Bootstrapped Artificial Neural Networks for the seismic analysis of structural systems

E. Ferrario[1], N. Pedroni[2], E. Zio[2,3] and F. Lopez-Caballero[4]

[1]Laboratory LGI, Université Paris-Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France
[2]Chair System Science and the Energy Challenge, Fondation Electricité de France (EDF), CentraleSupélec, Université Paris-Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France
[3]Department of Energy, Politecnico di Milano, Italy
[4]Laboratory MSSMat, UMR CNRS 8579, CentraleSupélec, Université Paris-Saclay, Grande Voie des Vignes, 92290 Chatenay-Malabry, France

Abstract

We look at the behavior of structural systems under the occurrence of seismic events with the aim of identifying the fragility curves. Artificial Neural Network (ANN) empirical regression models are employed as fast-running surrogates of the (long-running) Finite Element Models (FEMs) that are typically adopted for the simulation of the system structural response. However, the use of regression models in safety critical applications raises concerns with regards to accuracy and precision. For this reason, we use the bootstrap method to quantify the uncertainty introduced by the ANN metamodel. An application is provided with respect to the evaluation of the structural damage (in this case, the maximal top displacement) of a masonry building subject to seismic risk. A family of structure fragility curves is identified, that accounts for both the (epistemic) uncertainty due to the use of ANN metamodels and the (epistemic) uncertainty due to the paucity of data available to infer the fragility parameters.

Keywords: seismic risk, structure, fragility curve, Artificial Neural Network, epistemic uncertainty, bootstrap, confidence intervals

Nomenclature

a : ground motion level

b : index of the bootstrap training data sets or of the bootstrapped regression models, $b = 1, …, B$

B : number of the bootstrap training data sets or of the bootstrapped regression models

CP : coverage probability

$D = \{(\boldsymbol{Y}_n, \delta_n), n = 1, …, N\}$: entire data set

$D_{train} = \{(\boldsymbol{Y}_n, \delta_{FEM}(\boldsymbol{Y}_n)), n = 1, …, N_{train}\}$: training data set

$D_{train,b}$: bootstrap training data set, $b \in \{1, …, B\}$

$D_{val} = \{(\boldsymbol{Y}_n, \delta_{FEM}(\boldsymbol{Y}_n)), n = 1, …, N_{val}\}$: validation data set

$D_{test} = \{(\boldsymbol{Y}_n, \delta_{FEM}(\boldsymbol{Y}_n)), n = 1, …, N_{test}\}$: test data set

E = network performance (energy function)

$f(\boldsymbol{Y}_n, \boldsymbol{w})$: regression function

$f_b(\boldsymbol{Y}_n, \boldsymbol{w}_b)$: bootstrapped regression function, $b \in \{1, …, B\}$

F : fragility curve

F_b : fragility curve built on the basis of the b-th bootstrapped regression function, $b \in \{1, …, B\}$

$\underline{F}$: lower bound fragility curve due to the paucity of data

$\bar{F}$: upper bound fragility curve due to the paucity of data

$\underline{\underline{F}}$: lower bound fragility curve due to the model and the paucity of data

$\bar{\bar{F}}$: upper bound fragility curve due to the model and the paucity of data

h : optimal number of hidden neurons

I_{Arias} : Arias intensity

j : index of the inputs $\boldsymbol{Y}$

L : likelihood function

M : number of input variables

n : index of the data in a given set

N : number of realization of the seismic event

NMW : normalized mean width

N_{test} : number of test data

N_{train} : number of training data

N_{val} : number of validation data

p : coverage indicator; $p = 1$ if the output is included in the confidence interval; $p = 0$ otherwise

pgv : Peak Ground Velocity

$PSA(T_{str})$: spectral acceleration at the first-mode period of the structure

$RMSE_{ANN}$: root mean square error of the ANN trained with the whole training data set D_{train}

$RMSE_{Boot}$: root mean square error of the bootstrap ensemble of ANNs

SI : spectral intensity

T_m : mean period

T_p : predominant period

$Tstr$: fundamental period of the structure

V_{s30} : average shear wave velocity in the upper 30 m

w : vector of parameters of the regression functions

w_b : vector of parameters of the b-th bootstrapped regression functions, $b \in \{1, …, B\}$

X : outcome of a seismic event (Bernoulli random variable)

x_n : realization of the Bernoulli random variable X_n

$\boldsymbol{x} = \{x_1, x_2, …, x_n, …, x_N\}$: vector of the realizations of the N Bernoulli random variable X_n, $n = 1, …, N$

y : ground motion IMs

$\boldsymbol{Y} = \{y_1, y_2, …, y_i, …, y_M\}$: vector of M uncertain input variables

$\boldsymbol{Z} = \{z_1, z_2, …\}$: vector of system responses

Greek Letters

α : median ground motion intensity measure (IM)

$\hat{\alpha}$: estimate of α

$\left[\underline{\alpha}^{100(1-\theta)\%}, \ \bar{\alpha}^{100(1-\theta)\%}\right]$: $100(1 - \theta)\%$ confidence interval of α

β : logarithmic standard deviation

$\hat{\beta}$: estimate of β

$\left[\underline{\beta}^{100(1-\theta)\%}, \ \bar{\beta}^{100(1-\theta)\%}\right]$: $100(1 - \theta)\%$ confidence interval of β

$1 - \gamma$: level of confidence for the bootstrap-based empirical PDFs

δ : target, maximal structural top displacement

$\delta(\boldsymbol{Y}_n)$: model output (maximal structural top displacement) in correspondence of the n-th input vector $\boldsymbol{Y}_n$

$\delta_{ANN}(\boldsymbol{Y}_n)$: estimate of the maximal structural top displacement obtained by the ANN

$\delta_{Boot_b}(\boldsymbol{Y}_n)$: estimate of the maximal structural top displacement given by one of the b-th bootstrapped regression functions, $b \in \{1, …, B\}$

$\bar{\delta}_{Boot}(\boldsymbol{Y}_n)$: average of the B estimates $\delta_{Boot_b}(\boldsymbol{Y}_n)$, $b = 1, …, B$

$\delta_{FEM}(\boldsymbol{Y}_n)$: true maximal structural top displacement computed by the FEM

δ^* : damage threshold

$\left[\underline{\delta}^{100(1-\gamma)\%}(\boldsymbol{Y}_n), \ \bar{\delta}^{100(1-\gamma)\%}(\boldsymbol{Y}_n)\right]$: $100(1 - \gamma)\%$ confidence interval of the quantity $\delta(\boldsymbol{Y}_n)$

$\varepsilon(\boldsymbol{Y}_n)$: Gaussian white noise

$1 - \theta$: level of confidence for parameters α and β

$\mu_\delta(\boldsymbol{Y}_n)$: nonlinear deterministic function

$\sigma_{Boot}^2(\boldsymbol{Y}_n)$: bootstrap estimate of the variance of $\sigma_f^2(\boldsymbol{Y_n})$

$\sigma_f^2(\boldsymbol{Y}_n)$: variance of the distribution of the regression function $f(\boldsymbol{Y}_n, w)$

