	Cumulative Absolute Velocity (No. 11) – Housner Intensity (No. 14) – Predominant Period (No. 19)
T10	Uniform Duration (No. 21) – Bracketed Duration (No. 22) – Significant Duration (No. 23)

responses, and this relationship can be reliably represented using simple mathematical expressions.

The correlation level of the SED parameter, similarly to other seismic parameters with high Shannon conditional information values, is significantly higher than that of the PGA parameter, which is currently used to represent site seismic hazard. Based on available data, this result suggests that this parameter could effectively be integrated to define probabilistic seismic hazard maps within performance-based design procedures aimed at controlling the non-linear response of structures.

7. ML models within performance-based seismic design

The most advanced seismic regulations foresee the possibility, in specific cases (e.g., the design of safety-critical structures, highly irregular buildings, base-isolated structures and structures designed for a high level of ductility), to perform non-linear dynamic analyses considering suites of accelerograms. Although this approach is theoretically the most general, it is characterized by numerical difficulties and long processing times. Implementation of ML models can simplify this approach and make it easier to use and more widely applicable. After identifying a seismic scenario in terms of magnitude, source-to-site distance and site geology, a perspective evolution can be the definition of a spectral response on the basis of specific site hazard studies [84] and a generation of several seismic excitations compatible with this spectrum. These seismic signals can be used for the training phase of a ML algorithm capable of predicting the seismic performance of a structure on varying one or more seismic parameters and main mechanical characteristics of the structural system. The advantage of such an approach is relevant: the estimation of the system performance would be carried out through numerically stable functions with dramatically

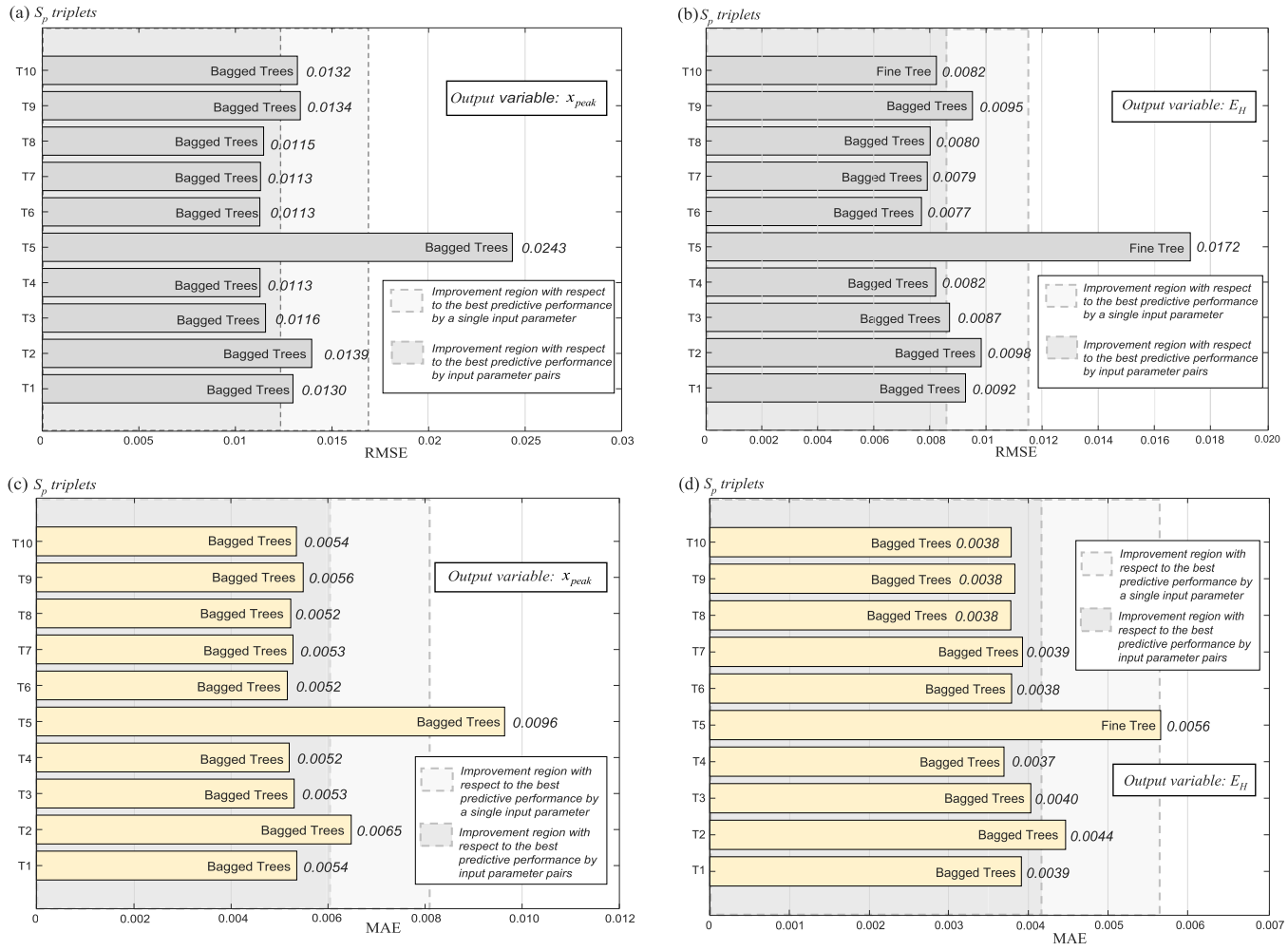


Fig. 27. Minimum RMSE for each considered seismic parameters triplets.

reduced computational times, allowing for a preliminary real-time assessment of the safety level of the structure.

Having this aim, the results of this study are useful to understand which set of seismic parameters are the most suitable to capture the information necessary for estimating the relative displacement and hysteretic energy demands of an event. Secondly, it is important to investigate the level of reliability that the ML model presents in terms of its predictive objective. It is therefore useful a comparative analysis between the predicted response, through the obtained Bagged trees model, with the response that the structural system exhibits in the case of natural ground motions with dynamic characteristics similar to those of the set used for the ML algorithm training phase.

In this context, it is also important to emphasize that the goal of ML modeling is to calibrate the algorithm based on the specific hazard characteristics of the reference site where the structure will be built. The present work has highlighted that ML model becomes more efficient when specific input parameters combinations are used to model the structural response. This means that a meaningful verification should focus on natural excitations whose seismic parameters, belonging to the HPSP set, have values close to those of the artificial records used for the training. Therefore, from the PEER database [85], ten seismic events (no-frequent records [86]) have been identified and properly scaled so that the HPSPs have values within the range defined by the synthetic accelerograms used to train the ML models. Table 4 reports the selected earthquake events, including the scale factors. These events have been considered to compute the non-linear dynamic responses of the 1000 SDOF systems through (a) numerical time integration of the motion equation using a step-by-step solution method, in comparison to (b) peak and energy responses prediction by means of the ML techniques (i.e., Bagged Trees model using the T6 input combination).

