National Technical University of Athens

School of Chemical Engineering

Computational Mechanics

A machine learning approach to earthquake response analysis of structural systems

Georgakis Panagiotis

Supervisor:

Vissarion Papadopoulos

Associate Professor, School of Civil Engineering, NTUA

ATHENS, October 2021

Acknowledgments

Completing the cycle of my postgraduate studies I would like to thank those who contributed to the completion of this dissertation.

I would like to thank my supervising Professor Vissarion Papadopoulos and Associate Professor Michalis Fragiadakis for their support, guidance and inspiration.

Specifically, I would like to express my deepest gratitude to Dr. Manolis Georgioudakis for the countless hours of collaboration, mentorship and for all the fruitful conversations and laughs during the elaboration of the present thesis. It was very kind from him to share his knowledge and experience with me and it has been truly a privilege working with him.

I would also like to sincerely thank Mr. Spyros Diamantopoulos for his endless support and contribution throughout all these years sharing.

Finally, I warmly thank all my professors for all their work through this postgraduate program and also my family and friends for all the patience and moral support.

Contents

Περίληψη

Η μη γραμμική ανάλυση απόκρισης (NRHA) είναι το πιο ρεαλιστικό εργαλείο αξιολόγησης σεισμικής απόκρισης κατασκευών που απαιτεί τη χρήση χρονοϊστοριών επιτάχυνσης ως είσοδο σε αριθμητικές προσομοιώσεις. Οι τεχνικές μηχανικής μάθησης κερδίζουν συνεχώς αυξανόμενο ενδιαφέρον στους τομείς της μηχανικής και μπορούν να αποτελέσουν ένα πολλά υποσχόμενο εργαλείο για αξιόπιστες προβλέψεις με την ικανότητά τους να εντοπίζουν γρήγορα και με ακρίβεια τάσεις ή μοτίβα μέσω πειραματικών ή τεχνητά δημιουργημένων δεδομένων. Σε αυτή τη μελέτη, προτείνεται μια διαδικασία μηχανικής μάθησης για την εκτίμηση της μη-γραμμικής απόκρισης μονοβάθμιων συστημάτων ως προς τη μέγιστη μετατόπισή τους. Η εφαρμοσιμότητα και η αποτελεσματικότητα της προτεινόμενης προσέγγισης αποδεικνύεται σε ένα μονοβάθμιο δομικό σύστημα με την αξιολόγηση της απόδοσης διαφορετικών μοντέλων μηχανικής μάθησης. Αποδεικνύεται ότι ελήφθησαν επαρκείς προβλέψεις μέσω της διαδικασίας επικύρωσης που μπορεί να λειτουργήσει ως εργαλείο αναφοράς και μπορεί να επεκταθεί για την εκτίμηση της απόκρισης πολυβάθμιων συστημάτων.

Abstract

Nonlinear response history analysis (NRHA) is the most realistic seismic performance assessment tool of structures which requires the use of recorded acceleration time-histories as input into numerical simulations. Machine learning (ML) techniques are constantly gaining increasing interest in engineering fields and can consist a promising tool for reliable predictions with their ability to quickly and accurately identify trends or patterns through experimental or artificially generated data. In this study, a machine-learning pipeline is proposed to estimate the nonlinear response analysis of single degree-of-freedom systems in terms of their maximum displacement. The applicability and efficiency of the proposed approach is demonstrated in a single degree-of-freedom (DOF) structural system by evaluating the performance of different ML models. It is shown that adequate predictions were obtained through the validation process which can act as a reference tool that can be extended to estimate the structural response of multi-DOF, as well.

List of Figures

List of Tables

1. Introduction

Nonlinear response history analysis (NRHA) is the most realistic seismic performance assessment tool of structures which requires the use of recorded acceleration time-histories as input into numerical simulations. Evidently, it can be computationally intensive when it is applied to multi degree-of-freedom (DOF) structural systems, particularly when dealing with parametric studies or/and incremental dynamic analyses. Various computational techniques have been proposed in the literature to reduce the computational cost of such analyses e.g. record simplification (Faroughi and Hosseini, 2011), record down-sampling, modified inverse Fourier transform, to name a few. Nowadays, machine learning (ML) techniques are constantly gaining increasing interest in engineering fields (Salehia and Burgueñoa, 2018) and can consist a promising tool for reliable predictions with their ability to quickly and accurately identify trends or patterns through experimental or artificially generated data.

In this study, we propose a robust machine learning pipeline to estimate the nonlinear response analysis of multi-DOF systems in terms of their maximum displacement/ductility aiming to eliminate the computational cost of the NRHA analysis. A pulse extraction process is used (Mavroeidis and Papageorgiou, 2003) to quantify the wavelet parameters of a ground motion records which, along with the material parameters of the structural system consist an adequate training data set. The applicability and efficiency of the proposed approach is demonstrated first in a single-DOF oscillator that is used as a benchmark example. A comparison study between various ML regression models, is followed by, in order to choose the efficient ML technique that will be used for the training process. It is shown that adequate predictions were obtained through the validation of various single-DOF benchmark structures which can act as a reference tool for an engineer in practice.

The remaining thesis is organized as follows. Chapter 2 provides a literature review of the nonlinear response analysis method that is will be used in this study. Furthermore, details about the central difference method are also provided and the bilinear hysteretic model is defined. Chapter 3 summarizes the method that a ground motion record can be represented as a sum of pulse-like wavelets which is called pulse extraction. Chapter 4 proposes a machine learning model that predicts the max displacement, evaluates different regression algorithms and investigates the impact of the input parameters in the models predictions. Finally, in Chapter 5 the extension to multi-DOF structural systems is applied using the models trained in Chapter 4.

2. Dynamic response analysis of structures

In this opening chapter, the structural dynamics problem is formulated for simple structures that can be idealized as a system with a lumped mass and a massless supporting structure. Linear systems as well as inelastic structures subjected to applied dynamic force or earthquake-induced ground motion are considered. Then a method for solving the differential equation governing the motion of the structure is presented. The chapter ends with a brief definition of the bilinear hysteretic model.

2.1. Single-DOF systems

For a linear system, the principal problem of structural dynamics that concerns structural engineers is the behavior of structures subjected to earthquake-induced motion of the base of the structure. The displacement of the ground is denoted by u_g, the total (or absolute) displacement of the mass by u^t, and the relative displacement between the mass and ground by u (Figure 1). At each instant of time these displacements are related by

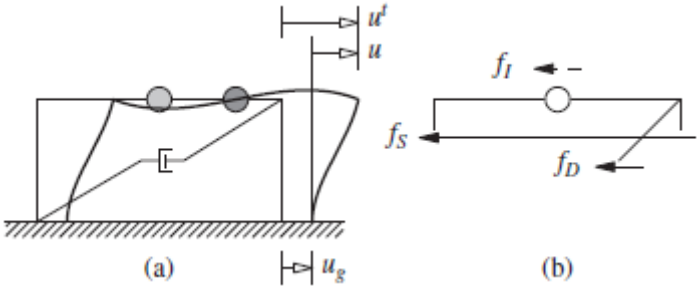


Figure 1: Single degree-of-Freedom System (Chopra A.K., 2011).

Both u^t and u_g refer to the same inertial frame of reference and their positive directions coincide.

From the free-body diagram including the inertia force f_I, shown in Figure 1b, the equation of dynamic equilibrium is

$$f_I + f_D + f_S = 0 \tag{2.1}$$

$$m\ddot{u}^t(t) + c\dot{u}(t) + ku(t) = 0 \tag{2.2}$$

where

$$u^t(t) = u_g(t) + u(t) \tag{2.3}$$

Only the relative motion u between the mass and the base due to structural deformation produces elastic and damping forces (i.e., the rigid-body component of the displacement of the structure produces no internal forces).

The relation that describes the equation of motion for each time moment t of the above system is:

$$\mathrm{m\ddot{u}}(t) + \mathrm{c\dot{u}}(t) + \mathrm{ku}(t) = -\mathrm{m\ddot{u}_g(t)} \tag{2.4}$$

where m, c and k are mass, damping and stiffness.

In the above relation what changes depending on the model to be used is the definition of the term ku(t).

2.2. Multi-DOF systems

The equation of motion of a dynamic problem is formulated by summing the elastic forces $\mathbf{F_E}$ of the equilibrium equation for the static problem with the inertia forces $\mathbf{F_I}$ and damping $\mathbf{F_D}$, so that at any given time this sum is equal to the externally imposed forces **P**:

$$\mathbf{F}_I(t) + \mathbf{F}_D(t) + \mathbf{F}_E(t) = \mathbf{P}(t) \tag{2.5}$$

By denoting **u**, $\dot{\mathbf{u}}$ and $\ddot{\mathbf{u}}$ the displacement, velocity and acceleration vectors respectively, the motion equation becomes:

$$\mathbf{M\ddot{u}}(t) + \mathbf{C\dot{u}}(t) + \mathbf{Ku}(t) = \mathbf{P}(t) \tag{2.6}$$

where **M**, **C** and **K** are the matrices of mass, damping and stiffness. In case the construction is seismically stimulated with a time history of accelerations $\ddot{\mathbf{u}}_g$**(t)** at its base, then the externally imposed loads are proportional to the mass and calculated as:

$$\mathbf{P}(t) = -\mathbf{Mr\ddot{u}_g}(t) \tag{2.7}$$

where **r** is the direction vector of seismic excitation, with its elements to are equal to 1 if the degree of freedom is in the same direction as that of the earthquake and with 0 in a different case.

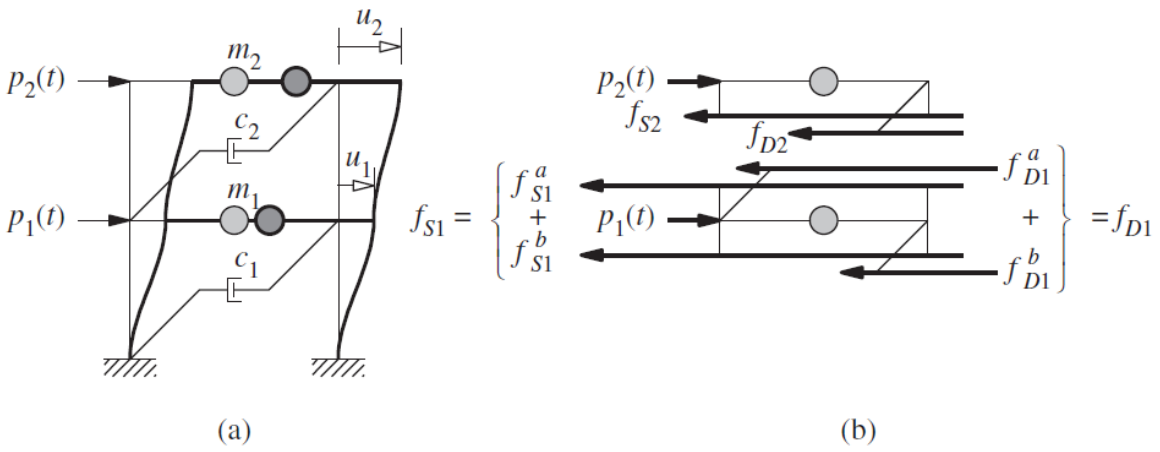


Figure 2: (a) Two-story shear frame; (b) forces acting on the two masses.

2.3. Solution of the dynamic system

There are two main types of methods for calculating the dynamics of a response of a construction, the numerical and the classical. The numerical solutions are divided into two categories, the Direct Integration methods of the motion equation and the Modal Methods. Direct Integration methods are divided into explicit methods (e.g Central Difference Method) and implicit methods (e.g Newmark Method). Out of these two, the central difference method will be adopted in this study.

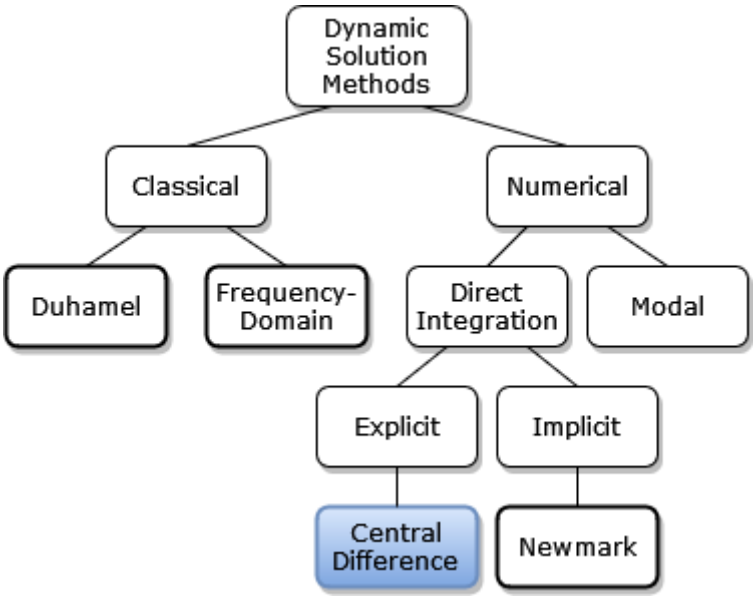


Figure 3: Dynamic solution methods tree chart.