$\sigma_\varepsilon^2(\boldsymbol{Y}_n)$: variance of $\varepsilon(\boldsymbol{Y}_n)$

$\Phi[\cdot]$: standard Gaussian cumulative distribution

Acronyms

ANN : Artificial Neural Network

CDF : Cumulative Distribution Function

CP : Coverage Probability

FEM : Finite Element Model

GA : Genetic Algorithm

IM : Intensity Measure

LGP : Local Gaussian Process

NMW : Normalized Mean Width

NPP : Nuclear Power Plant

PGA : Peak Ground Acceleration

PDF : Probability Density Function

RMSE : Root Mean Square Error

RS : polynomial Response Surface

SA : Spectral Acceleration

SPRA : Seismic Probabilistic Risk Assessment

SVMs : Support Vector Machines

1. Introduction

In the aftermath of the Fukushima nuclear accident, the 5-year project SINAPS@ (Earthquake and Nuclear Facilities: Ensuring and Sustaining Safety) has been launched in France in 2013. One of the key objectives of the project is the quantitative assessment of the behavior of Nuclear Power Plants (NPPs) under the occurrence of a seismic event. In the framework of this project, we made a preliminary study on the behavior of a structural system subject to seismic risk [Ferrario et al., 2015], with the aim of identifying the structure fragility curve, i.e., the conditional probability of damage of a component for any given ground motion level [EPRI, 2003]. In this work, we complete the previous analysis with the estimation of the associated uncertainties.

Within the framework of analysis considered, in general the actions, events and physical phenomena that may cause damages to a nuclear (structural) system are described by complex mathematical models, which are then implemented into computer codes to simulate the behavior of the system of interest under various conditions [USNRC, 2009; NASA, 2010]. In particular, computer codes based on Finite Element Models (FEMs) are typically adopted for the simulation of the system structural behavior and response: an example is represented by the Gefdyn code [Aubry et al., 1986].

In practice, not all the system characteristics can be fully captured in the mathematical model. As a consequence, uncertainty is always present both in the *values of the model input parameters and variables* and in the *hypotheses supporting the model structure*. This translates into variability in the model outputs, whose uncertainty must be estimated for a realistic assessment of the (seismic) risk [Ellingwood and Kinali, 2009; Liel et al., 2009].

For the treatment of uncertainty in risk assessment, it is often convenient to distinguish two types: randomness due to inherent variability in the system behavior (aleatory uncertainty) and imprecision due to the lack of knowledge and information on the system (epistemic uncertainty). The former is related to random phenomena, like the occurrence of unexpected events (e.g., earthquakes) whereas the latter arises from a lack of knowledge of some phenomena and processes (e.g., the power level in the nuclear reactor), and/or from the paucity of operational and experimental data available [Apostolakis, 1990; Ferson and Ginzburg, 1996; USNRC, 2009; Aven, 2010; Aven and Zio, 2011 Beer et al., 2014].

For uncertainty characterization, two issues need to be considered: first, the assessment of the system behavior typically requires a very large number (e.g., several hundreds or thousands) of FEM simulations under many different scenarios and conditions, to fully explore the wide range of uncertainties affecting the system; second, FEMs are computationally expensive and may require

hours or even days to carry out a single simulation. This makes the computational burden associated with the analysis impracticable, at times.

In this context, fast-running regression models, also called metamodels (such as Artificial Neural Networks (ANNs) [Zio, 2006; Cardoso et al., 2008; Beer and Spanos, 2009; Pedroni et al., 2010; Chojaczyk et al., 2015], Local Gaussian Processes (LGPs) [Bichon et al., 2008; Villemonteix et al., 2009] polynomial Response Surfaces (RSs) [Bucher and Most, 2008; Liel et al., 2009], polynomial chaos expansions [Ciriello et al., 2013; Sudret and Mai, 2015], stochastic collocations [Babuska et al., 2010], Support Vector Machines (SVMs) [Hurtado, 2007] and kriging [Bect et al., 2012; Dubourg and Sudret, 2014; Zhang et al., 2015]), can be built by means of input-output data examples to approximate the response of the original long-running FEMs without requiring a detailed physical understanding and modeling of the system process, and used for the seismic analysis. Since the metamodel response is obtained quickly, the problem of high computational times is circumvented. However, the use of regression models in safety critical applications like NPPs still raises concerns as regards the control of their accuracy and precision.

In this work, we use ANN-based metamodels to approximate the response of a detailed FEM. To evaluate the approximation introduced by using the ANN metamodels in replacement of the detailed FEM, we can adopt three main approaches for the estimation of ANN accuracy and precision: the bootstrap method, based on a resampling technique [Efron, 1979; Efron and Tibshirani, 1993]; the delta method, based on a Taylor expansion of the regression function [Rivals and Personnaz, 1998; Dybowski and Roberts, 2001]; the Bayesian approach, based on Bayesian statistics [Bishop, 1995; Dybowski and Roberts, 2001]. We adopt the first one, because it requires no prior knowledge about the distribution function of the underlying population, being a distribution-free inference method [Efron and Thibshirani, 1993], it is simple to implement and provides accurate uncertainty estimates [Zio, 2006].

The effectiveness of ANN and bootstrap methods in robustly quantifying the uncertainty associated with (safety-related) estimates (possibly obtained with small-sized data sets) has been thoroughly demonstrated in the open literature: see, e.g., [Efron and Thibshirani, 1993; Zio, 2006; Pedroni et al., 2010; Zio and Pedroni, 2011].

In this work, we originally apply bootstrapped ANNs in seismic risk analysis for the estimation of fragility curves and the quantification of the corresponding uncertainty. For the analysis, we consider the same case study employed in [Ferrario et al., 2015], which consider the structural damages of a masonry structure subject to a seismic event [Lopez-Caballero et al., 2011]. First, we use the ANN metamodel to estimate the maximal structural top displacement, and its uncertainty in the form of

bootstrap-based confidence intervals. Then, we derive the corresponding family of fragility curves for a given damage threshold, taking into account both the variability of the maximal structural top displacement within the confidence intervals and the epistemic uncertainty due to paucity of data available for the estimation of the fragility curve parameters.

The paper organization is as follows. In Section 2, the steps necessary to build the fragility curves are illustrated and the sources of uncertainty identified; in Section 3, a synthetic description of ANN metamodels is provided; in Section 4, the uncertainty introduced by the ANN metamodels is described; in Section 5, the bootstrap approach for the estimation of the ANN uncertainty is given; in Section 6, the procedure for the bootstrapped ANN-based estimation of fragility curves in the presence of uncertainties is provided; in Section 7, the case study and the main results of the analysis are presented; in Section 8, conclusions and future developments are provided.

2. Analysis of the failure behavior of structural systems: fragility curve estimation

Within the framework of Seismic Probabilistic Risk Assessment (SPRA), a fragility curve, F, is the conditional probability of failure (i.e., of exceeding a level of damage) of a component/structure for any given ground motion level (i.e., peak ground acceleration – PGA – or peak spectral acceleration – SA – at different component or structural frequencies) [EPRI, 2003]. It represents in probabilistic terms the seismic capacity of a component/structure, in terms of the maximum ground motion level (PGA or SA) that a component/structure can sustain without failure. Both PGA and SA are good indicators of structural damage; however, most SPRAs are based on PGA since most existing hazard studies have focused primarily on it [EPRI, 2003].