The results of these analyses are presented in Figs. 31–36. Figs. 31–33 illustrate the seismic response in terms of peak relative displacement (Fig. 31), hysteretic energy (Fig. 32) and Park & Ang damage index, D_{PA} (Fig. 33), for a single ground motion (i.e., #8 of Table 4). More specifically, the Park & Ang damage index [87] is a measure of the level of structural damage as a function of the peak relative displacement and hysteretic energy, expressed as follows:

$$D_{PA} = \frac{x_{peak}}{x_m} + \frac{\beta}{f_y \bullet x_m} \bullet E_H \tag{13}$$

where x_m is the relative displacement capacity of the system, f_y is the calculated yield strength and β is a parameter depending on the structural typology. In the following, x_m will be assumed to be five times the displacement of the system at the elastic limit (i.e., structural ductility equal to 5) and $\beta = 0.15$, a typical value for reinforced concrete structures.

The graphs on the left represent the seismic response obtained from the non-linear dynamic analyses as a function of the yield strength factor of the structural system, R_y for various values of the elastic period, T_0. The graphs on the right show the results predicted using the Bagged Trees model, considering the T6 combination of seismic parameters as input data. The comparison demonstrates how, for the seismic event #8 and without any information about the non-linear dynamic model, the

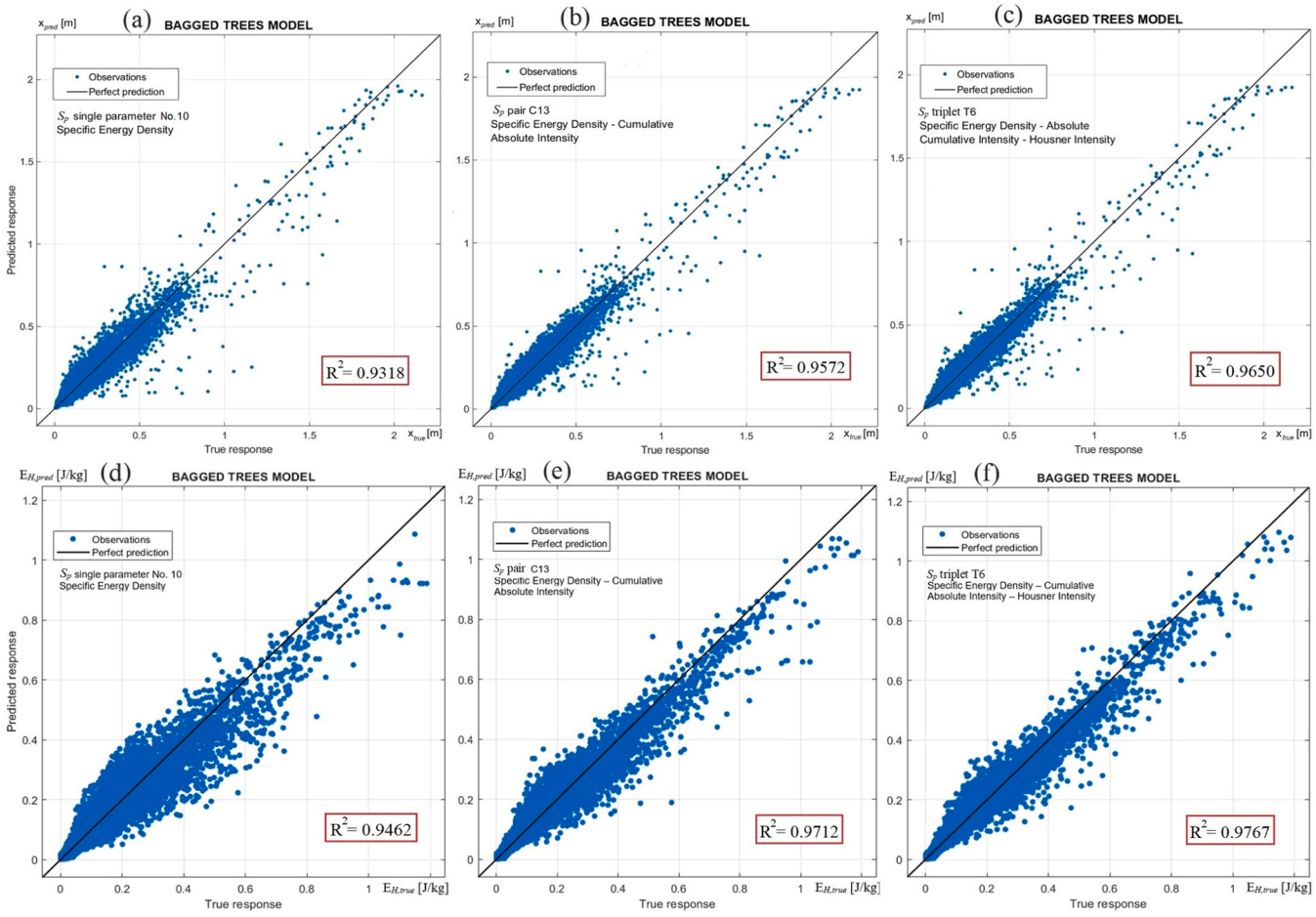


Fig. 28. Comparison between the predicted response and true response in Bagged Trees models: (a) x_{peak} - S_p= No. 10, (b) x_{peak} - C13, (c) x_{peak} - T6, (d) E_H - S_p= No. 10, (e) E_H - C13, (f) E_H - T6.

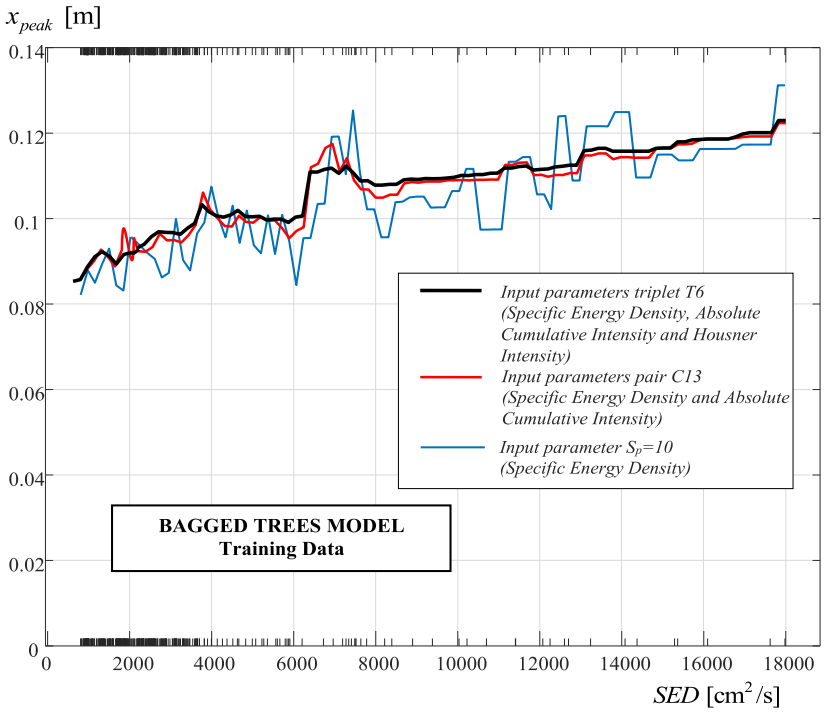


Fig. 29. Partial dependence plot comparison: seismic parameter ***SED*** vs peak relative displacement model prediction.