2.3.1. Linear systems

The central differences method is an explicit method and is based on a finite differences approximation of the derivatives with respect to the time of displacement, i.e. of velocity and acceleration.

The values of the vector **u**(t) at the times **u**(t + Δt) and **u**(t - Δt) can be approximated using the Taylor formula:

$$\mathbf{u}(\mathrm{t}+\Delta\mathrm{t}) = \mathbf{u}(\mathrm{t}) + \Delta t\dot{\boldsymbol{u}}(t) + \frac{1}{2}\Delta t^2\ddot{\mathbf{u}}(t) + \frac{1}{6}\Delta t^3\dddot{\mathbf{u}}(t) + \cdots \tag{2.8}$$

$$\mathbf{u}(\mathrm{t}-\Delta\mathrm{t}) = \mathbf{u}(\mathrm{t}) - \Delta t\dot{\boldsymbol{u}}(t) + \frac{1}{2}\Delta t^2\ddot{\mathbf{u}}(t) - \frac{1}{6}\Delta t^3\dddot{\mathbf{u}}(t) + \cdots \tag{2.9}$$

By subtracting the equations (2.5) and (2.6) we have:

If Δt is small, terms with factors Δt^3, Δt^5, ... can be omitted from relation (2.7) and its solution

$$\mathbf{u}(\mathrm{t}+\Delta\mathrm{t})-\mathbf{u}(\mathrm{t}-\Delta\mathrm{t})=2\Delta t\dot{\boldsymbol{u}}(t)+\frac{2}{6}\Delta t^3\ddot{\mathbf{u}}(t)+\cdots \tag{2.10}$$

with respect to $\dot{\boldsymbol{u}}(t)$ gives us the approximation of the first derivative at the time ***t:***

$$\dot{\boldsymbol{u}}(t)\approx\frac{\mathbf{u}(\mathrm{t}+\Delta\mathrm{t})-\mathbf{u}(\mathrm{t}-\Delta\mathrm{t})}{\mathbf{2}\Delta\mathrm{t}} \tag{2.11}$$

Then, adding up the relations (2.5) and (2.6) and omitting the with factors Δt^4, Δt^6,... its solution with respect to $\ddot{\mathbf{u}}(t)$ gives us the approximation of the second derivative at the time ***t:***

$$\ddot{\mathbf{u}}(t)\approx\frac{\mathbf{u}(\mathrm{t}+\Delta\mathrm{t})-2\mathbf{u}(t)-\mathbf{u}(\mathrm{t}-\Delta\mathrm{t})}{\Delta\mathrm{t}^2} \tag{2.12}$$

Using the symbols $\mathbf{u}(t) = \mathbf{u}_i$, $\mathbf{u}(t + \Delta t) = \mathbf{u}_{i+1}$ and $\mathbf{u}(t - \Delta t) = \mathbf{u}_{i-1}$, we can write the relations (2.8) and (2.9) as:

$$\dot{\boldsymbol{u}}(t)\approx\frac{\mathbf{u}_{i+1}-\mathbf{u}_{i-1}}{\mathbf{2}\Delta\mathrm{t}} \tag{2.13}$$

$$\ddot{\mathbf{u}}(t)\approx\frac{\mathbf{u}_{i+1}-2\,\mathbf{u}_i+\mathbf{u}_{i-1}}{\Delta t^2} \tag{2.14}$$

which are the approximations of the derivatives of the vector $\mathbf{u}_i$ with the central differences and after replacing them in the equation of motion at time t_i (Ex. 2.2) we receive:

$$\left(\frac{1}{\Delta t2}\,\mathbf{M}+\frac{1}{2\Delta t}\mathbf{C}\right)\mathbf{u}_{i+1}=\mathbf{P}_i-\left(\mathbf{K}-\frac{2}{\Delta t^2}\mathbf{M}\right)\mathbf{u}_i-\left(\frac{1}{\Delta t^2}\,\mathbf{M}-\frac{1}{2\Delta t}\mathbf{C}\right)\mathbf{u}_{i-1} \tag{2.15}$$

or

$$\widehat{\mathbf{K}}\mathbf{u}_{i+1}=\widehat{\mathbf{P}}_i \tag{2.16}$$

where:

$$\widehat{\mathbf{K}}=\frac{1}{\Delta t^2}\,\mathbf{M}+\frac{1}{2\Delta t}\mathbf{C} \tag{2.17}$$

and:

$$\widehat{\mathbf{P}}_i=\mathbf{P}_i-\left(\mathbf{K}-\frac{2}{\Delta t^2}\mathbf{M}\right)\mathbf{u}_i-\left(\frac{1}{\Delta t^2}\,\mathbf{M}-\frac{1}{2\Delta t}\mathbf{C}\right)\mathbf{u}_{i-1} \tag{2.18}$$

The unknown vector $\mathbf{u}_{i+1}$ at time t_{i+1} is calculated from equation 2.10, that is, from the state of equilibrium at time t_i (Eq. 2.2), without its use equilibrium state at time t_{i+1}. Hence the method of central differences is explicit.

In the first step of the iterative process, i.e to determine the vector $\mathbf{u}_1$, setting i = 0 in Ex. (2.15), we observe that the vector $\mathbf{u}_{-1}$ is required. To determine the vector $\mathbf{u}_{-1}$, we set i = 0 to equations (2.10) and (2.11) and we get:

$$\dot{\boldsymbol{u}}(t) \approx \frac{\mathbf{u}_{i+1} - \mathbf{u}_{-1}}{\mathbf{2}\Delta \mathrm{t}} \tag{2.19}$$

$$\ddot{\mathbf{u}}(t) \approx \frac{\mathbf{u}_{i+1} - 2\,\mathbf{u}_i + \mathbf{u}_{-1}}{\Delta t^2} \tag{2.20}$$

Solving Eq. (2.16) with respect to $\mathbf{u}_1$ and replacing it in Eq. (2.17) we have:

$$\mathbf{u}_{-1} \approx \mathbf{u}_0 - \Delta t\dot{\mathbf{u}}_0 + \frac{\Delta t^2}{2}\ddot{\mathbf{u}}_0 \tag{2.21}$$

The initial displacement vectors $\mathbf{u}_0$ and initial velocities $\dot{\mathbf{u}}_0$ are given, while from equation of motion at time $t_0 = 0$

$$\mathbf{M}\ddot{\mathbf{u}}_0 + \mathbf{C}\dot{\mathbf{u}}_0 + \mathbf{K}\mathbf{u}_0 = \mathbf{P}_0 \tag{2.22}$$

the resulting initial accelerations vector

$$\ddot{\mathbf{u}}_0 = \mathbf{M}^{-1}(\mathbf{P}_0 - \mathbf{C}\dot{\mathbf{u}}_0 + \mathbf{K}\mathbf{u}_0) \tag{2.23}$$

2.3.2. For nonlinear systems

The central difference method can be easily modified and applied to non-linear systems. Having the approaches of its derivatives vector $\mathbf{u}_i$ with the central differences, we replace them in the motion equation for at time t_i as formulated for nonlinear systems (Eq. 2.2) and we get:

$$\left(\frac{1}{\Delta t2}\,\mathbf{M}\, + \frac{1}{2\Delta t}\mathbf{C}\right)\mathbf{u}_{i+1} = \mathbf{P}_i + \frac{2}{\Delta t^2}\,\mathbf{M}\mathbf{u}_i - \left(\frac{1}{\Delta t^2}\,\mathbf{M}\, - \frac{1}{2\Delta t}\mathbf{C}\right)\mathbf{u}_{i-1} - \mathbf{F}_{s,\,i} \tag{2.24}$$

or:

$$\widehat{\mathbf{K}}\mathbf{u}_{i+1} = \widehat{\mathbf{P}}_i \tag{2.25}$$

where:

$$\widehat{\mathbf{K}} = \frac{1}{\Delta t2}\,\mathbf{M}\, + \frac{1}{2\Delta t}\mathbf{C} \tag{2.26}$$

and:

$$\widehat{\mathbf{P}}_i = \mathbf{P}_i - \left(\frac{1}{\Delta t^2}\,\mathbf{M}\, - \frac{1}{2\Delta t}\mathbf{C}\right)\mathbf{u}_{i-1} + \frac{2}{\Delta t^2}\,\mathbf{M}\mathbf{u}_i - \mathbf{F}_{s,\,i} \tag{2.27}$$

The above equations, if compared to those for linear systems, differ only in the definition of equivalent load $\widehat{\mathbf{P}}$. The resistance forces $\mathbf{F}_{s,\,i}$, appear explicitly, since they depend only on the response at time t_i and not from the unknown response at time t_{i+1}.

A python code snippet of the algorithm follows below:

```
for i in range(1, npts):
   if hardening != 1:
       fsp = fs[i-1]
       up = u[i-1]
       uincr = u[i]
       fs[i] = bilinear(k, hardening, uy, up, fsp, uincr)
   else:
       fs[i] = k*u[i]
   pp[i] = p[i] - a * uminus1[i] + b*u[i] - fs[i]
   uplus1[i] = pp[i] / kk
   if i < npts-1:
       uminus1[i + 1] = u[i]
       u[i + 1] = uplus1[i]
       udot[i + 1] = (uplus1[i] - uminus1[i])/(2*dt)
       udotdot[i + 1] = (uplus1[i] - 2 * u[i] + uminus1[i])/(dt**2)

   v = udot[:]
   a = udotdot
   umax = max(abs(u))
   vmax = max(abs(v))
   amaxRel = max(abs(a))
   fsmax = max(abs(fs))
```

2.3.3. Pushover analysis

The pushover analysis may be used to verify the structural performance of newly designed buildings and of existing buildings. It consists of applying monotonically increasing constant shape lateral load distributions to the structure under consideration. The structure model can be either 2D or 3D. In particular, EC8 states that for buildings with plan regularity, 2D analysis of single plane frames can be performed, while for buildings with plan irregularity a complete 3D model is necessary.

The N2 method was developed using a shear building model, i.e. a frame model with floors rigid in their planes. Furthermore, vertical displacements are typically neglected in the method and only the two horizontal ground motion components, x and y, are considered. Extension to the general case of a fully deformable frame is straightforward. The N2 method consists of applying two load distributions to the frame:

- A modal pattern, that is a load shape proportional to the mass matrix multiplied by the first elastic mode shape,

$$P^1 = M\varphi_1$$

- A uniform pattern, that is a mass proportional load shape,

$$P^2 = MR$$

where M is the mass matrix, φ1 is the first mode shape and R a vector of 1s corresponding to the degrees of freedom parallel to the application of the ground motion and 0s for all other dofs. In the N2 method φ1 is normalized so that the top floor displacement is 1, i.e. $\varphi_{1,n}=1$. The two load distributions are schematically shown in Figure 4. The applied lateral load distributions are increased and the response is plotted in terms of base shear V_b vs. top floor displacement D (for example center of mass of the top floor). This is the so-called pushover curve or capacity curve.

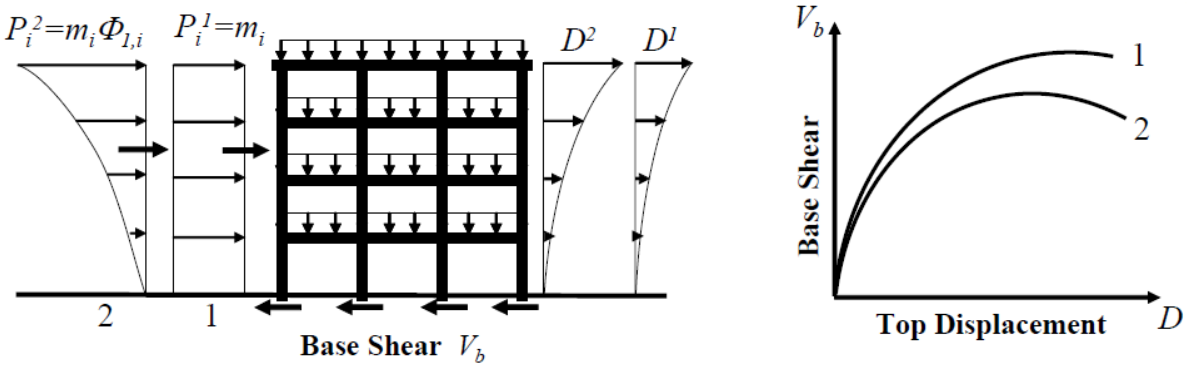


Figure 4: Load distribution for pushover analysis according to EC8 and pushover response curve (ZSoilr.PC 070202 report).