It is standard practice to model the seismic capacity by a lognormal probability distribution and evaluate its parameters for critical failure modes by the Fragility Analysis [EPRI, 2003; Ulrich et al., 2014], though any measure of ground motion intensity can be used. The use of the lognormal distribution is justified by its mathematical simplicity and by the fact that it has been widely used for several decades in earthquake engineering: actually, it often fits the observed distributions of quantities of interest, such as ground motion intensity measures (IMs) [Porter, 2015].

In this paper, we retain the lognormal model assumption and we infer the associated parameters by the maximum likelihood method, taking into account the epistemic uncertainty in their estimates due to the paucity of data available.

In particular, we construct the fragility curve through the following three main steps:

    (a) structural system modeling;

(b) structural system behavior simulation;

(c) fragility curve estimation.

In step (a), a mathematical model of the system is built to quantify its performance indicator. A quantitative model for seismic risk analysis may be viewed as composed of three main elements: a vector $Y = \{y_1, y_2, \ldots, y_j, \ldots, y_M\}$ containing all the uncertain input variables (e.g., the ground motion IMs); a computer code to simulate the behavior of the system of interest; and an output vector $Z = \{z_1, z_2, \ldots\}$ describing the system responses (e.g., the maximal structural top displacement).

In step (b), the mathematical model is implemented in a computer code and used to simulate system behavior under different uncertain operational and accidental conditions: for N vectors Y_n, $n = 1, \ldots, N$, of possible model input values (i.e., N seismic events in this case), N model outputs $\delta(Y_n)$, $n = 1, \ldots, N$, are computed. In order to deeply explore the wide range of uncertainties, a very large number (e.g., several thousands) of simulations is typically needed.

Finally, in step (c) the fragility curve F is estimated from the data (i.e., the N maximal structural top displacements $\delta(Y_n)$, $n = 1, \ldots, N$) generated at step (b) above.

The mathematical formulation of the fragility is the following [EPRI, 2003]:

$$F = P(\delta > \delta^*|a) = \Phi\left[\frac{1}{\beta}log\left(\frac{a}{\alpha}\right)\right], \tag{1}$$

where δ is the maximal structural top displacement, δ^* is the damage threshold, a is the ground motion level (e.g., PGA), $\Phi[\cdot]$ is the standard Gaussian cumulative distribution function (CDF), and α and β are the parameters of the lognormal distribution of the seismic capacity, where α is the median ground motion IM, i.e., the median ground acceleration capacity (e.g., median PGA), and β the logarithmic standard deviation.

Given the model outputs $\delta(Y_n)$, $n = 1, \ldots, N$, of step (b) above, the parameters α and β in equation (1) can be estimated through the maximum likelihood method [Shinozuka et al., 2000; Saez et al., 2011] by analyzing those outputs $\delta(Y_n)$, $n = 1, \ldots, N$, that exceed the damage threshold of interest δ^*. The likelihood function L that has to be maximized to infer parameters α and β is:

$$L(x; \alpha, \beta) = \prod_{n=1}^{N} [F_n]^{x_n}[1 - F_n]^{1-x_n}, \tag{2}$$

where $x = \{x_1, x_2, \ldots, x_n, \ldots, x_N\}$ is the vector of the N realizations of a Bernoulli random variable X representing the outcome of a seismic event that can be classified into two mutually exclusive ways: $x_n = 0$, if the component/structure sustains the damage threshold δ^* (i.e., $\delta(Y_n) < \delta^*$) for a given ground motion level a, and $x_n = 1$, otherwise (i.e., $\delta(Y_n) > \delta^*$); F_n, $n = 1, \ldots, N$, is the fragility value in correspondence of the n-th seismic event.

Two sources of uncertainty are involved in step (c), due to:

- aleatory uncertainty due to randomness in the structural seismic capacity, described by a lognormal distribution;

- epistemic uncertainty due to the finite amount of data available to infer the parameters (α and β) of the lognormal distribution.

The seismic capacity is characterized by aleatory uncertainty: actually, the structural response results from randomness in the ground motion IM; thus, even with a large data set available, this uncertainty is irreducible. The classical way to address the uncertainty due to randomness is to collect data about the random phenomenon of interest (i.e., the ground motion) and perform a statistical analysis to identify the probability density function (PDF) that best captures the variability of the available data. The analyst's choice of a particular shape of the PDF (i.e., lognormal) represents a source of epistemic uncertainty and can be progressively reduced by the analyst as the size of the available data set increases [Apostolakis, 1990]. In this work, we assume that the chosen lognormal PDF is a well-fitted model for the random seismic capacity (i.e., its shape acceptably represents the variability of the uncertain seismic capacity).

The (epistemic) uncertainty upon the distribution parameters (α and β) depends on the quantity of data available: the higher the number N of seismic events, the more accurate the estimates of the parameters. To account for this type of epistemic uncertainty, we consider the $(1 - \theta)\% = 95\%$ confidence intervals (i.e., $\theta = 0.05$) of the α and β estimates, i.e., $\left[\underline{\alpha}^{100(1-\theta)\%}, \ \bar{\alpha}^{100(1-\theta)\%} \right]$ and $\left[\underline{\beta}^{100(1-\theta)\%}, \ \bar{\beta}^{100(1-\theta)\%} \right]$, respectively. Letting the parameters vary in their corresponding confidence intervals, a family of fragility curves is thereby obtained. Then, the upper and lower bounds, $\underline{F}$ and $\bar{F}$, respectively, of these curves can be identified to provide a measure of the epistemic uncertainty associated with F due to the paucity of data.

A final remark is in order. Traditionally, computer codes based on FEMs are adopted for the simulation of structural systems behavior and response. Since FEMs are computationally expensive (e.g., they may require hours or even days to carry out a single simulation), the computational cost associated with the analysis may be prohibitive.

One possibility to overcome this computational issue is to resort to fast-running regression models (metamodels) to replace the detailed, long-running FEMs with acceptable approximation. In this work, we adopt ANNs (see Section 3) to reproduce the nonlinear relation between the vector $\mathbf{Y}$ of M inputs representing the different characteristics of the seismic event, like Peak Ground Velocity (*pgv*), Arias Intensity (*I_{Arias}*), Spectral Intensity (*SI*), etc., and one output, that is the maximal structural top

displacement δ, i.e., $\mathbf{Z} = \{z_1\} = \delta$. The estimates provided by ANNs are affected by uncertainty (see Section 4) that can be quantified in terms of confidence intervals by the bootstrap approach (see Sections 5 and 6).

3. Artificial Neural Networks for nonlinear regression

Given a finite set of input/output data examples (patterns), $D = \{(Y_n, \delta(Y_n)), n = 1, \ldots, N\}$ where Y_n is a vector of input variables, $Y_n = \{y_{1,n}, y_{2,n}, \ldots, y_{j,n}, \ldots, y_{M,n}\}$, and $\delta(Y_n)$ is the output target (assumed mono-dimensional for simplicity of illustration), it is possible to perform the task of nonlinear regression, i.e., to estimate the nonlinear relationship between the vector of input variables and the output target.