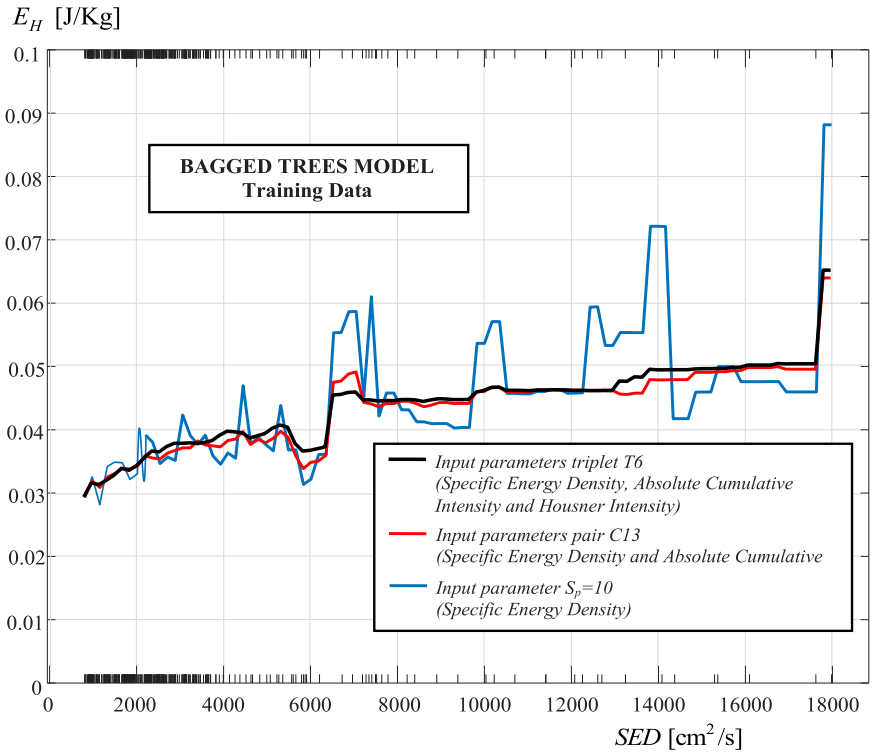


Fig. 30. Partial dependence plot comparison: seismic parameter ***SED*** vs hysteretic energy model prediction.

Table 4
Selected ground motion events.

#	Earthquake Name	Year	Station Name	Magnitude	Mechanism	Scale factor
1	Iwate_Japan	2008	IWT010	6.9	Reverse	1.0
2	Iwate_Japan	2008	Tamati Ono	6.9	Reverse	1.5
3	Iwate_Japan	2008	Semine Kurihara City	6.9	Reverse	1.5
4	Chuetsu-oki_ Japan	2007	Kawaguchi	6.8	Reverse	2.2
5	Chuetsu-oki_ Japan	2007	NIGH01	6.8	Reverse	1.0
6	Darfield_New Zealand	2010	DFHS	7.0	Strike slip	0.7
7	Loma Prieta	1989	San Jose - Santa Teresa Hills	6.93	Reverse Oblique	1.5
8	Landers	1992	North Palm Springs	7.28	Reverse	1.5
9	Imperial Valley−06	1979	Cerro Prieto	6.53	Reverse	1.5
10	Manjil_ Iran	1990	Abbar	7.37	Strike slip	0.7

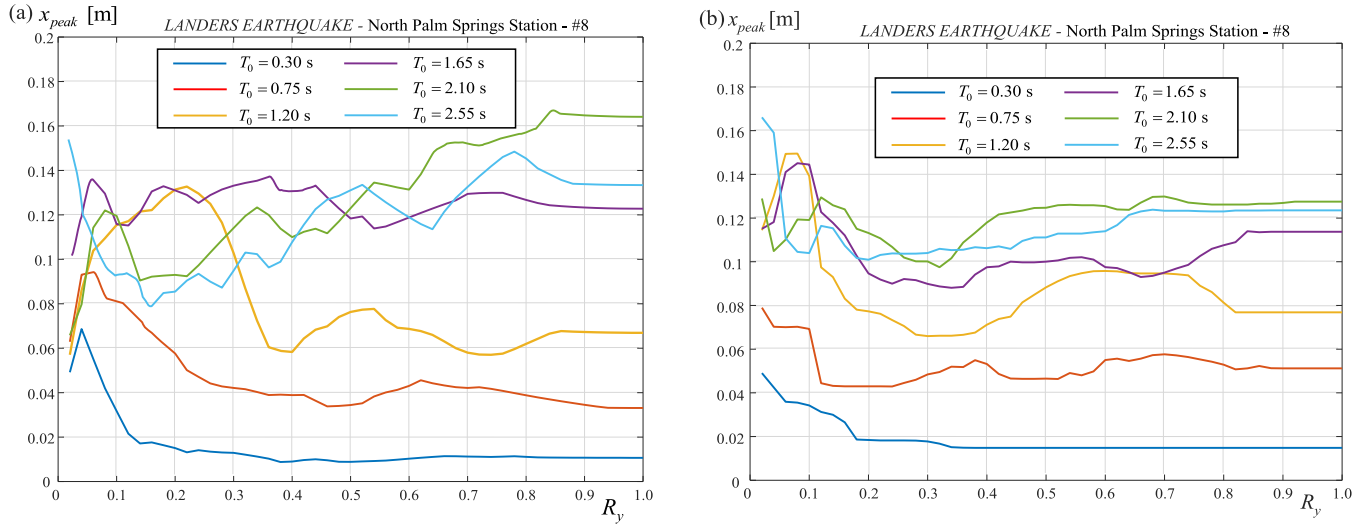


Fig. 31. Comparison between peak relative displacement response obtained by (a) dynamic analysis and (b) ML model.

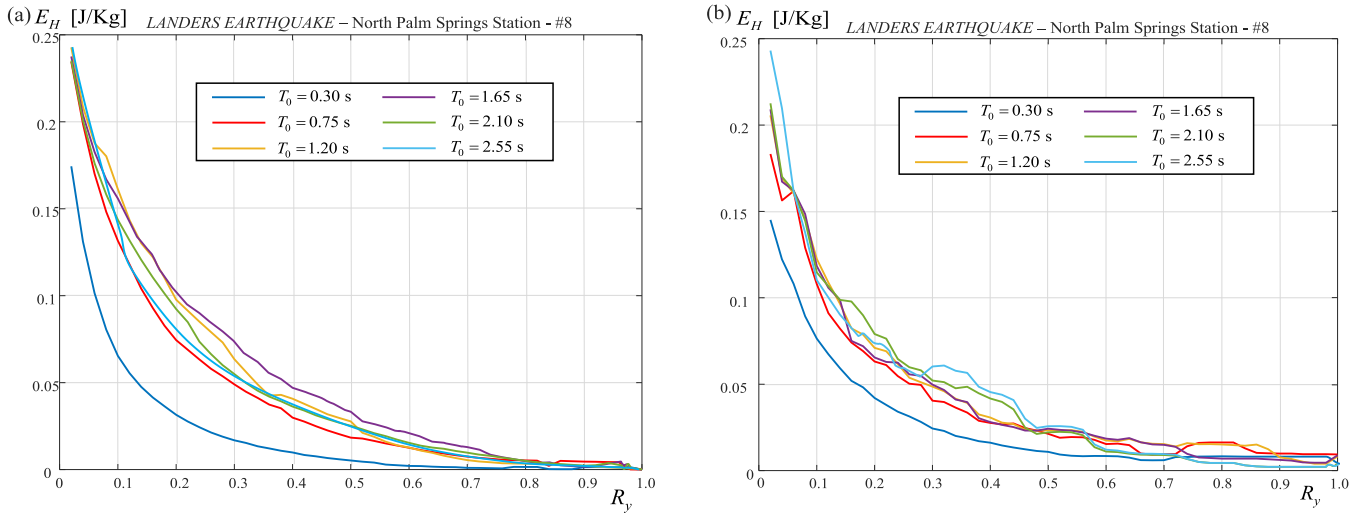


Fig. 32. Comparison between hysteretic energy response obtained by (a) dynamic analysis and (b) ML model.