The N2 procedure transforms the response of the MDOF system into the response of an equivalent SDOF system.

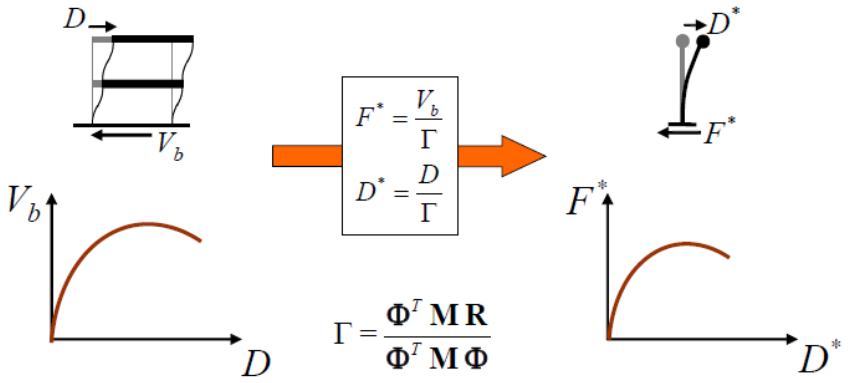


Figure 5: Capacity curve, transformation from response of MDOF to equivalent SDOF.

In order to compare the capacity curve to the demand curve given by the design spectrum, the nonlinear pushover curves of the SDOF are approximated by elastic-perfectly plastic (or bilinear) curves. A target displacement is assumed, and equal energy is assumed between bilinear and nonlinear pushover curves. This simple procedure is illustrated in Figure 6.

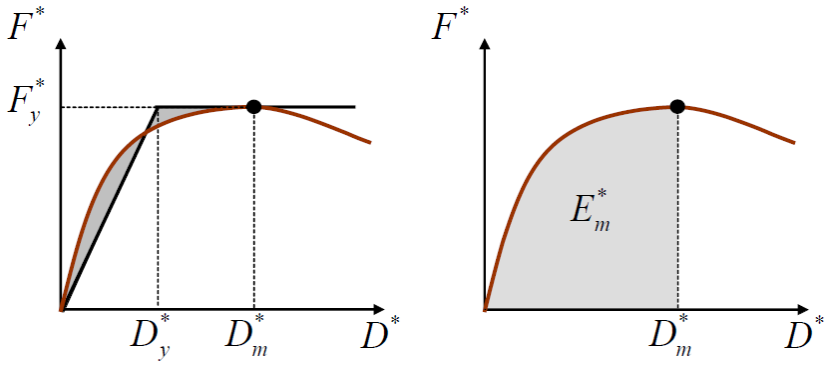


Figure 6: Bilinearization of the capacity curve of SDOF

The bilinearization of Figure 6 gives the yield force and the yield displacement

$$D^*_y = 2\left(D^*_m - \frac{E^*_m}{F^*_y}\right) \tag{2.28}$$

which allow the initial elastic period to be computed as:

$$T^* = 2\pi\sqrt{\frac{m^* D^*_y}{F^*_y}} \tag{2.29}$$

2.4. Bilinear hysteretic model

This chapter will refer to the Bilinear Hysteretic Model and specifically in its wording concerning the stress-strain relationship and which is similarly used for a single degree-of-freedom system. In the past it was one of the most common models for non-linear dynamic analysis. Most useful used mainly to describe the behavior of steel.

The Bilinear Model is generally described by:

• the elastic branch before yielding

• the post-elastic branch after yielding

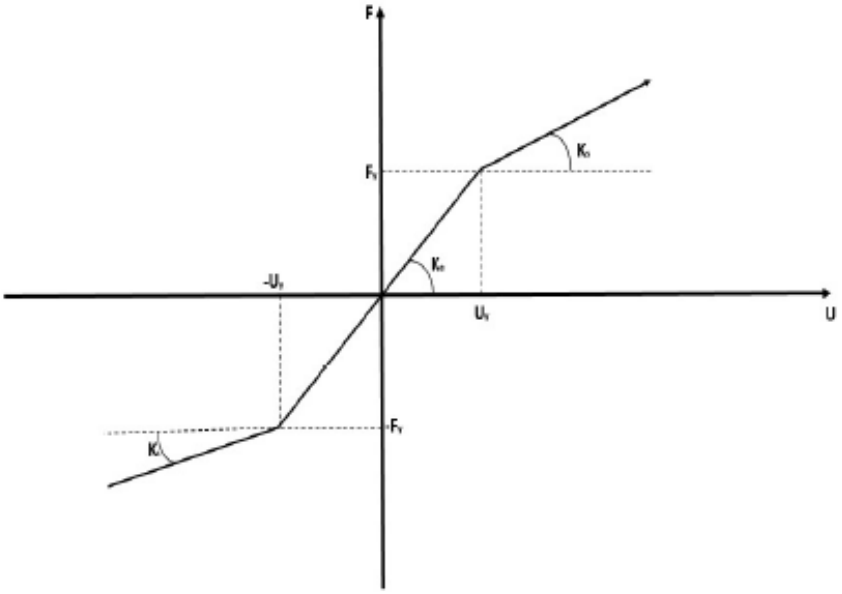


Figure 7: Bilinear Hysteretic Model

$$F = K_{elastic} \cdot u \text{ , για } -u_y \leq u \leq u_y \tag{2.30}$$

$$F = sign(u)\,\mathrm{K}_{\mathrm{elastic}}\,u_y + sign(u)K_s(|u| - u_y) \text{ , για } u \leq -u_y \text{ και } u \geq u_y \tag{2.31}$$

where:

- $K_{elastic}$: elastic stiffness
- u_y : yield displacement
- K_s : post-elastic stiffness
- F_y : yield strength

A code sample follows below:

```
def bilinear(kel, b1, uy, up, fsp, u):
   """Bilinear Hysteric Law model"""
   a1 = fsp - kel * uy
   a2 = (b1/(b1 - 1)) * (fsp - kel * up)
   a = max(a1, a2)
   sel = fsp + (u - up) * kel
   h = sel - a
   q = abs(h) - kel * uy

   if q <= 0:
       fs = sel
   else:
       depl = (q*(1 - b1))/kel
       fs = sel - np.sign(h) * kel * depl
   return fs
```

3. Ground motion simulation

The study of the ability to replace recordings with their pulse simulations requires the creation of the pulse extraction code which detects and extracts the required number of pulses from a near field ground motion. A basic method of pulse extraction is that introduced by Mavroeidis and Papageorgiou (2003), that consists a mathematical representation of near-field motions relying on designing a composite wavelet (based on Gabor wavelet). Although various wavelets have been proposed in the literature, only a limited number of them are popular and frequently used in practice. The most common wavelets are summarized in Table 1 along with their analytical expressions, input parameters, and associated references.

Table 1: Common Analytic Wavelets Used in Seismology

Wavelet	**Analytical expression**
Gabor	$f(t) = Ae^{-(\frac{2\pi f_p}{c})^{2t^2}} \cos[2\pi f_p t + m]$
Berlage	$f(t) = AH(t)t^n e^{-\left(\frac{2\pi f_{p/\gamma}}{c}\right)^t} \cos[2\pi f_p t + m]$
Generalized Rayleigh	$f(t) = A(-1)^k \frac{e^{i(u+\frac{\pi}{2})}}{(i+\frac{2\pi f_p t}{k})^{k+1}}$
Kupper	$f(t) = A\left[\sin(m^{\pi t}) - \frac{m}{m+2}\sin(m+2^{\pi t})\right] for\ 0 < t < T$
Ricker	Three loop (symmetric): $f(t) = A(1-2\pi^2 {f_p}^2 t^2)e^{-(\pi f_p)^2 t^2}$ Two loop (antisymmetric) $f(t) = Ate^{-(2\pi f_p)^2 t^2}$

where:

A: amplitude

f_p: is the prevailing frequency

γ: oscillatory character

u: phase delta

c: damping coefficient

n: asymmetry of envelope function

T: duration

m: controls the number of half-cycles

H(t): Heaviside unit step function

3.1. Simulation of a simple model for near field pulses

The analytical model of choice should satisfy the conditions and possess the following properties:

1. The synthetic wavelet should be expressed by a properly parameterized simple mathematical expression that, with a minimum number of input parameters that have an

unambiguous physical interpretation, allows as much flexibility as is necessary to represent, reasonably accurately, near-source pulses.

2. The synthetic wavelet should be capable of simulating as many as possible of the near-source records (preferably all of the records in Table 1).

3. The mathematical expression of the wavelet should be such that it facilitates derivation of closed-form expressions of its spectral characteristics in the form of Fourier transform and response spectra. Such closed-form expressions make considerably easier the parametric study of the response of structures to near-source pulses.

From the wavelets listed in Table 2, the analytical model that fulfills most of the aforementioned conditions and has the potential, if slightly modified, to entirely meet our requirements is the Gabor wavelet. This signal is the product of a harmonic oscillation and a bell-shaped function (i.e. Gaussian envelope). The Gabor wavelet is defined by the four parameters identified at the beginning of this section as the key features that determine the waveform characteristics of the near-source velocity pulses; namely, A, f_P, ν, and γ (see Table 1) define the amplitude, prevailing frequency, phase, and oscillatory character of the signal, respectively. The Berlage wavelet is similar to the Gabor signal because both of them consist of an amplitude-modulated harmonic. Furthermore, the common input parameters of the two signals have the same physical meaning. However, the Berlage wavelet is characterized by an additional free parameter (parameter n; see Table 1) that controls the skewness of its envelope function. This extra degree of freedom, although theoretically useful, is not of great practical importance. The need for a skewed (i.e., nonsymmetric) envelope may be accommodated (to a certain extent) by varying the phase angle of the amplitude-modulated harmonic. Furthermore, the existence of the factor t^n in the mathematical expression of the Berlage wavelet introduces additional complexity in the analytical derivations to follow, without significant benefits.

The three-loop symmetric and two-loop antisymmetric Ricker wavelets are not adequate to describe a broad range of near-fault velocity pulses; only a small number of recorded ground motions exhibit purely symmetrical or anti-symmetrical characteristics. There is no unique mathematical expression for “arbitrary” (meaning not perfectly symmetrical or antisymmetrical) Ricker wavelets; depending on the waveform to be obtained, different polynomial functions should be selected (Ricker, 1944, 1945). Hosken (1988) discussed the applicability and usefulness of the Ricker wavelet, focusing on the drawbacks of the signal. The author also proposed a technique to generalize the Ricker wavelet by generating signals in between the antisymmetric two-loop and the symmetric three-loop Ricker wavelets. Besides the complexity in the mathematical expressions of the generalized Ricker wavelets, it is apparent that a two-parameter model is far too constrained to allow accurate fitting of the fairly complicated recorded near-fault velocity pulses.

Like the Ricker wavelet, the original Rayleigh signal (e.g., Hudson, 1980) is a two-parameter model inflexible for most synthetic seismogram applications. Hubral and Tygel (1989) overcame this deficiency by deriving a generalized analytic wavelet based on the Rayleigh signal. The real or imaginary part of the mathematical expression of this generalized wavelet may be used to generate synthetic velocity pulses. Even though this is not a difficult task, the resulting formulas are quite complicated for further derivations (e.g., differentiation and integration to obtain acceleration and displacement time histories, respectively; response of an SDOF system; etc.).

On the other hand, the analytical expression of the Küpper wavelet is straightforward and its input parameters have a clear physical meaning. However, the waveforms generated by the Küpper signal are either symmetric or antisymmetric for odd or even m values, respectively. The lack of a phase parameter does not facilitate generation of “arbitrary” signals. This is an important limitation previously addressed with respect to the Ricker wavelet.