It can be assumed that the target $\delta(Y_n)$, $n \in \{1, \ldots, N\}$, is related to the input vector Y_n by an unknown nonlinear deterministic function $\mu_\delta(Y_n)$ corrupted by a Gaussian white noise $\varepsilon(Y_n)$ [Bishop, 1995], i.e,

$$\delta(Y_n) = \mu_\delta(Y_n) + \varepsilon(Y_n) ; \qquad \varepsilon(Y_n) \sim N(0, \sigma_\varepsilon^2(Y_n)) \qquad (3)$$

In the present case of damage of a structural system subject to seismic risk, the vector Y_n, $n \in \{1, \ldots, N\}$, contains the ground motion IMs, the nonlinear deterministic function $\mu_\delta(Y_n)$ is given by the complex long-running FEM code, the variable $\delta(Y_n)$ represents the maximal structural top displacement and the noise $\varepsilon(Y_n)$ represents the error introduced by the numerical method employed to calculate $\mu_\delta(Y_n)$.

The objective of the regression task by ANN is to estimate $\mu_\delta(Y_n)$ through a regression function $f(Y_n, w)$, that depends on the set of ANN parameters w. For the sake of clarity, in the rest of the paper we will refer to δ_{ANN} and δ_{FEM} to indicate the ANN and the FEM outputs, respectively.

ANNs are computing devices inspired by the function of the nerve cells in the brain [Bishop, 1995]. They are composed of many parallel computing units (called neurons or nodes) arranged in different layers and interconnected by weighed connections (called synapses). Each of these computing units performs few simple operations and communicates the results to its neighbouring units. From a mathematical viewpoint, ANNs consist of a set of nonlinear (e.g., sigmoidal) basis functions with adaptable parameters w that are adjusted by a training process on many different input/output data examples available in the data set $D_{train} = \{(Y_n, \delta_{FEM}(Y_n)), n = 1, \ldots, N_{train}\}$, where D_{train} is a subset of the entire data set D, as explained at the end of this Section. The training process is an iterative process of regression error minimization [Rumelhart et al., 1986]; it consists in optimizing the network performance E computed as, e.g., the average squared error between the network outputs $\delta_{ANN}(Y_n)$ and the targets $\delta_{FEM}(Y_n)$, $n = 1, \ldots, N_{train}$:

$$E = \frac{1}{N_{train}} \sum_{n=1}^{N_{train}} \left(\delta_{ANN}(\mathbf{Y}_n) - \delta_{FEM}(\mathbf{Y}_n) \right)^2 \qquad (4)$$

ANNs have been demonstrated to be universal approximants of continuous nonlinear functions (under mild mathematical conditions) [Cybenko, 1989], i.e., in principle, an ANN model with a properly selected architecture can be a consistent estimator of any continuous nonlinear function. The particular type of ANN considered in this paper is the classical feed-forward ANN composed of three layers (input, hidden and output, see Figure 1) and trained by the error back-propagation algorithm.
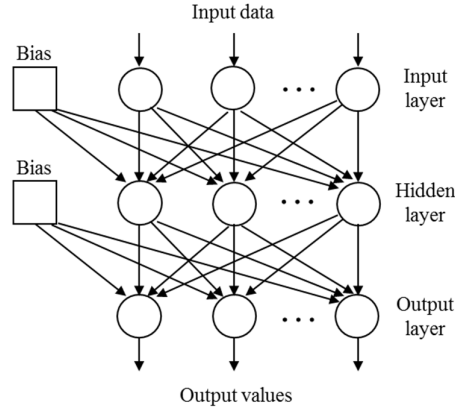

Figure 1. Scheme of a three-layered feed-forward Artificial Neural Network.

The number of nodes in the input layer equals the number of M input variables $\mathbf{Y}_n = \{y_{1,n}, y_{2,n}, \ldots, y_{j,n}, \ldots, y_{M,n}\}$, which significantly affect the output; the number of nodes in the output layer is determined by the number of quantities of interest to the problem (in this case one: the maximal structural top displacement, $\delta(\mathbf{Y}_n)$); the number of nodes in the hidden layer in general is kept as low as possible, since the higher the number of hidden nodes, the higher the number of parameters ($\mathbf{w}$) to be estimated and the stronger the requirements on the data needed for the estimation. In general, an ANN with too few hidden nodes does not succeed in learning the training data set; vice versa, an ANN with too many hidden nodes learns the training data set very well, but it does not have generalization capability.

Typically, the entire set of N input-output data is divided into three subsets:

- a training (input/output) data set ($D_{train} = \{(\mathbf{Y}_n, \delta_{FEM}(\mathbf{Y}_n)), n = 1, \ldots, N_{train}\}$), used to calibrate the adjustable parameters ($\mathbf{w}$) of the ANN regression model, for best fitting the FEM data;

- a validation (input/output) data set ($D_{val} = \{(\mathbf{Y}_n, \delta_{FEM}(\mathbf{Y}_n)), n = 1, \ldots, N_{val}\}$, made of patterns different from those of the training set), used to monitor the accuracy of the ANN model during the training procedure. In practice, the validation error is computed on the validation set at different iterative stages of the training procedure: at the beginning of training, this value

11

decreases as does the error computed on the training set; later in the training, if the ANN regression model starts overfitting the data, the error calculated on the validation set starts increasing and the training process must be stopped [Bishop, 1995];

- a test (input/output) data set ($D_{test} = \{(Y_n, \delta_{FEM}(Y_n)), n = 1, \ldots, N_{test}\}$), not used during ANN training and validation, needed to evaluate the network generalization capability in the presence of new data.

Further details about ANN regression models are not reported here for brevity; the interested reader may refer to the cited references and the copious literature in the field.

4. Artificial Neural Network uncertainty

In regression models, there are two types of prediction that are computed in correspondence of a given input Y_n, $n \in \{1, \ldots, N\}$: 1) the model structure $f(Y_n, w)$, that is the estimate of the deterministic function $\mu_\delta(Y_n)$, and 2) the model output $\delta(Y_n)$ (eq. 3). In this work, we focus on the first one and associate to this estimate the corresponding measure of confidence.