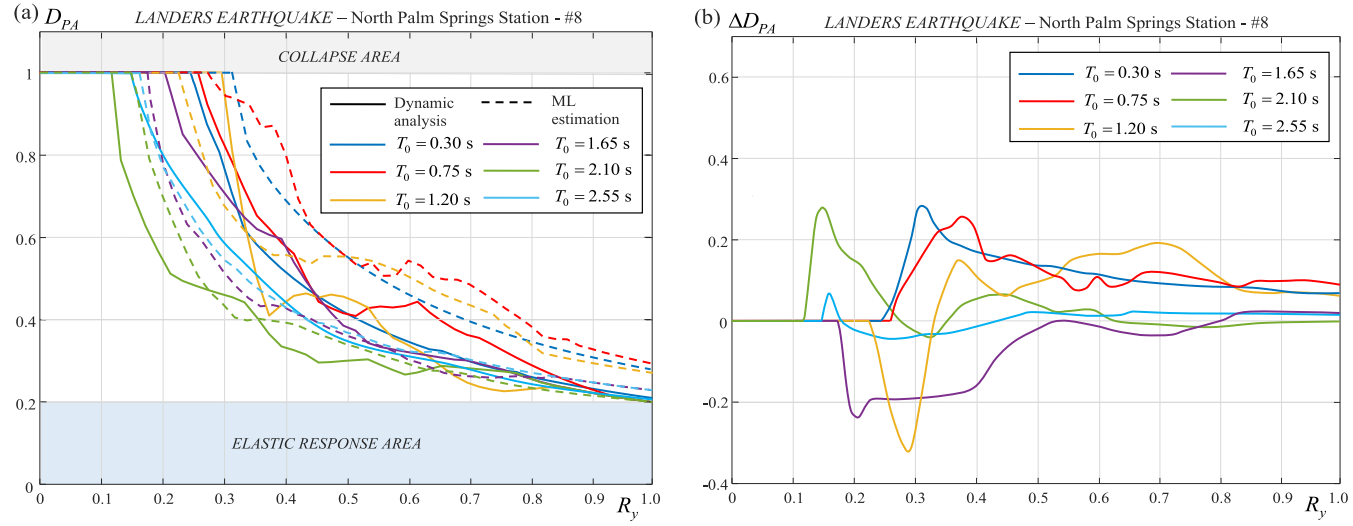


Fig. 33. Comparison between Park&Ang damage index obtained by dynamic analysis and ML model (a) and representation of the damage estimation error (b).

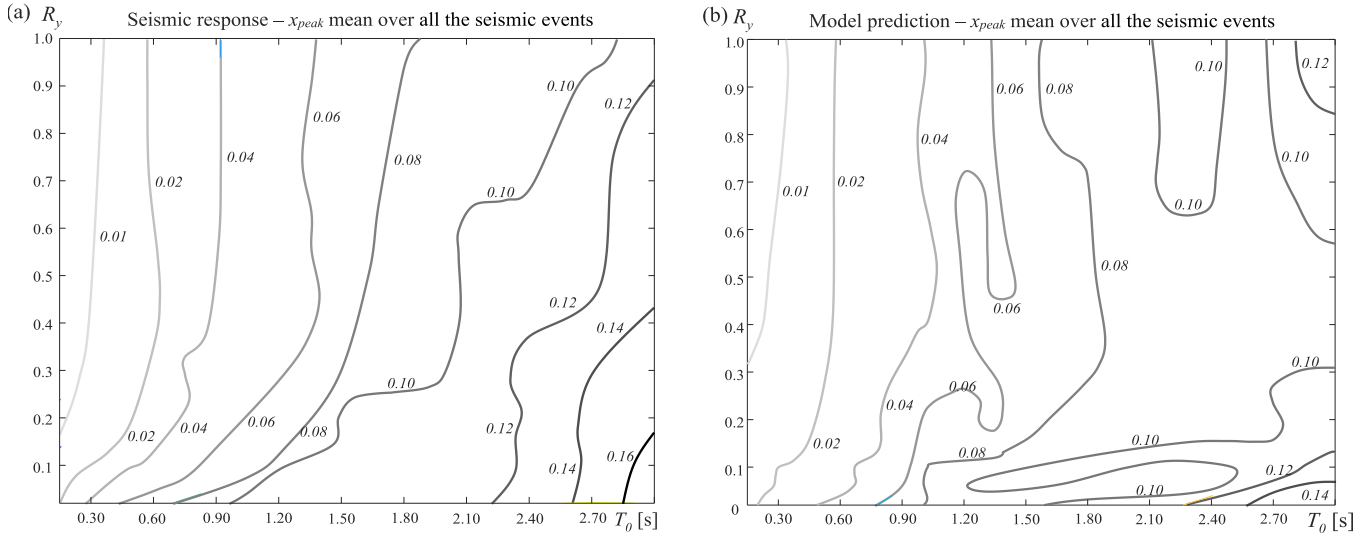


Fig. 34. Comparison between mean value of non-linear seismic response (x_{peak}) obtained by (a) dynamic analysis and (b) ML model – contour maps.