3.2. Mavroeidis and Papageorgiou wavelet

Based on the above, the Gabor wavelet is the analytical model (among those listed in Table 2) that better serves our needs. However, no closed-form solution can be derived for the response of an SDOF system when excited by synthetic ground motions generated by the Gabor signal. Such a closed-form expression would greatly facilitate parametric analyses, a matter of great importance to earthquake engineers. The difficulties in the derivations are caused by the exponential function (i.e., Gaussian envelope) included in the mathematical expression of the Gabor wavelet. Therefore, to overcome this difficulty, we propose an analytical model that retains the advantages of the Gabor wavelet (e.g. number of parameters, physical interpretation of them, simple mathematical expression, large flexibility in synthetic waveforms), while at the same time yields a closed-form expression for the response of the SDOF system subjected to synthetic ground motions generated by the model. To that effect, we have replaced the Gaussian envelope of the Gabor wavelet by another symmetric bell-shaped function that possesses a simpler analytical expression. Namely, a shifted haversed sine function (i.e., an elevated cosine function) is used to replace the Gaussian envelope, while the harmonic oscillation part remains the same. Thus, the proposed analytical signal is expressed by

$$f(t) = A\frac{1}{2}\left[1 + cos\left(2\pi\frac{f_{\mathrm{p}}t}{\gamma}\right)\right]cos(2\pi\, f_{\mathrm{p}}t + v) \tag{3.1}$$

The following remarks can be made pertaining to the M&P wavelet:

•The shifted haversed sine function is a periodic function; consequently, it does not produce an envelope with a single hump like the Gaussian function of the Gabor wavelet. This problem is easily resolved by limiting the time interval of the signal as follows:

$$-\frac{\gamma}{2f_{\mathrm{p}}} \leq \mathrm{t} \leq \frac{\gamma}{2f_{\mathrm{p}}} \tag{3.2}$$

The period of the harmonic oscillation should be smaller than the period of the envelope represented by the elevated cosine function in order to produce physically acceptable signals; that is,

$$\frac{1}{f_{\mathrm{p}}} \leq \frac{\gamma}{f_{\mathrm{p}}} \Rightarrow \gamma > 1 \tag{3.3}$$

It is convenient for the calibration of the model to introduce a time shift, t_0, in equation (3.1) to precisely define the epoch of the envelope’s peak. This parameter is frequently introduced

in all models listed in Table 2 as a feature that provides extra flexibility to the signal, allowing its translation along the time axis. Thus,

$$t \Rightarrow t - t_0 \tag{3.4}$$

The combination of equations (1) and (2) yields the formulation of the proposed analytical model for the near-fault ground velocity pulses:

$$\begin{aligned} v(t) &= \frac{A}{2}\left[1 + COS\left(2\,\pi\frac{f_p}{\gamma}\right)(t - t_0)\right] cos[2\pi f_p(t - t_0) + v], \\ t_0 - \gamma f_p &\le t \le t_0 + \gamma f_p, \mu\varepsilon\ \gamma > 1 \\ v(t) &= 0, otherwise \end{aligned} \tag{3.5}$$

Parameter A controls the amplitude of the signal, f_P is the frequency of the amplitude-modulated harmonic (or the prevailing frequency of the signal), v is the phase of the amplitude-modulated harmonic (i.e., v=0 and v=±π/2 define symmetric and antisymmetric signals, respectively), c is a parameter that defines the oscillatory character (i.e.,zero crossings) of the signal (i.e., for small c the signal approaches a deltalike pulse; as c increases the number of zero crossings increases), and t_0 specifies the epoch of the envelope's peak.

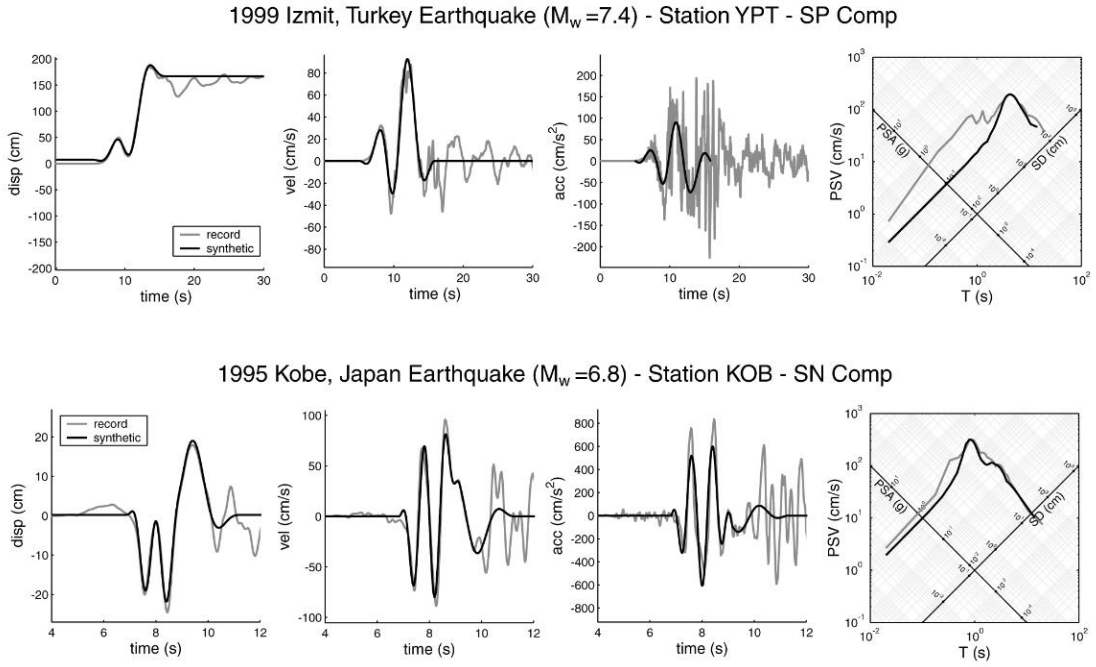


Figure 8: M&P model fitted to recorded motions (top) two, synthetic pulses have been combined (Izmit, Turkey, 1999) and (bottom), three synthetic pulses have been combined to generate the illustrated synthetic time series (1995 Kobe, Japan).

To demonstrate that the M&P wavelet produces almost identical pulses with the Gabor wavelet and therefore justify the substitution of the Gaussian envelope with the elevated cosine function, both analytical models have been used to generate synthetic signals that simulate the fault-normal velocity pulse recorded at station E06 during the 1979 Imperial Valley, California, earthquake. The comparison is illustrated in Figure 8; the harmonic oscillation (i.e., the carrier signal) is the same for both pulses, while their envelope functions are very similar. The synthetic signals produced by both models reproduce the longer-period portions of the observed ground motion.

A significant feature of the M&P wavelet is the objective definition of the pulse duration based on model input parameters. In the literature, no unique method exists to define the duration or period of the velocity pulse, even though this parameter is extensively used. Many researchers determine the pulse duration using the zero crossings of the pulse waveform. Others prefer the utilization of the peak value of the velocity response spectrum to indirectly define the pulse period. In many other cases, no explanation is provided regarding the estimation of this parameter. The M&P wavelet features an objective definition of the pulse duration (T_P) compatible with the physical aspects of the problem as the inverse of the prevailing frequency (f_P) of the signal; that is,

$$T_P = 1/\ f_P \tag{3.6}$$

The analytical expressions for the ground acceleration and displacement time histories compatible with the ground velocity given by equation (3.5) are

$$a(t)=\begin{cases} -\dfrac{A\pi f_p}{\gamma}\left[\begin{array}{l}\sin\left(\dfrac{2\pi f_p}{\gamma}(\mathrm{t}-\mathrm{t}_0)\right)\cos\left[2\pi f_p(t-t_0)+v\right]\\ +\gamma\cdot\sin[2\pi f_p(t-t_0)+v\left[1+\cos\left(\dfrac{2\pi f_p}{\gamma}(\mathrm{t}-\mathrm{t}_0)\right)\right]\end{array}\right], \mathrm{t}_0-\dfrac{\gamma}{2f_p}\le t\le t_0+\dfrac{\gamma}{2f_p},\gamma>1 \\ 0, otherwise \end{cases} \tag{3.7}$$

$$\mathrm{d}(t)=\begin{cases} \dfrac{A}{4\pi f_p}\left[\begin{array}{l}\sin\left[2\pi f_p(\mathrm{t}-\mathrm{t}_0)+\mathrm{v}\right]+\dfrac{1}{2}\dfrac{\gamma}{\gamma-1}\sin\left[\dfrac{2\pi f_p(\gamma-1)}{\gamma}(\mathrm{t}-\mathrm{t}_0)+v\right]\\ +\dfrac{1}{2}\dfrac{\gamma}{\gamma+1}\sin\left[\dfrac{2\pi f_p(\gamma+1)}{\gamma}(\mathrm{t}-\mathrm{t}_0)+v\right]\end{array}\right]+C \\ \dfrac{A}{4\pi f_p}\dfrac{\gamma}{\left(1-\gamma^2\right)}\sin\left(v-\pi\gamma\right)+C, t<t_0-\dfrac{\gamma}{2f_p} \\ \dfrac{A}{4\pi f_p}\dfrac{\gamma}{\left(1-\gamma^2\right)}\sin\left(v+\pi\gamma\right)+C, t<t_0+\dfrac{\gamma}{2f_p} \end{cases}, \mathrm{t}_0-\frac{\gamma}{2f_p}\le t\le t_0+\frac{\gamma}{2f_p},\gamma>1 \tag{3.8}$$

The constant displacement values for $t\ <\ t_0-\gamma/\ 2f_p$ and $t\ >\ t_0+\gamma/\ 2f_p$ were specified so that the displacement time histories satisfy continuity condition at $t\ =\ t_0-\gamma/\ 2f_p$ and $t\ =\ t_0+\gamma/\ 2f_p$.

When the velocity pulse is integrated to obtain the closed-form displacement, an as yet unspecified constant, C, appears in the derived displacement expression, that is, which for simplification reasons can be taken as zero.

3.3. A characteristic application of the pulse extraction

The pulse extraction algorithm has been applied to the ground motions of the database created in Section 4.1. It is clearly shown that each separate pulse is shifted in time (Figure 9) and the cumulative velocity signal is fitted adequately to the original ground motion (Figure 10).

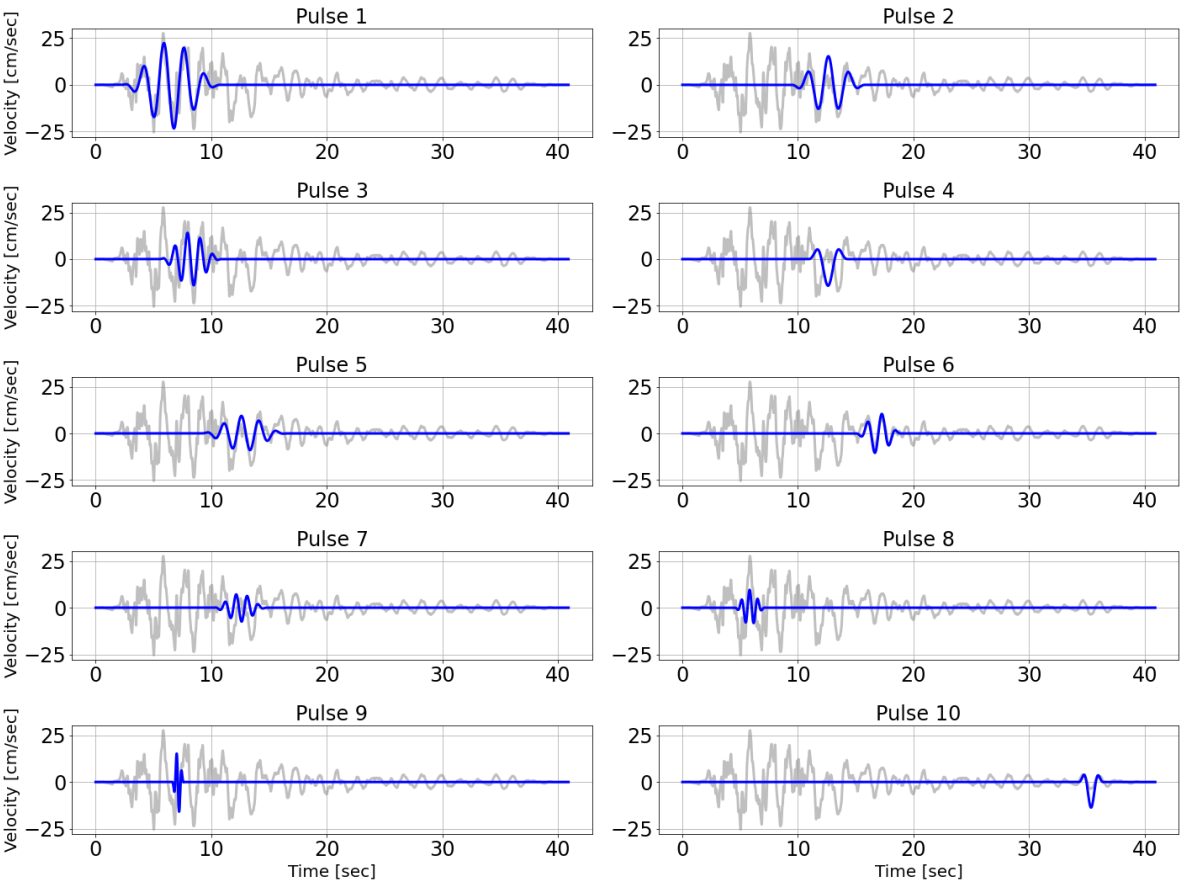


Figure 9: Velocity plots for 10 pulses extraction.