In this respect, the uncertainty related to the identification of the parameters w has to be taken into account. This is mainly due to three reasons:

- The set D_{train} adopted to train the network cannot be exhaustive due to the limited number of input/output patterns that do not fill all the input volume of interest. As a consequence, different training sets D_{train} can give rise to different sets of network parameters w and, then, to different sets of regression functions, i.e., to a distribution of regression functions. The quantification of the accuracy of the estimate $f(Y_n, w)$ of the true deterministic function, $\mu_\delta(Y_n)$, entails the assumption of a distribution for the *regression error $f(Y_n, w) - \mu_\delta(Y_n)$*. Its variance, $E[f(Y_n, w) - \mu_\delta(Y_n)]^2$, with respect to all possible training data set is given by the sum of two terms [Bishop, 1995; Zio, 2006]:
  - the square of the bias, $\{E[f(Y_n, w)] - \mu_\delta(Y_n)\}^2$, that is the square of the difference between the expected value of the distribution of the regression functions, $E[f(Y, w)]$, and the true deterministic function, $\mu_\delta(Y_n)$;
  - the variance, $\sigma_f^2(Y_n)$, of the distribution of the regression functions:

$$\sigma_f^2(Y_n) = \{f(Y_n, w) - E[f(Y_n, w)]\}^2 \,. \tag{5}$$

Since the bias can be considered negligible with respect to the second term, the accuracy of the estimate $f(Y_n, w)$ is given by the variance $\sigma_f^2(Y_n)$ of the distribution of the regression functions [Bishop, 1995; Zio, 2006]:

$$E[f(Y_n, w) - \mu_\delta(Y_n)]^2 = \sigma_f^2(Y_n) \tag{6}$$

- The choice of the network architecture is inappropriate. If the number of hidden neurons is too small or too big, the network may not have the expected generalization capabilities (see Section 3). A trade-off can be achieved by controlling the number of parameters, and the training procedure, e.g., by early stopping of the training [Bishop, 1995; Zio, 2006].
- The global minimum of the error function (eq. 4) is not achieved: the minimization algorithm may get stuck in a local minimum and/or the training may be stopped before reaching the minimum [Bishop, 1995; Zio, 2006].

Thus, due to the uncertainties, for given model input parameters/variables Y_n, $n \in \{1, \ldots, N\}$, the model output $\delta(Y_n)$ can vary within a range of possible values (e.g., confidence interval). For N vectors Y_n, $n = 1, \ldots, N$, of model inputs (i.e., N seismic events in this case), N model outputs $\delta(Y_n)$, $n = 1, \ldots, N$, can be computed with the associated N confidence intervals $\left[\underline{\delta}^{100(1-\gamma)\%}(Y_n), \; \bar{\delta}^{100(1-\gamma)\%}(Y_n) \right]$, $n = 1, \ldots, N$, with $(1 - \gamma)$ level of confidence.

5. Bootstrap approach for ANN uncertainty estimation

To quantify the model uncertainty introduced by the ANN empirical regression models (see Section 4), the confidence intervals of the regression error $f(Y_n, w) - \mu_\delta(Y_n)$, where $f(Y_n, w)$ is the regression function given by the ANN and $\mu_\delta(Y_n)$ is the deterministic function of the FEM, have to be estimated. One method is the bootstrap approach, that is capable of quantifying the model uncertainty by considering an ensemble of ANNs built on different data sets that are sampled with replacement (bootstrapped) from the original one [Zio, 2006]. From each bootstrap data set, a bootstrapped regression model is built and the model output of interest can be computed. Different bootstrap data sets give rise to a distribution of regression functions and so to a PDF of the model output. Thus, the model uncertainty of the estimates provided by the ANNs can be quantified in terms of confidence intervals of the obtained model output PDF by the bootstrap algorithm. An important advantage of this approach is that it provides confidence intervals for a given model output, without making any model assumption (e.g., normality). However, the computational cost could be high when the training data set and the number of parameters in the regression model are large [Pedroni et al., 2010].

The operative steps to identify the confidence intervals of the distribution of the regression error are detailed in the following:

1) Divide the entire available data set of N input/output patterns into training, validation and test data sets, as $D_{train} = \{(Y_n, \delta_{FEM}(Y_n)), n = 1, \ldots, N_{train}\}$, $D_{val} = \{(Y_n, \delta_{FEM}(Y_n)), n = 1, \ldots, N_{val}\}$

and D_{test} = {$(Y_n, \delta_{FEM}(Y_n))$, $n = 1, \ldots, N_{test}$}, respectively, where the outputs $\delta_{FEM}(Y_n)$ are generated through the FEM code.

2) Generate B (e.g., $B = 100$ in this work) bootstrap training data sets, $D_{train,b}$, $b = 1, \ldots, B$, by sampling with replacement from the original training data set D_{train}. Each set $D_{train,b}$, $b \in \{1, \ldots, B\}$, is composed by the same number N_{train} of the original training data set; however, due to the sampling with replacement some of the input/output patterns in D_{train} will appear more than once in $D_{train,b}$, $b \in \{1, \ldots, B\}$, whereas some will not appear at all.

3) Build the bootstrapped regression models $f_b(Y_n, w_b)$, $b = 1, \ldots, B$, on the basis of the training data sets $D_{train,b}$, $b = 1, \ldots, B$, generated at the previous step 2) and the validation data set D_{val} = {$(Y_n, \delta_{FEM}(Y_n))$, $n = 1, \ldots, N_{val}$} of step 1) (Figure 2). The training data sets are used to calibrate the parameters w_b, $b = 1, \ldots, B$, and the validation data set to monitor the accuracy of the network performance (see Section 3).

4) Use the regression models of step 3) to compute the estimates $\delta_{Boot_b}(Y_n)$, $b = 1, \ldots, B$, of the model output $\delta_{FEM}(Y_n)$ on a new data set D_{test} = {$(Y_n, \delta_{FEM}(Y_n))$, $n = 1, \ldots, N_{test}$}. By so doing, N_{test} bootstrap-based empirical PDFs for the quantity $\delta_{FEM}(Y_n)$, $n = 1, \ldots, N_{test}$, are produced. In correspondence of a new input Y_n, $n \in \{1, \ldots, N_{test}\}$, the bootstrap estimate $\bar{\delta}_{Boot}(Y_n)$ is given by the average of the B regression functions $\delta_{Boot_b}(Y_n)$, $b = 1, \ldots, B$, $n \in \{1, \ldots, N_{test}\}$:

$$\bar{\delta}_{Boot}(Y_n) = f(Y_n, \{w_b, b = 1, 2, \ldots, B\}) = \frac{\sum_{b=1}^{B} f_b(Y_n, w_b)}{B} = \frac{\sum_{b=1}^{B} \delta_{Boot_b}(Y_n)}{B} \qquad (7)$$

and the bootstrap estimate $\sigma_{Boot}^2(Y_n)$ of the variance $\sigma_f^2(Y_n)$ in equation (5) is given by:

$$\sigma_{Boot}^2(Y_n) = \frac{\sum_{b=1}^{B} \left[\delta_{Boot_b}(Y_n) - \bar{\delta}_{Boot}(Y_n)\right]^2}{B-1}, n \in \{1, \ldots, N_{test}\} \qquad (8)$$

as illustrated in Figure 3.

5) Identify the confidence intervals (with confidence level $(1 - \gamma)$) from the bootstrap-based empirical PDF for the quantity $\delta_{FEM}(Y_n)$, $n \in \{1, \ldots, N_{test}\}$, obtained in step 4) as follows. For each test $n \in \{1, \ldots, N_{test}\}$ compute the two-sided $100(1 - \gamma)\%$ confidence intervals for the estimates $\delta_{Boot_b}(Y_n)$, $b = 1, \ldots, B$, by ordering the estimates $\delta_{Boot_b}(Y_n)$, $b = 1, \ldots, B$, in increasing values and identifying the $(100\gamma/2)$th and $100(1- \gamma/2)$th quantiles of the bootstrap-based empirical PDF of the quantity $\delta_{FEM}(Y_n)$ as the closest values to the $(B\gamma/2)$th and $B(1 - \gamma/2)$th elements, respectively. In the following, the confidence interval of $\delta_{FEM}(Y_n)$, $n \in \{1, \ldots, N_{test}\}$, are referred to as $\left[\underline{\delta}^{100(1-\gamma)\%}(Y_n), \bar{\delta}^{100(1-\gamma)\%}(Y_n)\right]$.