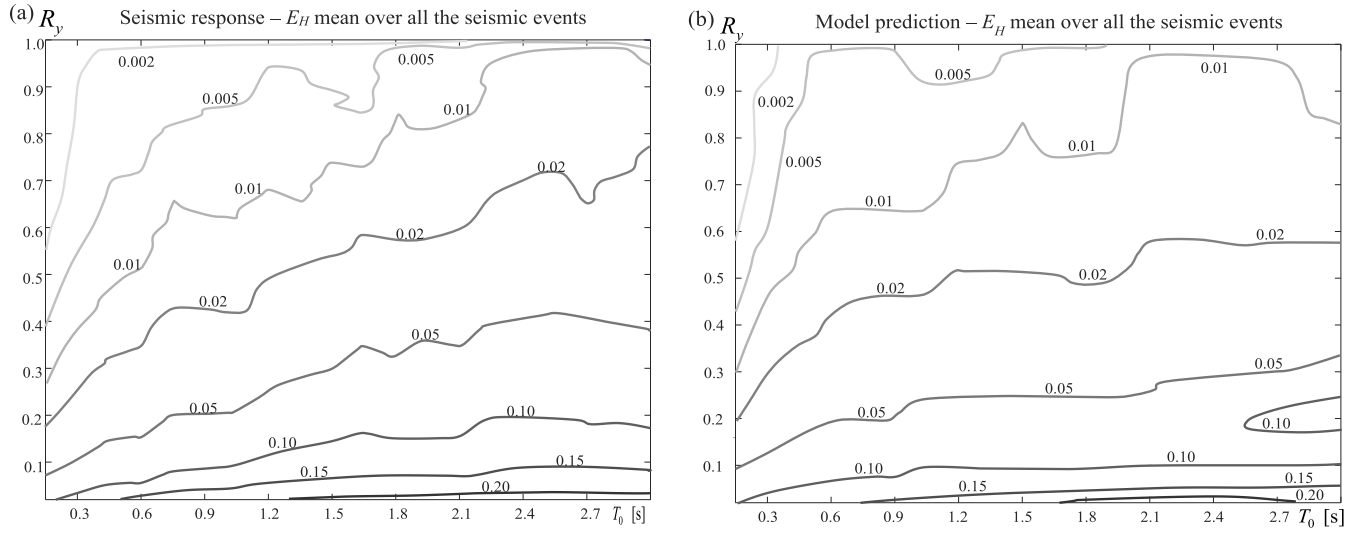


Fig. 35. Comparison between mean value of non-linear seismic response (E_H) obtained by (a) dynamic analysis and (b) ML model – contour maps.

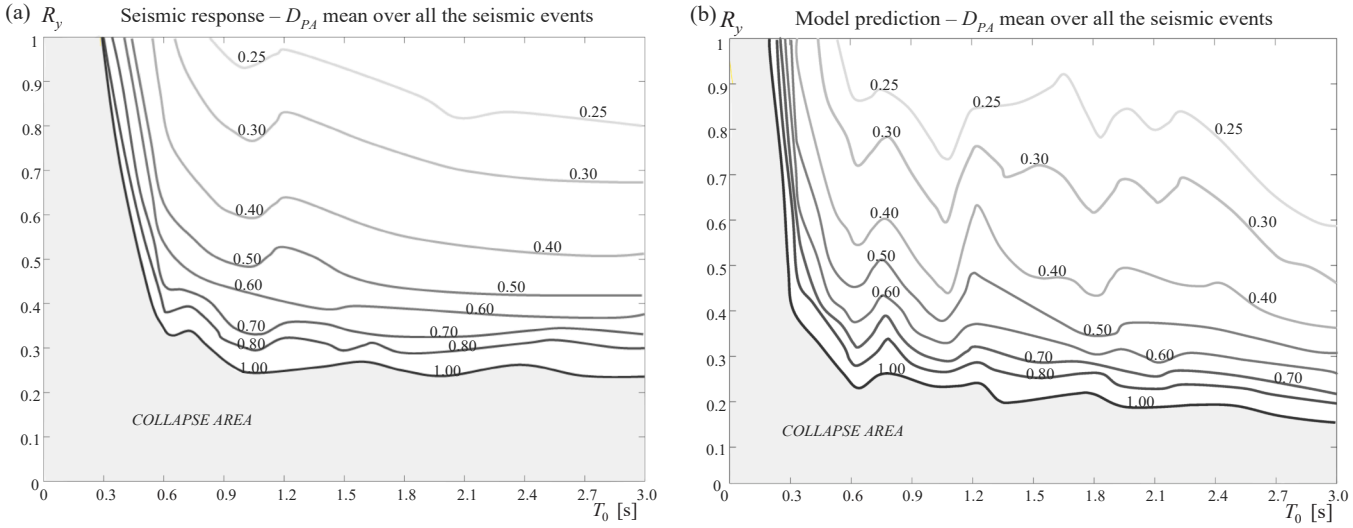


Fig. 36. Comparison between mean value of non-linear seismic response (D_{PA}) obtained by (a) dynamic analysis and (b) ML model – contour maps.

ML model can capture the relationship between input and output data. It predicts peak relative displacements (Fig. 31) with maximum errors within 15–25 % compared to the traditionally calculated seismic response. These errors are larger in the case of flexible structures with lower yield strength values, while they decrease for structural systems with higher stiffness and resistance. An even better level of prediction is observed in the case of hysteretic energy (Fig. 32), where the ML model reliably captures the effect of the yield strength factor, showing a tendency to underestimate the values by about 5 % for longer periods.

This observation also applies to the Park & Ang index in Fig. 33,

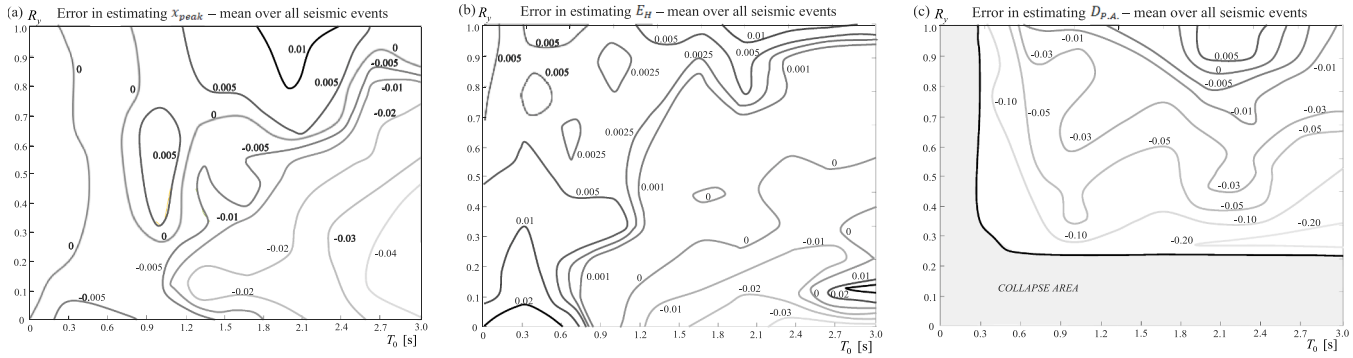


Fig. 37. Error committed by ML best model in estimating the peak displacement (a), hysteretic energy (b) and Park&Ang damage index (c).

where a comparison between the computed, $D_{P.A.,com}$, and estimated, $D_{P.A.,est}$, values of this index is carried out through a direct comparison of its trend as the reduction coefficient varies. In Fig. 33(b), the error in Park&Ang damage index estimation, $\Delta D_{P.A.} = D_{P.A.,com} - D_{P.A.,est}$, is shown. This graph highlights how the estimation error is limited. Moreover, a slight overestimation is observed for lower-period values, while higher-period values tend to be slightly underestimated. The case of $T_0 = 1.20s$ exhibits intermediate behavior, with an underestimation for lower values of the reduction coefficient and an overestimation of the damage level for $R_y > 0.35$. Similar results are obtained for the entire set of natural seismic excitations analyzed.