Separate pulse (blue) vs Original velocity time-history (grey).

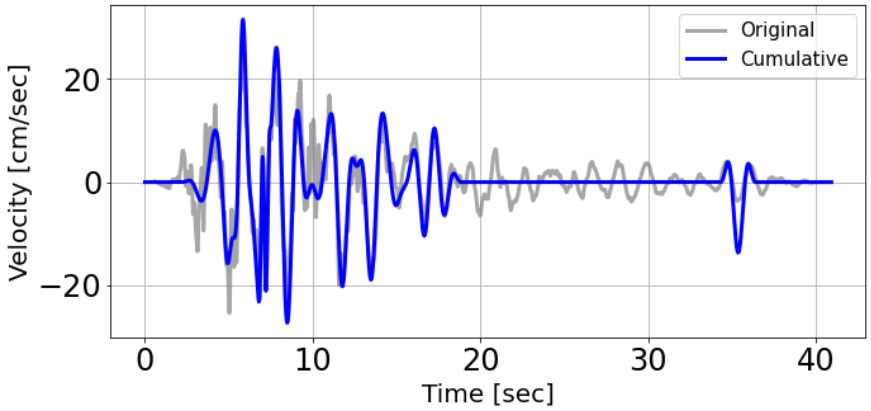


Figure 10: Original vs Cumulative velocity plots for 10 pulses extraction.

4. Case study: a machine-learning approach

In this chapter, a machine learning pipeline is proposed in order to estimate the non-linear response analysis of structural systems. This is carried out in terms of maximum displacements aiming to eliminate the computational cost of the NRHA analysis. A pulse extraction process is used to quantify the wavelet parameters of a ground motion records which, along with the material parameters of the structural system consist an adequate training data set. In the following sections the applicability and performance of the proposed approach is demonstrated in single-DOF structural systems.

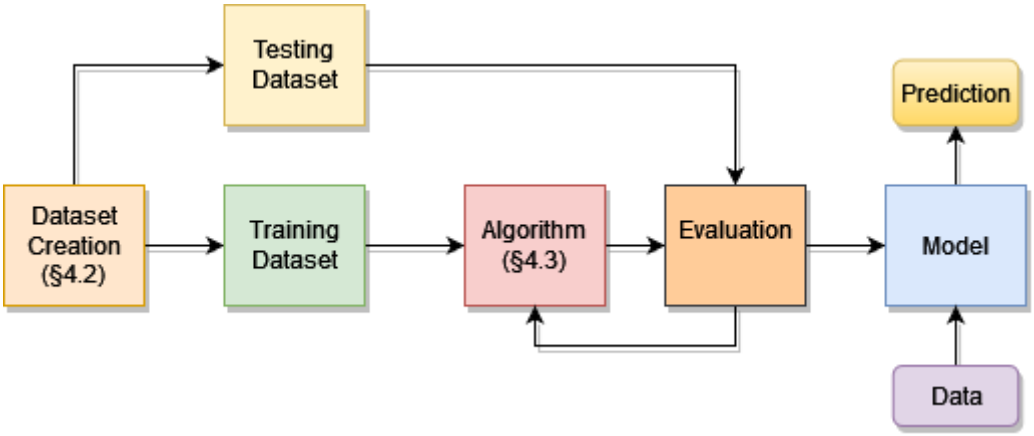


Figure 11: Machine learning pipeline flowchart

4.1. Ground motion records selection

A large number of ground motion records is necessary for the generation of the dataset that will be used for the training and the evaluation of the ML models. For this reason, a database was generated consisting of 1716 near-field ground motions records (distance from the rapture<60 km), obtained from the Next Generation Attenuation for Western US (NGA-West2) (PEER 2017). The NGA-West2 ground motion database includes a very large set of ground motions recorded in worldwide shallow crustal earthquakes in active tectonic regimes. The database has one of the most comprehensive sets of meta-data, including different distance measure, various site characterizations, earthquake source data, etc. Among of the directions of motion available, only the two horizontal signals were kept. For each record, the acceleration time-history signal and the time-step were stored in the database, along with a unique identification number of the record. The geographical locations of the ground motions are shown in Figure 12 below. In addition, histograms of characteristic ground motion parameters such as the peak ground acceleration (PGA), the magnitude, the fault rapture area and the shear-wave velocity (V_s^{30}) are provided in Figure 13.

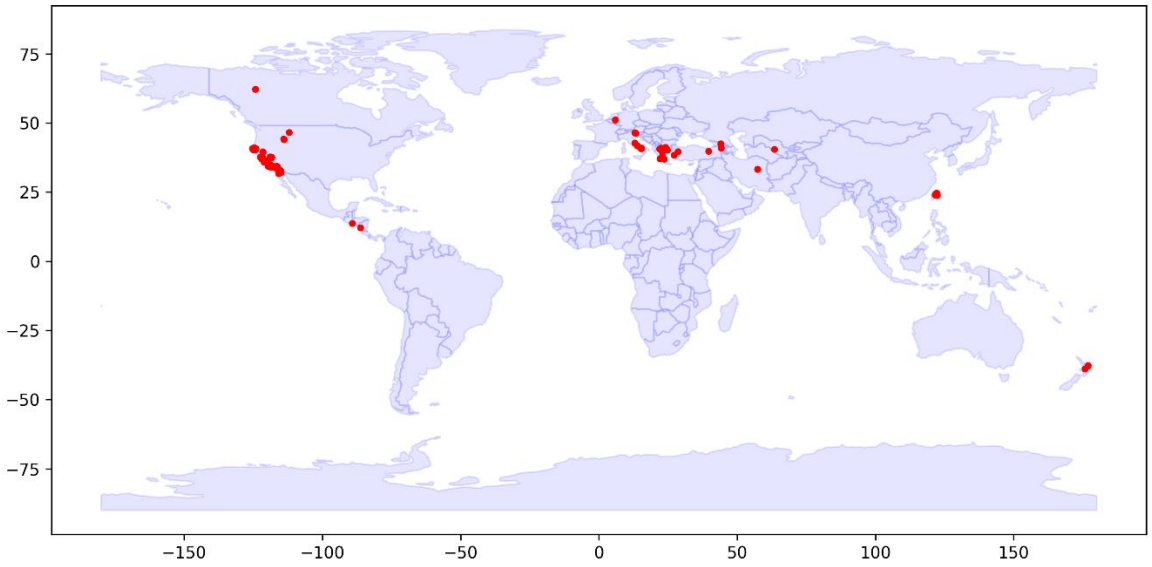


Figure 12: Geographical location of the ground motion records of the database

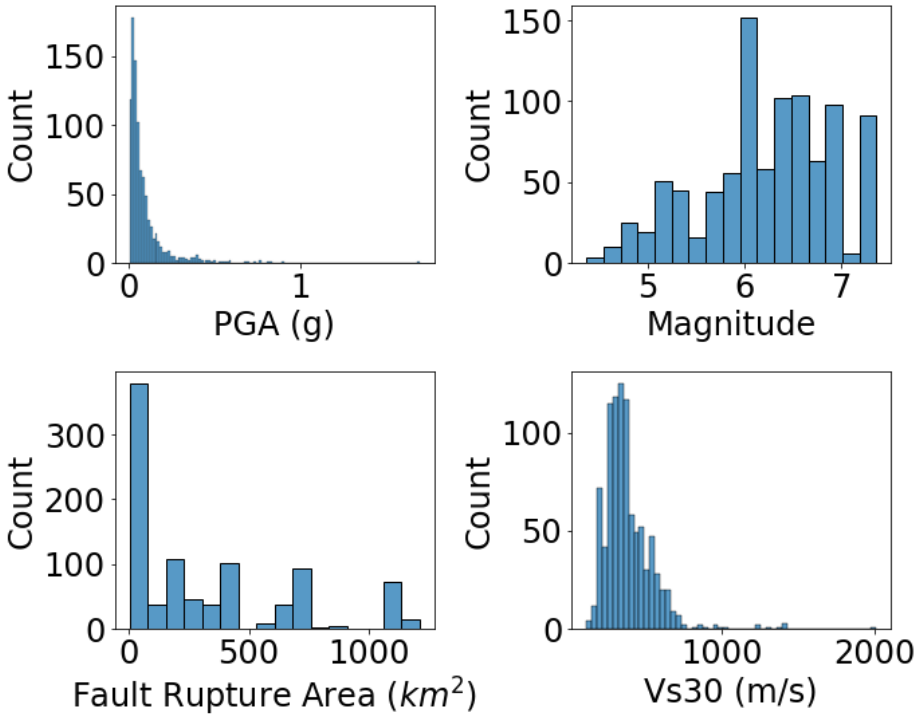
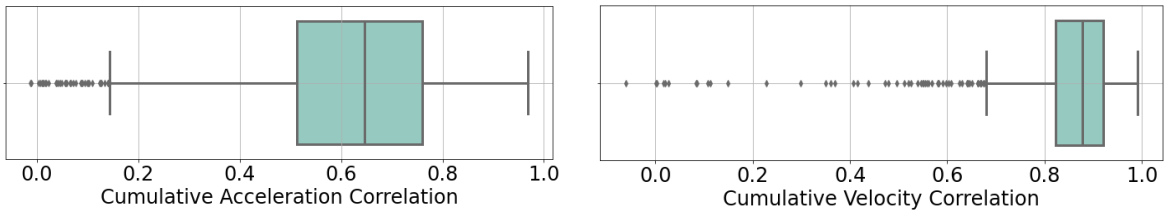


Figure 13: Histograms of characteristic ground motion parameters of the database

Having a strong correlation of the cumulative velocity with the original ground motion record signal is also important. Based on Table 2, descriptive statistics for the cumulative correlations are produced. It shown that the cumulative correlation of the acceleration signal with the original ground motion acceleration is generally slightly lower than velocity's. This is justified

because the pulse extraction algorithm's targeted series is the velocity time history due to being less noisy than the acceleration one. In order to achieve better fitting of the simulated pulse-like signals with the original ground motion records we will filter the generated dataset by selecting only those with cumulative velocity correlation greater than 0.8. This reduces the total number of records in our dataset from 1716 to 1405.



(a) (b)

Figure 14: Boxplots of (a) Cumulative Acceleration Correlation and (b) Cumulative Velocity Correlation

	Cumulative Acceleration Correlation	Cumulative Velocity Correlation
count	1716	1716
mean	0.62	0.86
std	0.19	0.11
min	0.00	0.00
25%	0.51	0.82
50%	0.65	0.88
75%	0.76	0.92
90%	0.85	0.95

Table 2: Descriptive statistics for the Cumulative Acceleration Correlation and Cumulative Velocity Correlation.

4.2. Dataset creation

Using the ground motion record database that was generated in Section 4.1, the dataset which will be used for the training of the ML models can be created. This dataset is artificially created by combining bilinear oscillator response analysis results i.e. maximum displacement u_{max} of the SDOF system as well as the pulse representation parameters of each near-field ground motion record.

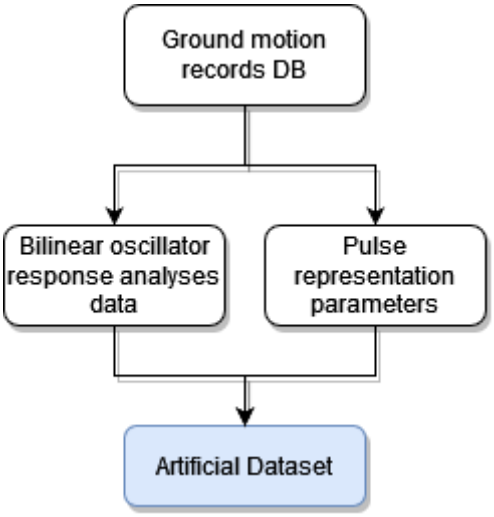


Figure 15: Dataset creation flowchart

4.2.1. Bilinear oscillator response analysis

In order to calculate the maximum displacement u_{max} for each ground motion record of the database nonlinear history response analysis is conducted. This was achieved by using the Central Difference Method that was described in Section 2.3.1. The ranges of the input parameters (T, F_y) are shown in Table 3. For the hardening parameter k of the bilinear model (Section 0) zero value is considered. Also, the damping ratio ζ was taken as 5% for all analyses. Thus, 9 x 5 x 1405 = 63.225 data points were generated.