To determine the quality of the confidence intervals identified in step 5), two quantitative measures can be taken into account [Ak et al., 2013]: the coverage probability (CP) and the normalized mean

width (NMW). The former represents the probability that the true values $\delta_{FEM}(Y_n)$, $n = 1, \ldots, N_{test}$, are included in the corresponding confidence intervals, whereas the latter quantify the extension of the confidence intervals and it is usually normalized with respect to the minimum and maximum values of $\delta_{FEM}(Y_n)$, $n = \{1, \ldots, N_{train}\}$. They are conflicting measures since the wider NMW the larger CP, and in practice it is important to have large coverage and small width.

The CP is given by [Ak et al., 2013]:

$$CP = \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} p_n, \tag{9}$$

where $p_n = 1$ if $\delta_{FEM}(Y_n) \in \left[\underline{\delta}^{100(1-\gamma)\%}(Y_n), \ \bar{\delta}^{100(1-\gamma)\%}(Y_n) \right]$; otherwise $p_n = 0$, for $n = 1, \ldots$, N_{test}.

The NMW is computed as [Ak et al., 2013]:

$$NMW = \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \frac{\left[\bar{\delta}^{100(1-\gamma)\%}(Y_n) - \underline{\delta}^{100(1-\gamma)\%}(Y_n) \right]}{\delta_{max} - \delta_{min}} \tag{10}$$

where δ_{max} and δ_{min} are the maximum and minimum values of the FEM outputs, respectively.

The estimates given by the bootstrapped ANNs, $\bar{\delta}_{Boot}(Y_n)$, are in general more accurate than the estimate of the best ANN, $\delta_{ANN}(Y_n)$, in the bootstrap ensemble of ANNs (that is the one trained with the original data set, D_{train}, as shown in step 1) above) [Zio, 2006; Zio and Pedroni, 2011]. Actually, the ANNs ensemble has diverse (higher) generalization capabilities than the best ANN. This can be verified through the estimation of the Root Mean Square Error (RMSE) for the bootstrap ensemble of ANNs ($RMSE_{Boot}$) and the best ANN ($RMSE_{ANN}$) with respect to the true output $\delta_{FEM}(Y_n)$ of the FEM, as follows:

$$RMSE_{Boot} = \sqrt{\frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \left(\bar{\delta}_{Boot}(Y_n) - \delta_{FEM}(Y_n) \right)^2} \tag{11}$$

$$RMSE_{ANN} = \sqrt{\frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \left(\delta_{ANN}(Y_n) - \delta_{FEM}(Y_n) \right)^2} \tag{12}$$
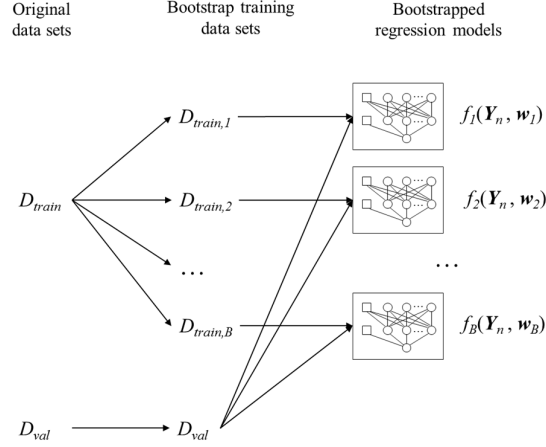
Figure 2: Exemplification of step 3 of the bootstrap method of Section 5: construction of the bootstrapped regression models given the training and validation data sets.
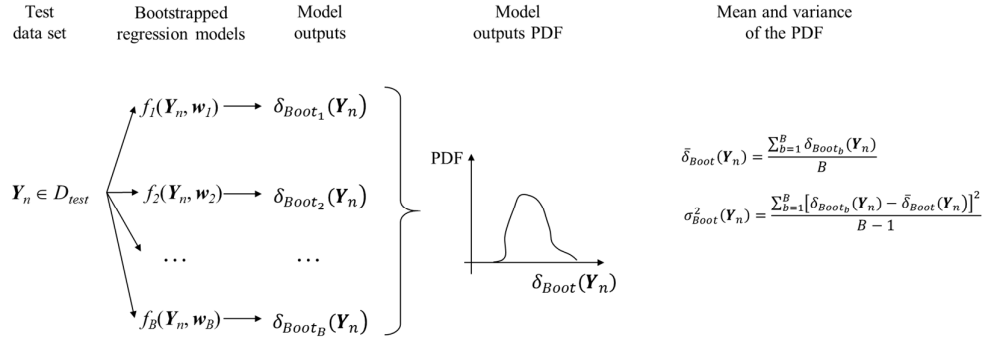
Figure 3: Exemplification of step 4 of the bootstrap method of Section 5: construction of the bootstrap-based empirical PDF of the model output for a new input Y_n of the test data set.

6. Bootstrapped ANN-based estimation of fragility curves in the presence of uncertainties

The complete procedure is as follows (Figure 4):

1) given an ensemble of B properly trained ANNs $\{f(Y_n \,|w_b): b = 1, 2, ..., B\}$, for N randomly generated vectors Y_n, $n = 1, 2, …, N$, of model inputs (i.e., N seismic events), evaluate the corresponding model outputs $\delta_{Boot_b}(Y_n)$, $b = 1, 2, …, B$, $n = 1, 2, …, N$ (see Section 5);

2) for each ANN $b = 1, 2, …, B$, of the ensemble, evaluate the parameters $\hat{\alpha}_b$ and $\hat{\beta}_b$ of the fragility curves and the corresponding CIs, $[\underline{\alpha_b}^{100(1-\theta)\%}(Y_n), \bar{\alpha}_b^{100(1-\theta)\%}(Y_n)]$ and $[\underline{\beta_b}^{100(1-\theta)\%}(Y_n), \bar{\beta}_b^{100(1-\theta)\%}(Y_n)]$, respectively, by MLE (see Section 2);

16

3) for each ANN of the ensemble, letting parameters α and β range within the corresponding CIs $[\underline{\alpha}_b^{100(1-\theta)\%}(\mathbf{Y}_n), \bar{\alpha}_b^{100(1-\theta)\%}(\mathbf{Y}_n)]$ and $[\underline{\beta}_b^{100(1-\theta)\%}(\mathbf{Y}_n), \bar{\beta}_b^{100(1-\theta)\%}(\mathbf{Y}_n)]$ (step 2 above), generate a family of fragility curves (under the lognormal assumption) and identify the corresponding upper and lower bounds, $\underline{F}_b$ and $\bar{F}_b$, respectively;

4) identify the extreme upper and lower fragility curves, $\underline{F}$ and $\bar{\bar{F}}$, respectively, that envelop *all* the bootstrapped fragility curves $\{(\underline{F}_b, \bar{F}_b): b = 1, 2, \dots, B\}$ thereby obtained.