Figs. 34–36 illustrate contour maps representing the average values of peak relative displacement (Fig. 34), hysteretic energy (Fig. 35) and Park & Ang damage index (Fig. 36) calculated by (a) time integration and (b) Bagged Trees model with the T6 input parameter triplet, for all 1000 SDOF systems subjected to the 10 natural events. In addition, Fig. 37 shows the estimation error obtained using the same model for the three quantities considered. It is observed that the maps related to the peak relative displacement (Fig. 34) exhibit similar trends for structural systems with periods less than 1.5 s. Beyond this threshold, the ML numerical models tend to underestimate the peak relative displacement (average error less than 5 %). Regarding the hysteretic energy and Park & Ang damage index, the maps indicate that the considerations made concerning seismic excitation #8 can also be extended to the averages calculated across all seismic events. With respect to the dynamic analyses, the ML model slightly overestimated the hysteretic energy. A significant underestimation is only observable for very low resistance values and high periods (Fig. 37(b)). The collapse region ($D_{PA} > 1$) appears less extensive, particularly, in the range of longer periods. It is worth emphasizing that the errors in the damage index remain below 10 %, except for a very limited region of the $T_0 - R_y$ domain (Fig. 37(c)), thus enabling a rapid and effective assessment of this parameter for equivalent SDOF structural systems.

Similar results are obtained when considering the Bagged Trees model trained with other input combinations of the seismic parameters (i.e., T3, T4, T7, T8, C13). Conversely, less promising results are observed using other algorithms or when the training is conducted with less informative input seismic parameters combination, confirming the relevance of the proposals. It follows that the proposals represent a promising initial step to integrate ML techniques into the context of the performance-based seismic design. If the training phase is properly calibrated with respect to the site seismic hazard and seismic parameters are selected within the HPSP set, the model is capable to estimate the performance of an equivalent structural system, reducing significantly the computational effort and using simple modelling of the relationship between seismic responses and high predictive seismic parameters. This highlights the potential for enhancing the efficiency of the performance-based seismic design through a judicious incorporation of ML methodologies.

8. Conclusions

The use of supervised learning algorithms in the field of structural engineering, and, in general, the potential offered by ML, is a topic that has recently attracted the interest of the scientific community. While these techniques appear to be promising in terms of numerical results, they pose challenges due to being black-box models, compromising the physical understanding of the problem and the control of information necessary for a correct design approach.

This contribution presents the results of a comprehensive investigation with a dual objective: firstly, the identification and validation of a numerical technique based on the concept of the conditional information that allows to identify the best seismic parameters characterizing a seismic signal able to be the most useful for modeling seismic responses through techniques based on supervised learning algorithms; secondly, to study the ML algorithms that can be effectively used in the context of a performance-based seismic design procedure, leading to the definition of the ML model with the best capability to predict the non-linear responses of structural systems. These seismic parameters can be interpreted as the physical quantity with the most complete information to enable the model for a reliable prediction of the system response.

The results demonstrate that using seismic parameters, with a higher value of Shannon conditional information, allows supervised learning algorithms to generate models with better performance in terms of seismic response prediction. Furthermore, the family of algorithms called "Tree", specifically the "Bagged Tree", leads to models capable of making more reliable predictions of the non-linear dynamic response, in terms of both peak relative displacement and hysteretic energy, of a wide class of equivalent SDOF systems. The results also highlight that five out of the twenty-three considered seismic parameters, provided as single input data for the ML model, can be classified as HPSP for their superior quality of the performance predictions. It is also noteworthy that the HPSP set is composed of the seismic parameters involving velocity, displacement and parameters with an energetic nature. Instead, the use of seismic parameters directly related to acceleration or time is less effective, highlighting the low effectiveness of the seismic hazard definition if ML models will be considered in the context of performance-based design. Using a combination of multiple seismic parameters, as input data, significantly improves the predictive performance, reducing the model RMSE and MAE values by over 30 %. Pairs and triplets containing at least one element from the HPSP set demonstrate quite always a superior predictive performance.

Finally, the seismic response in terms of peak relative displacement, hysteretic energy and Park & Ang damage index, obtained from the time integration of the non-linear dynamic equations, is compared to the behavior predicted by the best ML models using one of the most predictive combinations of the seismic parameters (i.e., T6). This comparison was performed using natural ground motions properly scaled to match seismic parameter values similar to those of the artificial records used during the ML algorithm's training phase. The results demonstrate that the ML model can estimate the performance level of an equivalent SDOF system while significantly reducing computational effort by using energy-based parameters with high conditional informative content with respect to seismic response. This represents a promising approach for integrating ML models into performance-based design procedures regarding both the choice of the best seismic parameters in the definition of seismic hazard maps and safety assessment. In detail, the proposals can guide engineers to a better selection of one or more intensity measures for seismic reliability assessment of a specific structure or substitute non-linear dynamic simulations in case of classes of systems or for simplified multi-risk analysis.

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.

Acknowledgments

This study was carried out within the «Data fusion based digital twins for structural safety assessment» project – funded by European Union – Next Generation EU within the PRIN 2022 program (D.D. 104–02/02/2022 Ministero dell'Università e della Ricerca). This manuscript reflects only the authors' views and opinions and the Ministry cannot be considered responsible for them.

This study was also carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan—NRRP, Mission 4, Component 2, Investment 1.3—D.D. 1243 2/8/2022, PE0000005).

This work is part of the research activity developed by the authors within the framework of the “PNRR”: MOST – Sustainable Mobility National Research Center - SPOKE 7 “Coopera-tive Connected and Automated Mobility and Smart Infrastructures” - WP4" and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1033 17/06/2022).

Data availability

Data will be made available on request.