Table 3: Lower, upper bound and step values for SDOF bilinear oscillator

Variable	Lower Bound	Upper Bound	Step
T [s]	0.1	0.9	0.1
F_y [% of SDOF mass]	0.1	0.5	0.1

4.2.2. Pulse representation of near-field ground motions

Ground motion records cannot be directly used as data points for the training of machine learning models. For this reason, it is necessary for each record to be converted into a single-row format, that contains all the important information and characteristics of the ground motion. This can be succeeded by replacing the original records with their pulse representations using the pulse extraction method that was described in Chapter 3.

Each ground motion record of the database will be represented by a cumulative signal of pulses. In order to indicate the accuracy of the fitting to the original signal the correlation of the cumulative signal of velocities is used. In this study, we will examine the cases of representation of the ground motion signal with 10 and 2 pulses. Highest number of pulses always leads to higher computational cost in compensation of higher correlation between the original and the simulated signal.

Applying the pulse extraction algorithm each ground motion record using 10 pulses, we get the parameters A_p, T_p, γ, v, and t_b for each pulse (50 parameters in total), as shown in Table 3 below.

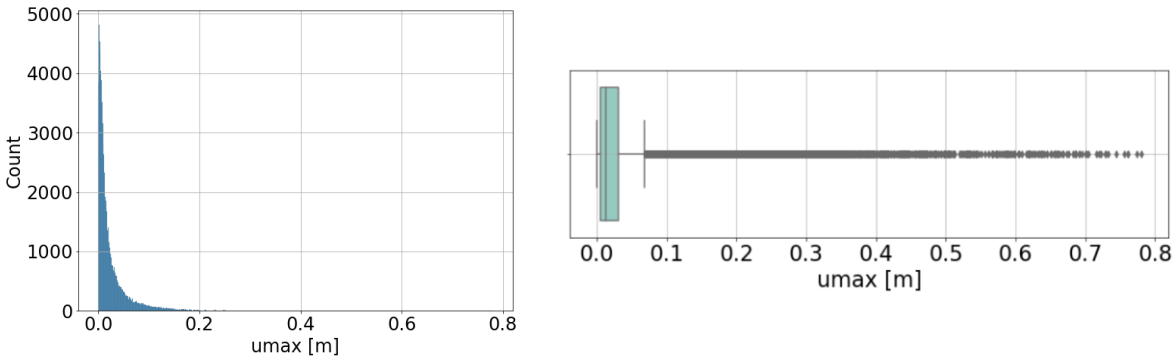
Table 3: Pulse extraction algorithm results for a ground motion record using 10 pulses

	A_p	T_p	γ	v	t_b	Cumulative Velocity Correlation
1	0.05	4.15	5	170	10.38	0.51
2	0.08	1.75	4	150	3.15	0.72
3	0.05	1.48	4	35	2.73	0.76
4	0.03	1.68	10	25	8.40	0.81
5	0.03	2.88	6	355	8.49	0.85
6	0.02	1.49	10	20	7.15	0.86
7	0.08	0.68	2	260	0.68	0.88
8	0.03	0.88	7	55	3.25	0.89
9	0.01	5.00	4	210	11.00	0.90
10	0.03	2.07	3	15	2.79	0.91

4.2.3. Exploratory data analysis

In order to understand better the training dataset, we will make use of histograms, boxplots and descriptive statistics. This procedure is also used as a visual method of detecting and removing outliers in the dataset which often lead to improve the models performance.

It is very important to understand the data of the predicted variable, check its distribution and search for possible outliers. As shown in the histogram in Figure 16a max displacement does not follow a normal distribution and according to the boxplot in Figure 16b values greater than 0.07m could be considered as “outliers”.



(a) (b)

Figure 16: (a) Histogram, (b) boxplot of the max displacement u_{max}

Regarding the input features of the dataset, inFigure 17, it is shown that all of the pulse parameters follow similar distributions for different number of pulses.

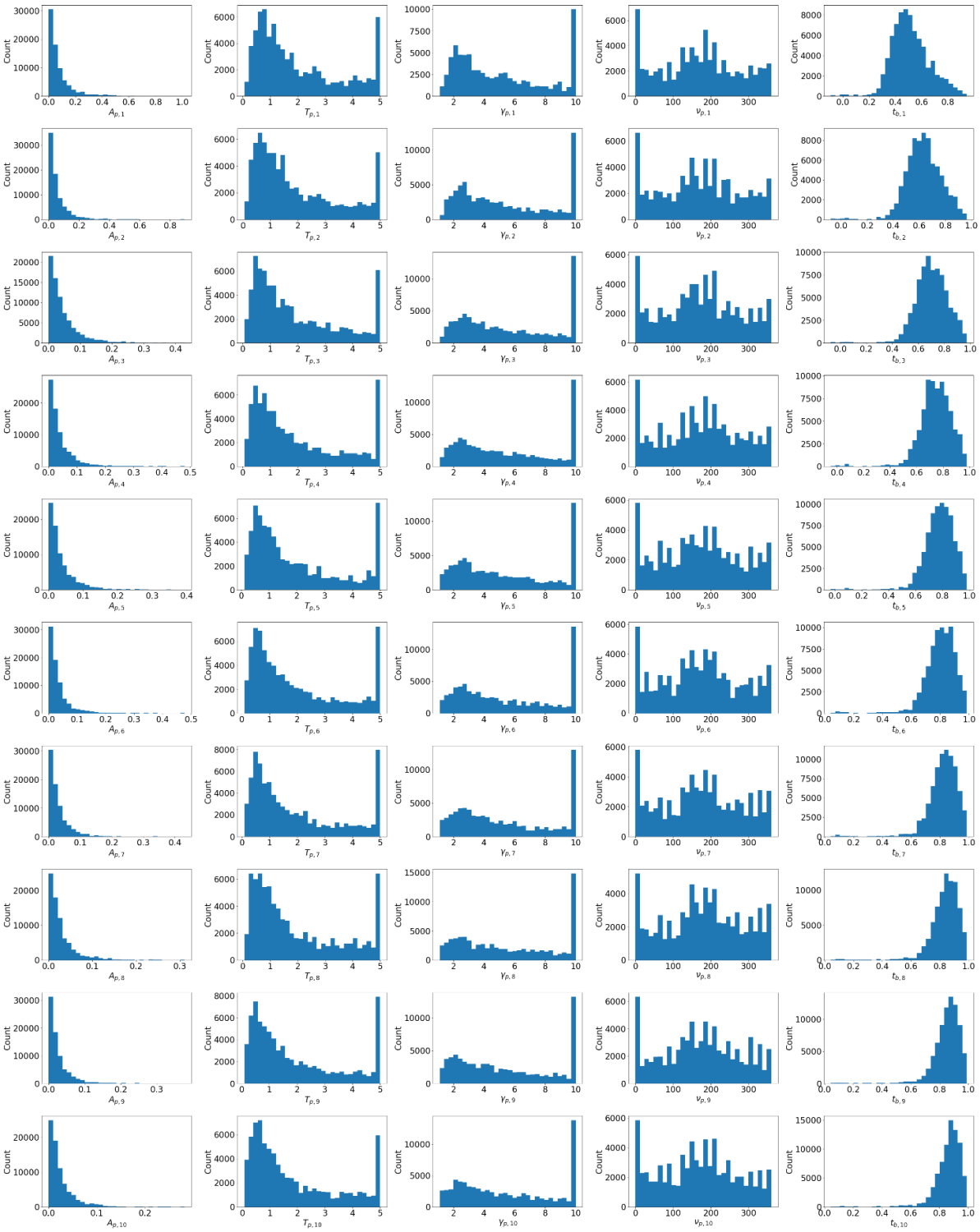


Figure 17: Histograms of the distribution of the pulse representation parameters

The interactions between the input features, T, F_y, A_p, T_p, γ_p, v_p, t_p, and the max displacement u_{max} are investigated using the scatter plots show in Figure 18 and the correlation heatmap shown in Figure 19. For the sake of simplicity only the first pulse parameters are shown. Based on the scatter plots in Figure 18, there is an indication that the variable A_p has an effect on the max displacement u_{max} which is also shown in the correlation heatmap. The

amplitude A_p gives the highest correlation between the rest of the input variables and it is expected to have the highest importance among the rest of the input variables.

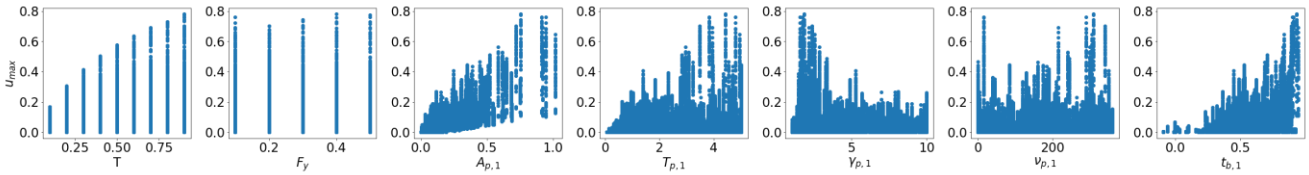


Figure 18: Scatter plots if the max displacement and the input variables of the model

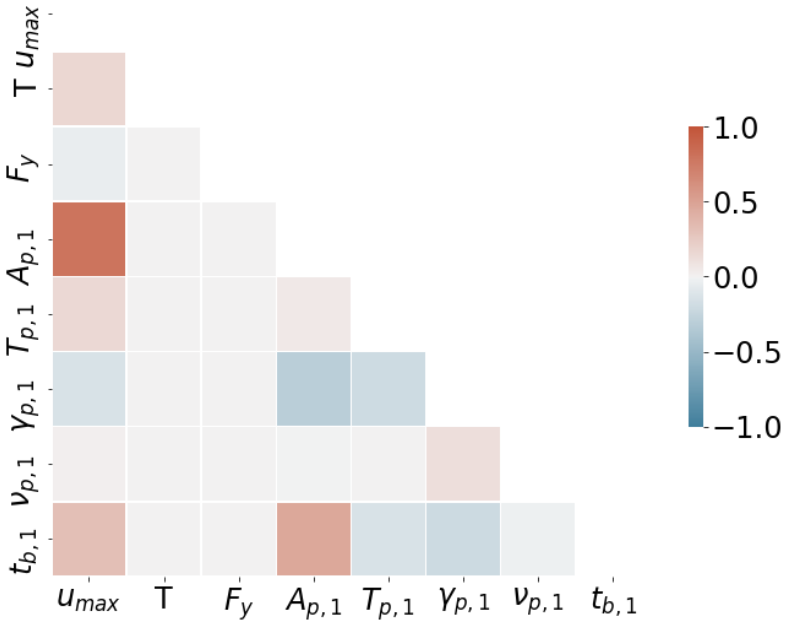


Figure 19: Correlation heatmap between the input and output variables of the model

4.3. ML model

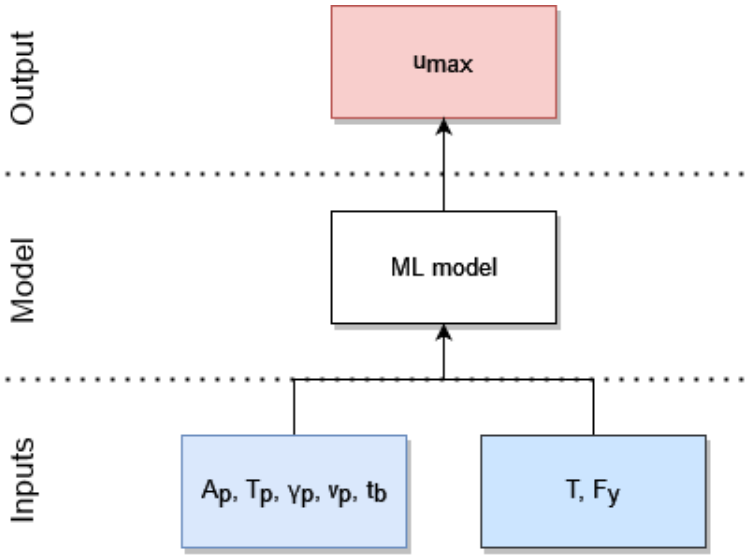


Figure 20: ML model definition

In order to train a model that has the ability to predict the max displacement u_{max} the dataset from Section 4.2 needs to be modified. The material parameters (T, F_y) of the single-DOF oscillator along with the pulse extraction parameters (A_p, T_p, γ_p, ν_p, t_b) will be used as input variables and the max displacement as the predictor variable. For the pulse extraction parameters, a parametric study will be conducted using through a range of 10 pulses. Also, a machine learning algorithm pool has been created in order to train, test and evaluate different ML models including:

- Ridge
- PLS Regression
- Decision Tree
- Random Forest
- XGBoost
- Artificial Neural Network

The results of the parametric study are shown in Figure 21. It can be observed that the algorithms with the best performance are the XGBoost and Random Forest. It can also be noticed that the performance of the models is generally stable when the number of pulses increases. This could be justified as the models are able to capture the important patterns in the data even when using only the first pulse for the representation of the ground motion signal. Algorithms such as Ridge and PLS Regression seem to have the highest errors as their linear nature is inadequate to capture the nonlinearities of the problem.