The curves produced at step 4) represent bounds that describe and take into account the epistemic uncertainty due to

1) the data for the extrapolation of the distribution parameters α and β of the seismic capacity (step (c) of Section 2) and

2) the model (the ANN) for the estimation of the model output δ (step (b) of Section 2).

Notice that in this case the runs of the (bootstrapped) ANNs are only needed to generate a set of model outputs $\delta_{Boot_b}(\mathbf{Y}_n)$, $b = 1, 2, \dots, B$, $n = 1, 2, \dots, N$, for *each* bootstrapped ANN. Then, on the basis of these new sets of data (produced by each ANN b of the ensemble), the parameters $\hat{\alpha}_b$ and $\hat{\beta}_b$ of the fragility curves and the associated CIs, $[\underline{\alpha}_b^{100(1-\theta)\%}(\mathbf{Y}_n), \bar{\alpha}_b^{100(1-\theta)\%}(\mathbf{Y}_n)]$ and $[\underline{\beta}_b^{100(1-\theta)\%}(\mathbf{Y}_n), \bar{\beta}_b^{100(1-\theta)\%}(\mathbf{Y}_n)]$, can be estimated by MLE. Notice that for *each given* ANN $b = 1, 2, \dots, B$, of the bootstrap, the CIs $[\underline{\alpha}_b^{100(1-\theta)\%}(\mathbf{Y}_n), \bar{\alpha}_b^{100(1-\theta)\%}(\mathbf{Y}_n)]$ and $[\underline{\beta}_b^{100(1-\theta)\%}(\mathbf{Y}_n), \bar{\beta}_b^{100(1-\theta)\%}(\mathbf{Y}_n)]$ mentioned above reflect the (epistemic) uncertainty related to the scarcity of the data available for the extrapolation of the distribution parameters $\hat{\alpha}_b$ and $\hat{\beta}_b$.

Moreover, it is worth stressing again that the need to resort to ANNs (and in general to metamodels) derives from the very high computational cost required by each FEM run (i.e., several hours), which would impair *further*, *more thorough* uncertainty and sensitivity analyses on the system model of interest, if needed. Also, the additional advantage of adopting *bootstrapped* ANNs instead of a *single* ANN is represented by the possibility of computing CIs $[\underline{\delta}^{100(1-\gamma)\%}(\mathbf{Y}_n), \bar{\delta}^{100(1-\gamma)\%}(\mathbf{Y}_n)]$ for the model outputs, in order to take into account *also* the epistemic (model) uncertainty associated with the use of ANN regression functions trained with a *finite-sized* set of data (see Section 5).

A final remark is in order with respect to step (2) of the algorithm above. The fact that here we adopt a *parametric* approach, i.e., we assume a lognormal probability distribution for the fragility curve and use the MLE method for the estimation of the corresponding parameters, does *not* absolutely impair the effectiveness and *general* applicability of our bootstrapped ANN-based approach. Actually, once the metamodels are properly trained and validated, (i) they enable detailed uncertainty and sensitivity

analyses of computationally expensive FEMs for structural risk assessment (if required); and (ii) they could be used to build the fragility curves of interest according to any other method available in the open literature (e.g., *non-parametric* approaches relying on *repeated* Monte Carlo-based pointwise estimations of the conditional failure probability for *many different* values of the selected earthquake intensity measure).
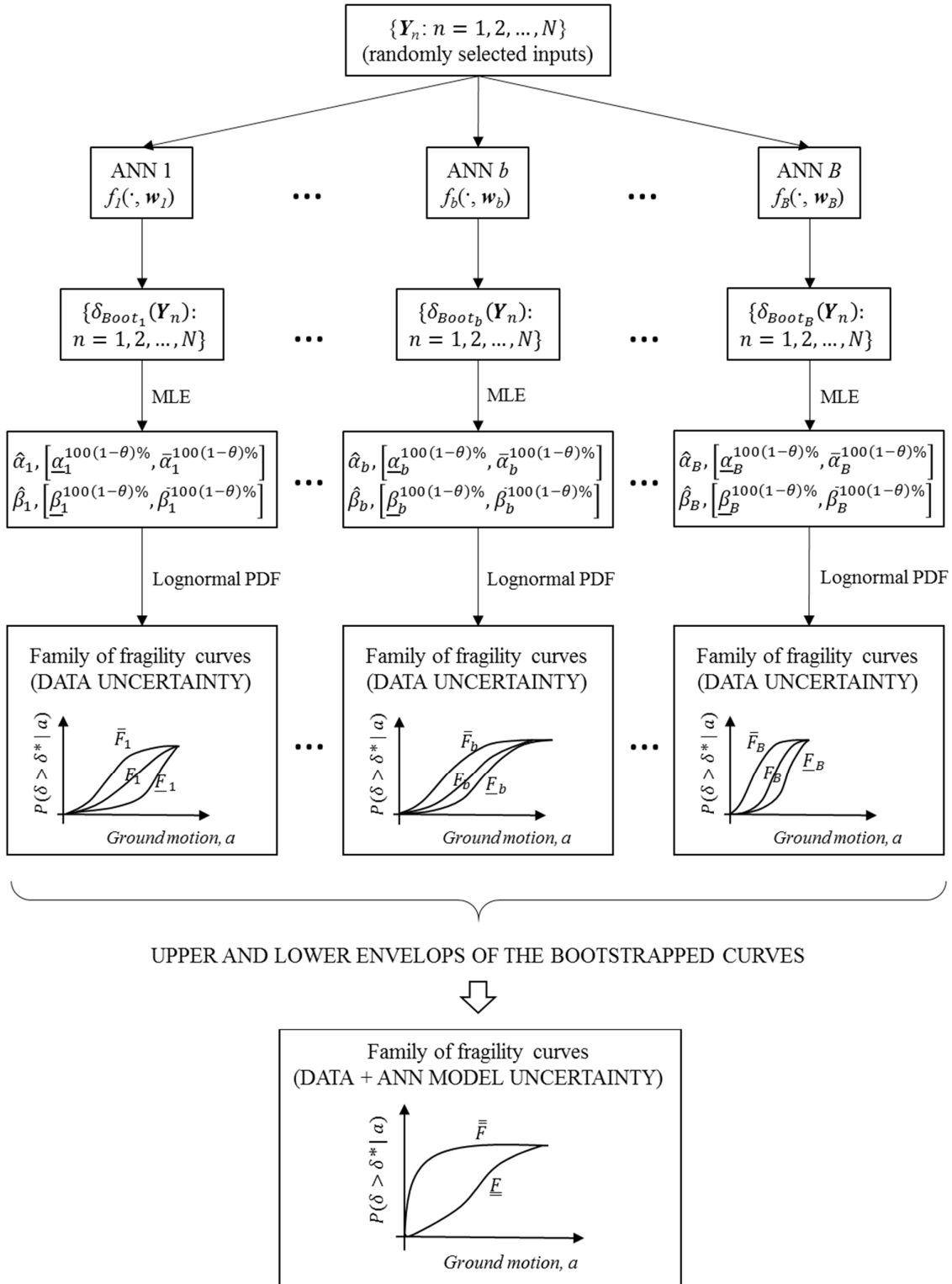
$\{\boldsymbol{Y}_n: n = 1, 2, \ldots, N\}$
(randomly selected inputs)