References

[1] Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27:379–423.
[2] Samuel AL. Some studies in machine learning using the game of checkers. II-Recent progress. Annu Rev Autom Program 1969;6:1–36.
[3] Martignon L. Information Theory. International Encyclopedia of the Social & Behavioral Sciences. Second Edition. Elsevier Inc; 2015. p. 106–9.
[4] J.D. Christensen Joel S., Accounting Theory: An Information Content Perspective (2002). McGraw-Hill/Irwin, 2003.
[5] Yazdani A, Salehi H, Shahidzadeh MS. A modified three-parameter lognormal distribution for seismic demand assessment considering collapse data. KSCE J Civ Eng 2018;22:204–12.
[6] Yazdani A, Nicknam A, Yousefi Dadras E, Eftekhari SN. Entropy-based sensitivity analysis of global seismic demand of concrete structures. Eng Struct 2017;146:118–26.
[7] I. Takahashi, T. Mukerji, G. Mavko, A strategy to select optimal seismic attributes for reservoir property estimation: Application of information theory. 1999 SEG Annual Meeting (Society of Exploration Geophysicists, 1999).
[8] Rundle JB, Giguere A, Turcotte DL, Crutchfield JP, Donnellan A. Global seismic nowcasting with shannon information entropy. Earth Space Sci 2019;6:191–7.
[9] Pasten D, Saravia G, Vogel EE, Posadas A. Information theory and earthquakes: depth propagation seismicity in northern Chile. Chaos, Solitons Fractals 2022;165.
[10] Tsioulou A, Galasso C. Information theory measures for the engineering validation of ground-motion simulations. Earthq Eng Struct Dyn 2018;47:1095–104.
[11] Ebrahimian H, Jalayer F, Lucchini A, Mollaioli F, Manfredi G. Preliminary ranking of alternative scalar and vector intensity measures of ground shaking. Bull Earthq Eng 2015;13:2805–40.
[12] Adeli H H. Neural networks in civil engineering: 1989-2000. Comput-Aided Civ Infrastruct Eng 2001;16(2):126–42.
[13] Kicinger R, Arciszewski T, De Jong K. Evolutionary computation and structural design: a survey of the state-of-the-art. Comput Struct 2005;83(23–24):1943–78.
[14] Saka MP, Geem ZW. Mathematical and metaheuristic applications in design optimization of steel frame structures: an extensive review. Math Probl Eng 2013:271031. (Article ID).
[15] Xie Y, Ebad Sichani M, Padgett JE, DesRoches R, R. The promise of implementing machine learning in earthquake engineering: a state-of-the-art review. Earthq Spectra 2020;36(4):1769–801.
[16] De Lautour OR, Omenzetter P. Prediction of seismic-induced structural damage using artificial neural networks. Eng Struct 2009;31(2):600–6.
[17] Oh BK, Park Y, Park HS. Seismic response prediction method for building structures using convolutional neural network. Struct Control Health Monit 2020;27(5):e2519.
[18] Gharehbaghi S, Yazdani H, Khatibinia M. Estimating inelastic seismic response of reinforced concrete frame structures using a wavelet support vector machine and an artificial neural network. Neural Comput Appl 2020;32:2975–88.
[19] Hait P, Sil A, Choudhury S. Seismic damage assessment and prediction using artificial neural network of RC building considering irregularities. J Struct Integr Maint 2020;5(1):51–69.
[20] Kim T, Kwon OS, Song J. Deep learning based seismic response prediction of hysteretic systems having degradation and pinching. Earthq Eng Struct Dyn 2023;52(8):2384–406.
[21] Liao Y, Tang H, Li R, Ran L, Xie L. Response Prediction for Linear and Nonlinear Structures Based on Data-Driven Deep Learning. Appl Sci 2023;13(10):5918.
[22] Hareendran SP, Alipour A. Prediction of nonlinear structural response under wind load using deep learning techniques. Appl Soft Comput 2022;129:109424.
[23] De Iuliis M, Miceli E, Castaldo P. Machine learning modelling of structural response for different seismic signal characteristics: a parametric analysis. Appl Soft Comput J 2024;164:112026. https://doi.org/10.1016/j.asoc.2024.112026.
[24] Won J, J, Shin J. Machine learning-based approach for seismic damage prediction method of building structures considering soil-structure interaction. Sustainability 2021;13(8):4334.
[25] Asgarkhani N, Kazemi F, Jakubczyk-Gałczyńska A, Mohebi B, Jankowski R. Seismic response and performance prediction of steel buckling-restrained braced frames using machine-learning methods. Eng Appl Artif Intell 2024;128:107388.
[26] Shahnazaryan D, O'Reilly GJ. Next-generation non-linear and collapse prediction models for short-to long-period systems via machine learning methods. Eng Struct 2024;306:117801.
[27] Payán-Serrano O, Bojórquez E, Carrillo J, Bojórquez J, Leyva H, Rodríguez-Castellanos A, Carvajal J, Torres J. Seismic Performance Prediction of RC. BRB SDOF Struct Using Deep Learn Intensity Meas INp AI 2024;5(3):1496–516.
[28] Nguyen HD, Dao ND, Shin M. Machine learning-based prediction for maximum displacement of seismic isolation systems. J Build Eng 2022;51:104251.
[29] Abdellatif B, Benazouz C, Ahmed M, Abdellatif B, Benazouz C, Ahmed M. Dynamic response estimation of an equivalent single degree of freedom system using artificial neural network and nonlinear static procedure. Res Eng Struct Mater 2023. https://doi.org/10.17515/resm2023.40me0818rs.
[30] Hammal S, Bourahla N, Laouami N. Neural-network based prediction of inelastic response spectra. Civ Eng J 2020;6(6):1124–35.
[31] Dwairi HM, Tarawneh AN. Artificial neural networks prediction of inelastic displacement demands for structures built on soft soils. Innov Infrastruct Solut 2022;7(1):4.
[32] Yinfeng D, Yingmin L, Ming L, Mingkui X. Nonlinear structural response prediction based on support vector machines. J Sound Vib 2008;311(3-5):886–97.
[33] Gharehbaghi S, Gandomi M, Plevris V, Gandomi AH. Prediction of seismic damage spectra using computational intelligence methods. Comp Struct 2021;253:106584.
[34] Demir A, Sahin EK, Demir S. Advanced tree-based machine learning methods for predicting the seismic response of regular and irregular RC frames. Structures, 64. Elsevier; 2024, 106524.
[35] 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.
[36] L. Luzi, R. Puglia, E. Russo & ORFEUS WG5, Engineering Strong Motion Database, version 1.0. *Istituto Nazionale di Geofisica e Vulcanologia, Observatories & Research Facilities for European Seismology* 2016.
[37] Mousavi SM, Sheng Y, Y, Zhu W, Beroza G.C GC. STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. IEEE Access 2019.
[38] Bakhshi H, Bagheri A, Ghodrati Amiri G, Barkhordari MA. Estimation of spectral acceleration based on neural networks. Proc Inst Civ Eng - Struct Build 2014;167(8):457–68.
[39] Khosravikia F, Clayton P, Nagy Z. Artificial neural network-based framework for developing ground-motion models for natural and induced earthquakes in Oklahoma, Kansas, and Texas. Seismol Res Lett 2019;90(2A):604–13.
[40] Cabalar AF AF, Cevik A. Genetic programming-based attenuation relationship: An application of recent earthquakes in turkey. Comp Geosci 2009;35(9):1884–96.
[41] Hamze-Ziabari SM, Bakhshpoori T T. Improving the prediction of ground motion parameters based on an efficient bagging ensemble model of M5# and CART algorithms. Appl Soft Comput J 2018;68:147–61.
[42] Akhani M, Kashani AR, Mousavi M, Gandomi AH. A hybrid computational intelligence approach to predict spectral acceleration. Measurement 2019;138:578–89.
[43] Kwon OS, Elnashai A. The effect of material and ground motion uncertainty on the seismic vulnerability curves of RC structure. Eng Struct 2006;28(2):289–303.
[44] Kazemi F, Asgarkhani N, Jankowski R. Machine learning-based seismic fragility and seismic vulnerability assessment of reinforced concrete structures. Soil Dyn Earthq Eng 2023;166(107761).
[45] Massumi A, Gholami F. The influence of seismic intensity parameters on structural damage of RC buildings using principal components analysis. Appl Math Model 2016;40(3):2161–76.
[46] Demertzis K, Kostinakis K, Morfidis K, Iliadis L. An interpretable machine learning method for the prediction of R/C buildings' seismic response. J Build Eng 2023;63.
[47] Hartmann CRP, Varshney PK, Mehrotra KG, Gerberich CL. Application of Information Theory to the Construction of Efficient Decision Trees. IEEE Trans Inf Theory 1982;28:565–77.
[48] Eurocode 8: “Design of structures for earthquake resistance, Part 1: General rules, seismic actions and rules for buildings, EN 1998-1: 2004”, CEN – European Committee for Standardization, Brussels, November 2004.
[49] Papageorgiou AS, Aki K. A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion. Part II. Appl Model *Bull Seismol Soc Am* 1983;73:953–78.
[50] Papageorgiou AS, Aki K. Sealing law of far-field spectra based on observed parameters of the specific barrier model. Pure Appl Geophys PAGEOPH 1985;123:353–74.
[51] W.B. Joyner, D.M. Boore, Measurement, characterization, and prediction of strong ground motion. In Earthquake Engineering and Soil Dynamics II, Proc. Am. Soc. Civil Eng. Geotech. Eng. Div. Specialty Conf., June 27–30, 1988, Park City, Utah (1988), pp. 43–102.
[52] Seismosoft. SeismoArtif - A computer program for generation of artificial accelerograms, [online]. 2020. Available at: ⟨https://seismosoft.com/⟩.
[53] Halldorsson B, Papageorgiou AS. Calibration of the specific barrier model to earthquakes of different tectonic regions. Bull Seismol Soc Am 2005;95:1276–300.
[54] Seismosoft. SEISMOAPPS – Technical Information Sheet, [online]. 2023. Available at: ⟨https://seismosoft.com/wp-content/uploads/prods/lib/SEISMOAPPS-Technical-Information-Sheet-ENG.pdf⟩.
[55] Rodolfo Saragoni G, Hart GC. Simulation of artificial earthquakes. Earthq Eng Struct Dyn 1973;2:249–67.
[56] Antikaev FF, Shebalin NV. Revision of Correlations between Level of Macroseismic Effect and Dynamic Parameters of Ground Movement. *Journal Engineering Seismology: Investigations on Seismic Hazard*. Nauka 1988;29:98–108.
[57] Bommer JJ, Pinho R. Adapting earthquake actions in Eurocode 8 for performance-based seismic design. Earthq Eng Struct Dyn 2006;35(1):39–55.
[58] Seismosoft, “SeismoSignal - A computer program for signal processing of time-histories”, 2016.