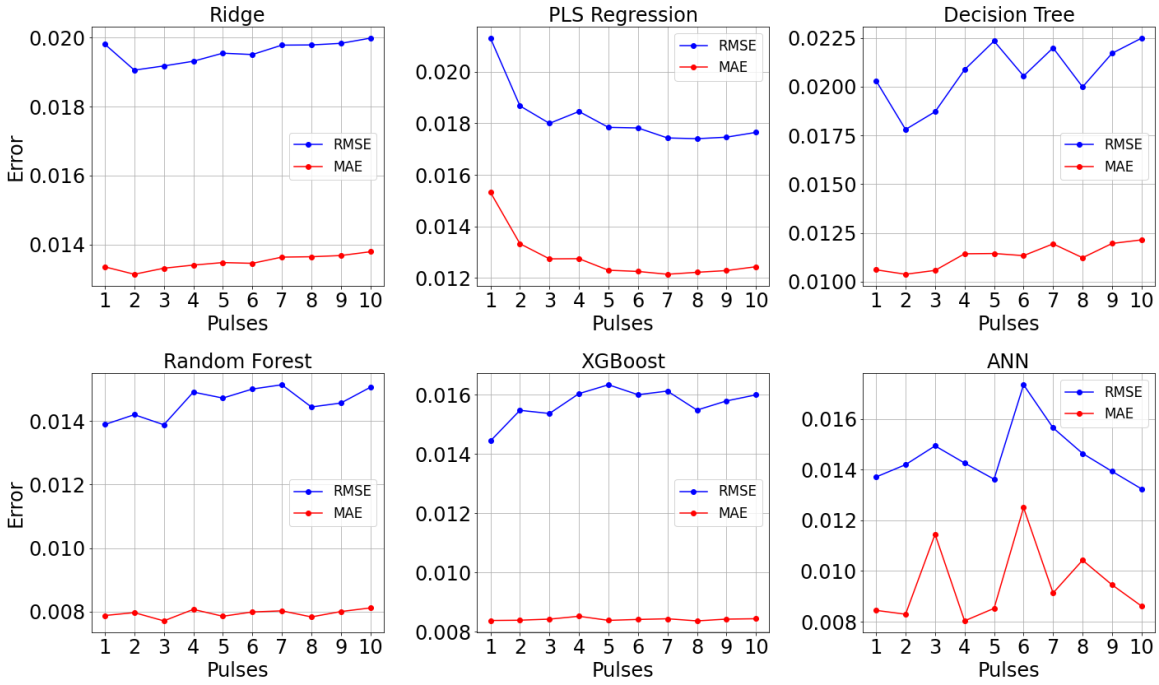


Figure 21: Performance of the ML models in the test set monitoring the RMSE and MAE metrics

In the following sections, the trained ML models using 1 and 10 pulses will be evaluated in more depth and the most important features of each model will be investigated.

4.3.1. 1 pulse

The parameters from 1 pulse will be used as inputs for the model training. Thus, the model uses 5 (parameters of each pulse) x 1 (pulses) + 2 (T and F_y) = 7 features as inputs.

The train set consists of 1124 records (80% of the dataset) and the test set of 281 records (20% of the dataset). The performance of each ML model can be shown through the root mean square error and mean absolute error metrics (Table 4) and the predictions vs real value scatter plots (Figure 22):

Table 4: Root Mean Square Error and Mean Absolute Error metrics for each ML model using 10 pulses.

Algorithm	RMSE	MAE
Ridge	0.0198	0.0134
PLS Regression	0.0213	0.0153
Decision Tree	0.0203	0.0106
Random Forest	0.0139	0.0079
XGBoost	0.0144	0.0084
ANN	0.0137	0.0084

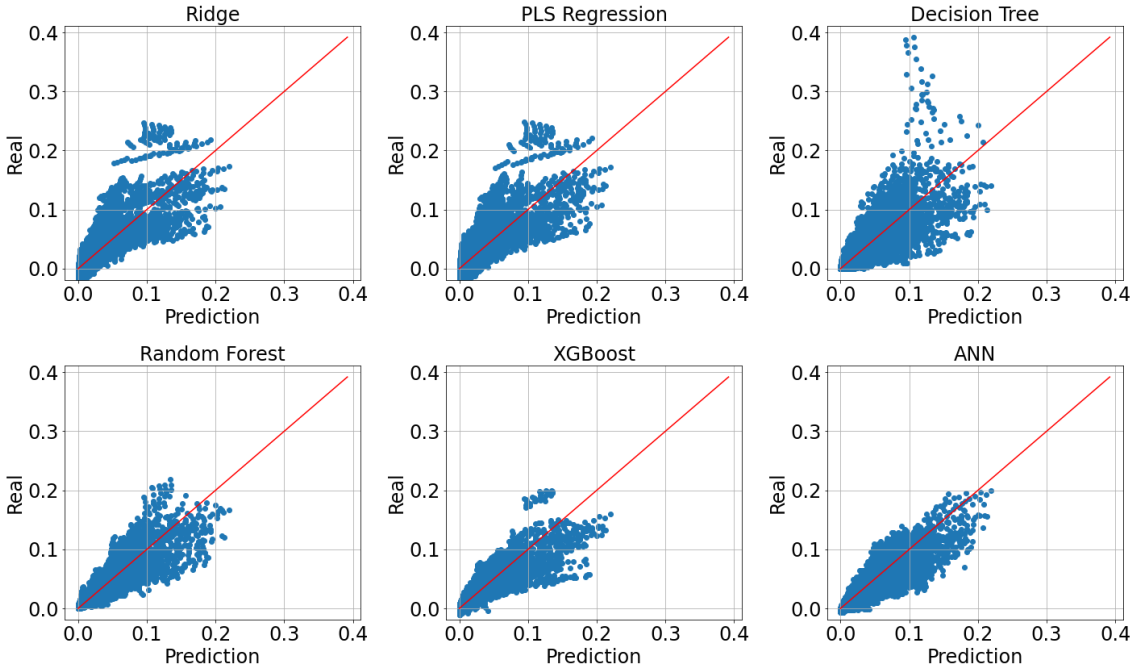


Figure 22: Scatter plots of the Predictions vs Real values of the max displacement using 1 pulse.

Based on the scatter plots in Figure 22 it is shown that all of the ML models have performed generally well and were able to give adequate prediction. Among them, the Random Forest and the Artificial Neural Network had the lowest RMSE and MAE metrics.

Feature Importance

Evaluating a models performance only by the error metrics is not enough in most situations. A model can give predictions with low error metrics but these predictions can be based on features with low importance according to the domain knowledge and laws.

Permutation feature importance is a model inspection technique that can be used for any fitted estimator when the data is tabular. This is especially useful for non-linear or opaque estimators. The permutation feature importance is defined to be the decrease in a model score when a single feature value is randomly shuffled. This procedure breaks the relationship between the feature and the target, thus the drop in the model score is indicative of how much the model depends on the feature. This technique benefits from being model agnostic and can be calculated many times with different permutations of the feature. Features that are deemed of low importance for a bad model (low cross-validation score) could be very important for a good model. Therefore, it is always important to evaluate the predictive power of a model using a held-out set (or better with cross-validation) prior to computing importances. Permutation importance does not reflect to the intrinsic predictive value of a feature by itself but how important this feature is for a particular model. Below we can see the permutation importance of the trained ML models.

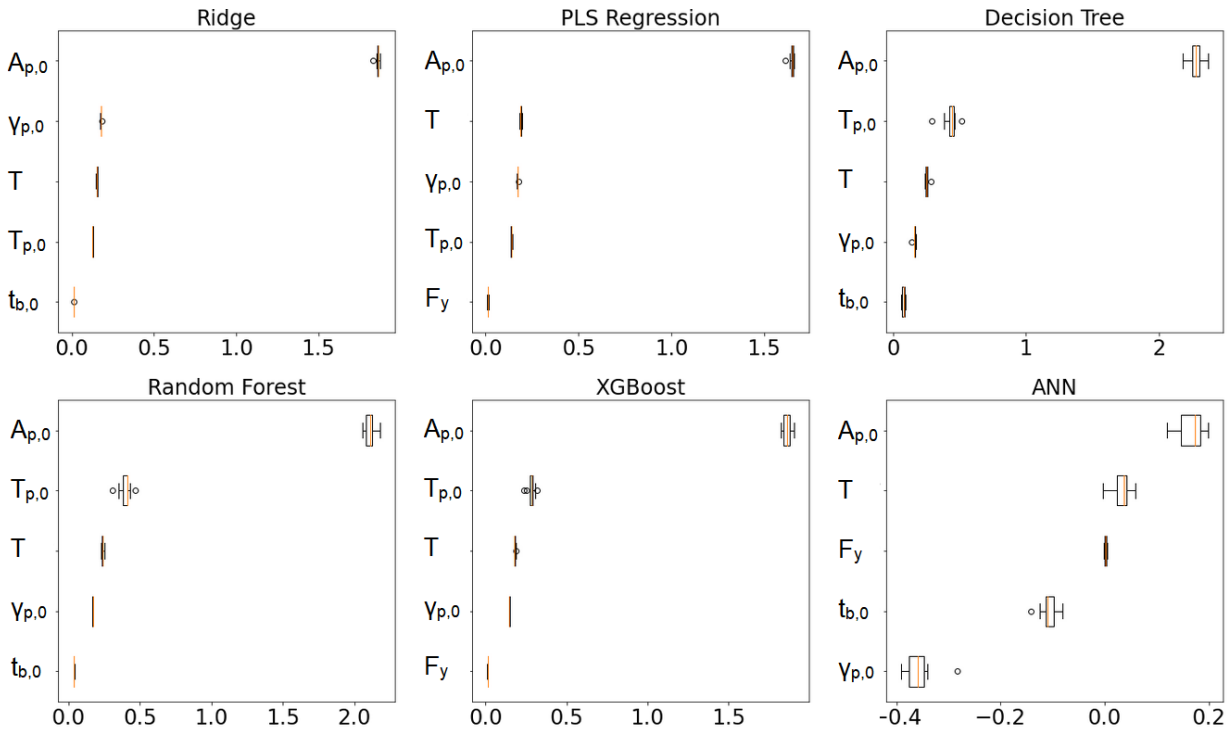


Figure 23: Permutation importance of the for each algorithm using 1 pulse.

Based on the earthquake engineering domain knowledge the most important features should be $A_{p,1}$, T and F_y. From Figure 23, it is shown that the models with the most acceptable feature importance are the Random Forest and the XGBoost model.

4.3.2. 10 pulses

The exact same training process is repeated but now keeping the parameters from all 10 pulses. So, the model uses 5 (parameters of each pulse) x 10 (pulses) + 2 (T and Fy) = 52 input features.

Table 5: Root Mean Square and Mean Absolute Error metrics for each ML algorithm using 10 pulses.

Algorithm	RMSE	MAE
Ridge	0.0200	0.0138
PLS Regression	0.0177	0.0124
Decision Tree	0.0225	0.0121
Random Forest	0.0151	0.0081
XGBoost	0.0160	0.0084
ANN	0.0132	0.0086

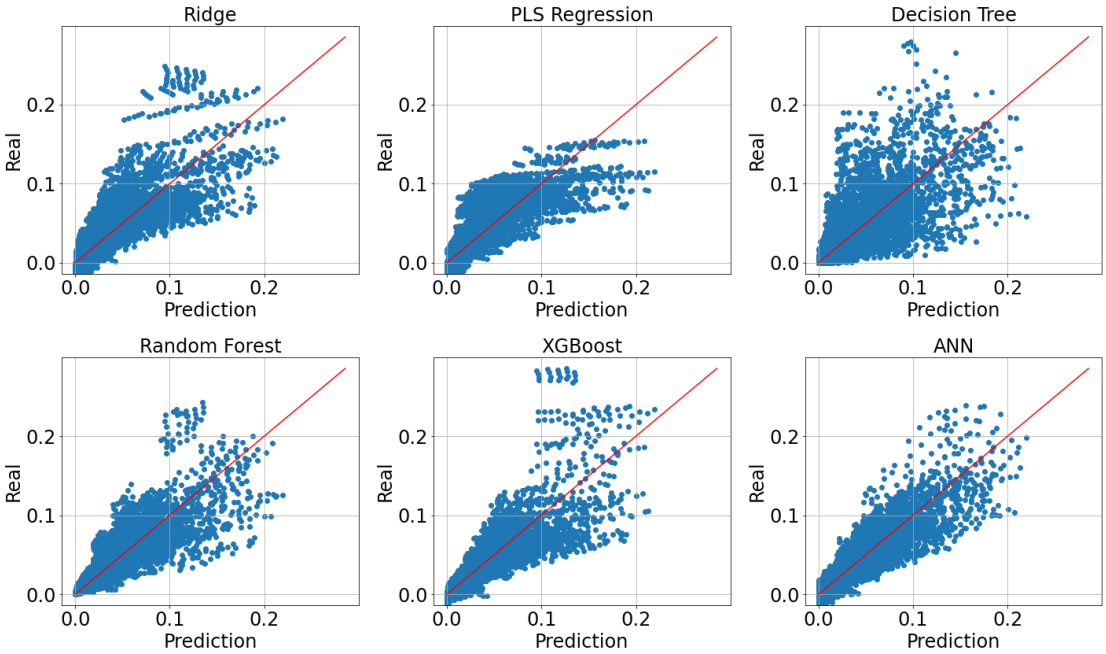


Figure 24: Scatter plots of the Predictions vs Real values of the max displacement using 10 pulses.

Again, based on the scatter plots in Figure 22 it is shown that all of the ML models have performed generally well and were able to give adequate prediction when using . Among them, the Random Forest and the Artificial Neural Network had the lowest RMSE and MAE metrics.

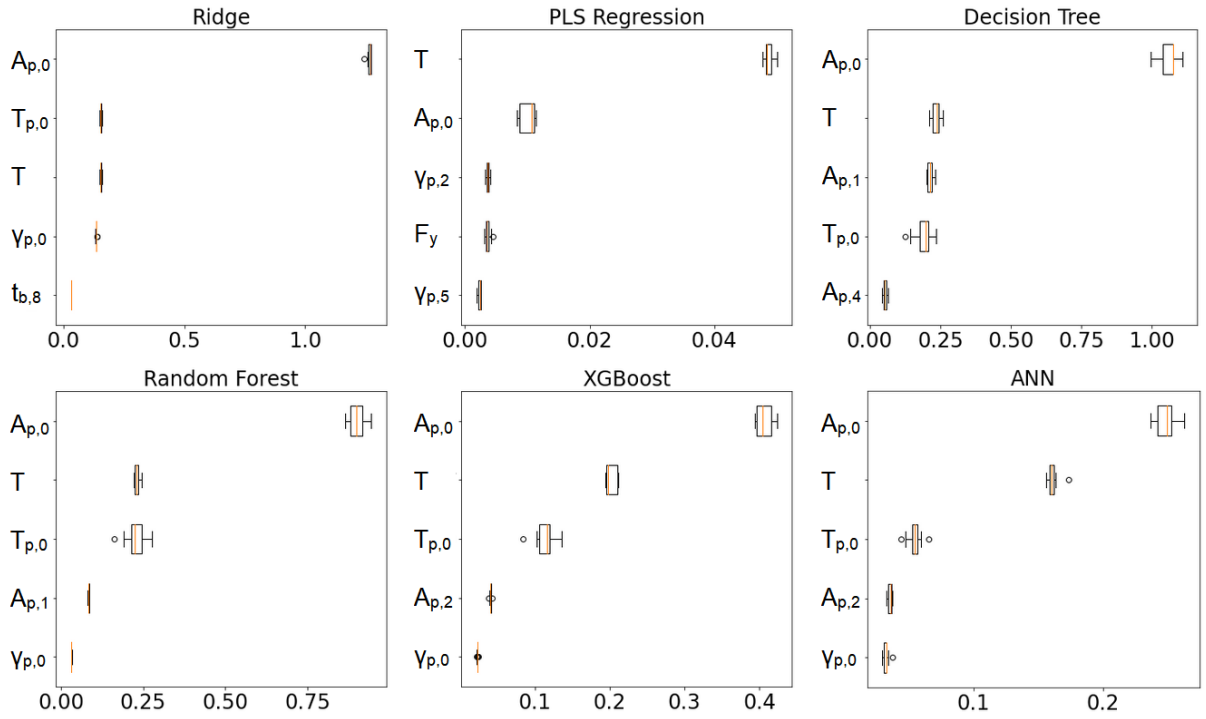


Figure 25: Permutation importance for each algorithm using 10 pulses

From the permutation importance above, it is shown that all models seem to prioritize the input features generally well, with the exception of the F_y which seems to have lesser importance for the models.

5. Application to Multi-DOF Structural Systems

In this section, the proposed methodology is extended to MDOF structural systems. A benchmark structure of four DOFs is used, a described in Figure 26.

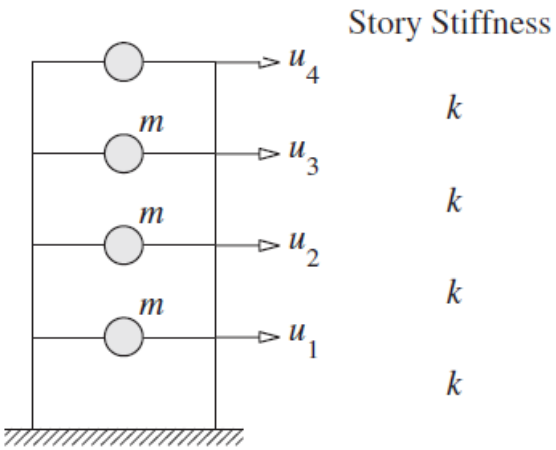


Figure 26: MDOF benchmark model

The max displacement for each ground motion record has been obtained through dynamic response analysis that was described in Section 2.3.2, using the OpenSees software. Then, the structural features, T and F_y, have been calculated using the Pushover analysis and the methodology described in Section 2.3.3. Finally, the predictions for the max displacement u_{max} were obtained by the models that were trained using 10 pulses for the representation of the ground motions. The error metrics are shown in Table 6 and the Real vs Predictions scatter plots are showin in Figure 26.

Table 6: Root Mean Square Error and Mean Absolute Error metrics for each ML model using 10 pulses.

Algorithm	RMSE	MAE
Ridge	0.0299	0.0182
PLS Regression	0.0368	0.0196
Decision Tree	0.0106	0.0027
Random Forest	0.0072	0.0028
XGBoost	0.0145	0.0086
ANN	0.0098	0.0067

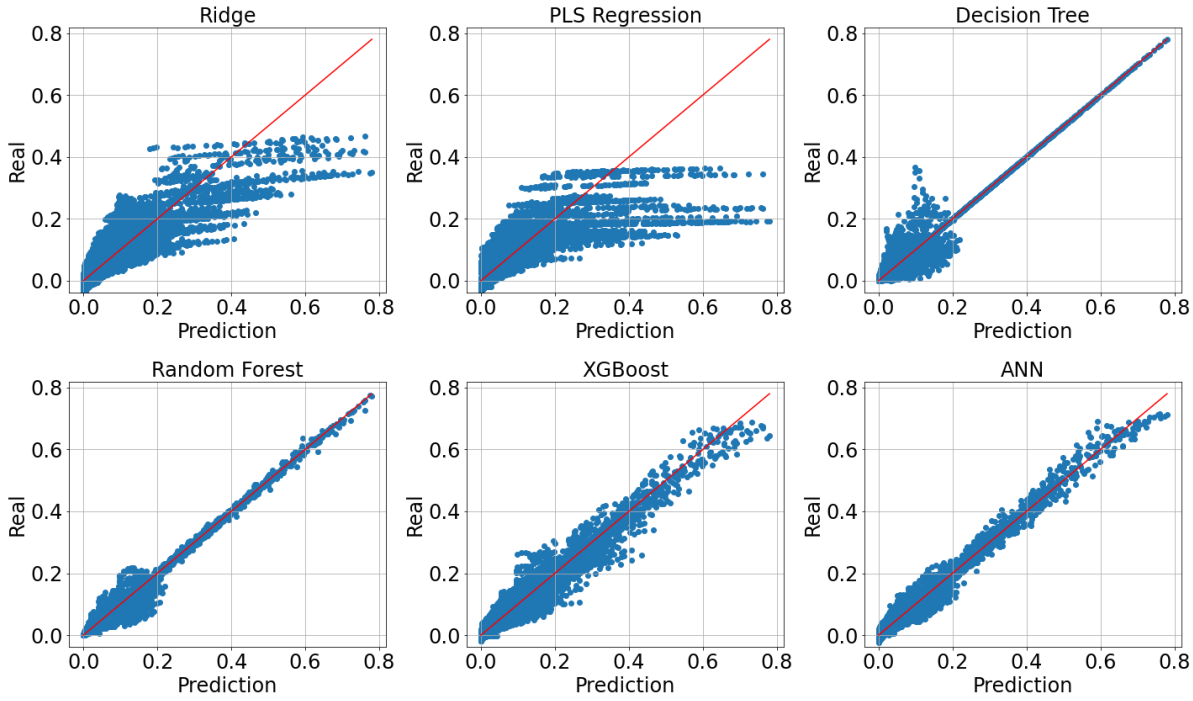


Figure 27: Scatter plots of the MDOF Predictions vs Real values of the max displacement using 10 pulses.

It is clearly shown, that all models performed really well, with the exception of Ridge and PLS Regression models. This is justified because their linear nature is not adequate to capture the nonlinearities of the problem. The model with the highest performance was the Random Forest model, achieving MAE lower than 3mm.

6. Conclusions and future work

As shown in this study ML models can be trained and give adequate predictions in terms of max displacement, both in SDOF structural systems. It is also shown based on the feature importance analysis that different machine-learning algorithms prioritize the input features differently, which plays a critical role in the final model selection.

The proposed methodology presented in this thesis can be extended and further developed for future work involving the following aspects:

- Artificial dataset can be generated also from results of MDOF structural systems. This incorporate to include additional features (i.e. number of DOFs, eigenperiods, etc.) to capture the structural behavior.
- Explore different ML algorithms for the model training.
- Neural Network tuning: different architectures and hyper-parameter optimization
- Dataset enhancement, by including additional ground motion records for training purposes.
- Incorporate additional input features (e.g. ground motion parameters)
- Explore other ground motion record representation techniques (e.g. using Convolutional Neural Networks)
- Use additional Machine Learning Interpretability methods (e.g. SHAP values explainers)

References

Faroughi and Hosseini (2011). Simplification of Earthquake Accelerograms for Quick Time History Analyses by using Their Modified Inverse Fourier Transforms. *Procedia Engineering,* **14**, 2872–2877.

Mavroeidis George P. and Papageorgiou Apostolos S. (2003). A Mathematical Representation of Near-Fault Ground Motions. *Bulletin of the Seismological Society of America, Vol. 93, No. 3, pp. 1099–1131.*

Salehia H. and Burgueñoa R. (2018). Emerging artificial intelligence methods in structural engineering, *Engineering Structures*, **171**, 170–189.

Static Pushover Analysis, ZSoilr.PC 070202 report.

Ζαννή Αναστασία (2020). Επεξεργασία σεισμικών καταγραφών κοντινού πεδίου και αντικατάσταση τους με ισοδύναμους παλμούς. Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο.

Ρεπούσης Νικηφόρος (2019). Μη γραμμικά μοντέλα. Διπλωματική εργασία, Εθνικό Μετσόβιο Πολυτεχνείο.

Permutation feature importance documentation (last visited 29/11/2021):

https://scikit-learn.org/stable/modules/permutation_importance.html