[59] Arias A. A measure of earthquake intensity. Seismic Design for Nuclear Power. Plants. Cambridge, MA: MIT Press; 1970. p. 438–69.
[60] Park YJ, Ang AH-S, Wen YK. Damage-limiting seismic design of buildings. Earthq Spectra 1987;3(1):1–26.
[61] John SA, Yuri R. Assessment of seismic input energy by means of new definition and the application to earthquake resistant design. Archit Eng 2016;1(4):26–35.
[62] Benjamin JR. A criterion for determining exceedance of the operating basis earthquake. Report No. EPRI NP-5930. USA.: Electrical Power Research Institute, Palo Alto, California; 1988.
[63] G.W. Housner, Spectrum intensities of strong-motion earthquakes. Proceedings of the Symposium on Earthquake and Blast Effects on Structures. EERI, Berkeley, CA, 1952.
[64] V. Thun, J. Lawrence, Earthquake ground motions for design and analysis of dams. Earthquake Engineering and Soil Dynamics II-Recent Advances in Ground-Motion Evaluation, 1988.
[65] O.W. Nuttli, The relation of sustained maximum ground acceleration and velocity to earthquake intensity and magnitude, Report 16, Misc. Paper S-73-1, 1979, US Army Waterways Experimental Station, Vicksburg, Mississippi, USA.
[66] R.P. Kennedy, S.A. Short, K.L. Merz, F.J. Tokarz, I.M. Idriss, M.S. Power, K. Sadigh, Engineering characterization of ground motion. Task I. Effects of characteristics of free-field motion on structural response (No. NUREG/CR-3805). Structural Mechanics Associates, Inc., Newport Beach, CA (USA); 1984 Woodward-Clyde Consultants, Walnut Creek, CA (USA).
[67] Kramer SL. Geotechnical Earthquake Engineering. Upper Saddle River, New Jersey, USA: Prentice Hall. Inc; 1996.
[68] Garg R, Vemuri JP, Subramaniam KVL. Correlating peak ground A/V Ratio with ground motion frequency content. Recent Advances in Structural Engineering, 2. Singapore: Publisher: Springer; 2019.
[69] Bommer JJ, Martínez-Pereira A. The effective duration of earthquake strong motion. J Earthq Eng 1999;3(2):127–72.
[70] Shibata A, Sozen M. Substitute structure method for seismic design in reinforced concrete, *Journal of Structural Division*. ASCE 1976;102(1):1–18.
[71] Applied Technology Council (ATC), Seismic evaluation and retrofit of concrete buildings, Report ATC-40, Applied Technology Council, 1996 Redwood City, CA, USA.
[72] Chopra AK, Goel RK. A modal pushover analysis procedure for estimating seismic demands for buildings. Earthq Eng Struct Dyn 2002;31(3):561–82.
[73] Proietti G, Pedone L, D'Amore S, Pampanin S. Inelastic response spectra for an integrated displacement and energy-based seismic design (DEBD) of structures. Front Built Environ 2023;9:1264033.
[74] Chopra AK. Dynamic of Structures: theory and applications to earthquake engineering. 5th edition. Pearson; 2017.
[75] De Iuliis M, Miceli E, Castaldo P. An energy framework to control viscoelastic semi-active devices in plan-wise one-way asymmetric systems. Struct Control Health Monit 2025. https://doi.org/10.1155/stc/7091316.
[76] Freedman D, Diaconis P. On the histogram as a density estimator: L2 theory. Z FüR Wahrscheinlichkeitstheorie und Verwandte– Geb 1981;57(4):453–76.
[77] Wyner AD. A definition of conditional mutual information for arbitrary ensembles. Inf Control 1978;38(1):51–9.
[78] James G, Witten D, Hastie T, Tibshirani, R R. An introduction to statistical learning. New York: Springer; 2013.
[79] Sheppard C. Tree-based machine learning algorithms: Decision trees, random forests, and boosting. Clint Sheppard 2019.
[80] Hamel L. Knowledge discovery with support vector machines. In: Larose Daniel T, editor. Wiley Series on Methods and Applications in Data Mining. Series Editor; 2009.
[81] Hastie T, Tibshirani R, Friedman J. The elements of Statistical learning. Data Mining, inference and prediction. Springer series in Statistics. Springer; 2009.
[82] Bengio Y, Courville A, Goodfellow IJ. Deep learning: adaptive computation and machine learning. Series: Adaptive Computation and Machine Learning series. MIT Press; 2016.
[83] Arroyo D, Ordaz M. On the estimation of hysteretic energy demands for SDOF systems. Earthq Eng Struct Dyn 2007;36(15):2365–82.
[84] Baker JW. Conditional mean spectrum: Tool for ground motion selection. J Struct Eng – ASCE 2011;137:322–31.
[85] Pacific Earthquake Engineering Research Center (2005) PEER Strong Motion Database on Line. Berkley, ⟨https://peer.berkeley.edu/peer-strong-ground-motion-databases⟩.
[86] Castaldo P, Miceli E. Optimal single concave sliding device properties for isolated multi-span continuous deck bridges depending on the ground motion characteristics. Soil Dyn Earthq Eng 2023;173:108128.
[87] Park YJ, Ang AHS. Mechanistic seismic damage model for reinforced concrete. J Struct Eng 1985;111(4):722–39.