ANN 1
$f_1(\cdot, \boldsymbol{w}_1)$

$\cdots$

ANN b
$f_b(\cdot, \boldsymbol{w}_b)$

$\cdots$

ANN B
$f_B(\cdot, \boldsymbol{w}_B)$

$\{\delta_{Boot_1}(\boldsymbol{Y}_n):$
$n = 1, 2, \ldots, N\}$

$\cdots$

$\{\delta_{Boot_b}(\boldsymbol{Y}_n):$
$n = 1, 2, \ldots, N\}$

$\cdots$

$\{\delta_{Boot_B}(\boldsymbol{Y}_n):$
$n = 1, 2, \ldots, N\}$

MLE

MLE

MLE

$\hat{\alpha}_1, \left[\underline{\alpha}_1^{100(1-\theta)\%}, \bar{\alpha}_1^{100(1-\theta)\%}\right]$
$\hat{\beta}_1, \left[\underline{\beta}_1^{100(1-\theta)\%}, \bar{\beta}_1^{100(1-\theta)\%}\right]$

$\cdots$

$\hat{\alpha}_b, \left[\underline{\alpha}_b^{100(1-\theta)\%}, \bar{\alpha}_b^{100(1-\theta)\%}\right]$
$\hat{\beta}_b, \left[\underline{\beta}_b^{100(1-\theta)\%}, \bar{\beta}_b^{100(1-\theta)\%}\right]$

$\cdots$

$\hat{\alpha}_B, \left[\underline{\alpha}_B^{100(1-\theta)\%}, \bar{\alpha}_B^{100(1-\theta)\%}\right]$
$\hat{\beta}_B, \left[\underline{\beta}_B^{100(1-\theta)\%}, \bar{\beta}_B^{100(1-\theta)\%}\right]$

Lognormal PDF

Lognormal PDF

Lognormal PDF

Family of fragility curves
(DATA UNCERTAINTY)

$P(\delta > \delta^* | a)$ $\bar{F}_1$ F_1 $\underline{F}_1$ *Ground motion, a*

$\cdots$

Family of fragility curves
(DATA UNCERTAINTY)

$P(\delta > \delta^* | a)$ $\bar{F}_b$ F_b $\underline{F}_b$ *Ground motion, a*

$\cdots$

Family of fragility curves
(DATA UNCERTAINTY)

$P(\delta > \delta^* | a)$ $\bar{F}_B$ F_B $\underline{F}_B$ *Ground motion, a*

UPPER AND LOWER ENVELOPS OF THE BOOTSTRAPPED CURVES

Family of fragility curves
(DATA + ANN MODEL UNCERTAINTY)

$P(\delta > \delta^* | a)$ $\bar{\bar{F}}$ $\underline{E}$ *Ground motion, a*

Figure 4: Exemplification of the procedure for the bootstrapped ANN-based estimation of fragility curves in presence of uncertainty.

7. Case study

The case study is taken from [Lopez-Caballero et al., 2011]. It deals with the non-linear soil influence on the seismic response of a two-story masonry structure founded on a rigid shallow foundation. In Section 7.1, the description of the specific system studied is given; in Section 7.2, the results of the uncertainty analysis are provided, together with some critical considerations.

7.1. Case study description

The masonry building analyzed in this work is illustrated in Figure 5. The total height of the building is 5.4 m, the width is 5.0 m and the thickness is 0.16 m. With these characteristics, the fundamental period of the structure (*Tstr*) is equal to 0.19 s. This structure is modelled using three different kinds of elements: beam-columns and diagonal struts describing the structural behavior, and strengthless solid elements to represent the masonry mass. The frame structural elements are modelled by plastic hinge beam-column elements. The behavior of this structure is simulated on the basis of non-linear dynamic FE analysis. Further details about the masonry characteristics and the FEM used are not reported here for brevity sake; the interested reader is referred to [Lopez-Caballero et al., 2011].
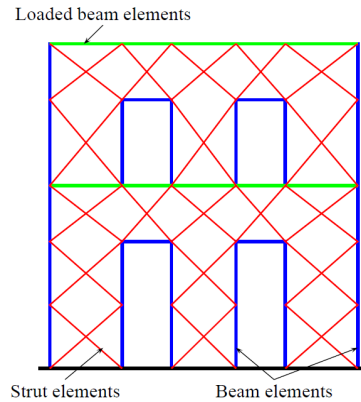

Figure 5: Building scheme.

In order to define appropriate ground motions for the non-linear dynamical analysis, a selection of *N* = 168 recorded accelerograms from the Pacific Earthquake Engineering Research Center (PEER) database have been used as input to the model. The events range between 5.2 and 7.6 in magnitude and the recordings have site-to-source distances from 15 to 50 km, and dense-to-firm conditions (i.e., 360 m/s $< V_{s30} <$ 800 m/s, where V_{s30} is the average shear wave velocity in the upper 30 m).

The information carried by each single earthquake signal has been synthetized by thirteen IM parameters [Lopez-Caballero et al., 2011]. In this work, we take into account just the five IM parameters reported in Table 1 that have been selected by the authors in [Ferrario et al., 2015] by performing a Genetic Algorithm (GA)-based wrapper feature selection aimed at identifying the subset

of important inputs that maximize the ANN performance. These selected parameters form the input vector $\mathbf{Y}_n = \{y_{1,n}, y_{2,n}, \ldots, y_{M=5,n}\}$, $n \in \{1, \ldots, N = 168\}$.

Table 1: IM earthquake parameters (model inputs: $\mathbf{Y} = y_1, y_2, \ldots, y_{M=5}$).

	y_1	I_{Arias}	Arias intensity
	y_2	$PSA(T_{str})$	Spectral acceleration at the first-mode period of the structure
$\mathbf{Y}$	y_3	T_m	Mean period
	y_4	T_p	Predominant period
	y_5	SI	Spectral intensity

In the following, we refer to the model inputs by their numbers (i.e., y_1, y_2, y_3 …) instead of their name (I_{Arias}, $PSA(T_{str})$, T_m, …) for brevity.

7.2. Results

ANNs are used in order to approximate the nonlinear seismic response of the masonry building of interest. A data set composed by $N = 168$ input/output patterns is available for the analysis. In order to train, validate and test the ANN, the N data are partitioned as follows: 70% in the training set (i.e., $N_{train} = 118$ data), 15% in the validation set (i.e., $N_{val} = 25$ data) and 15% in the test set (i.e., $N_{test} = 25$ data).

In Section 7.2.1, the results of the estimation of the accuracy of the ANN by the bootstrap approach (Section 5) is reported; then, in Section 7.2.2 the computation of the structure fragility curves under uncertainty (Section 6) is illustrated.

7.2.1. ANN accuracy

To estimate the ANN accuracy, the bootstrap approach of Section 5 has been applied. The analysis has been performed with respect to four different sizes of the training and validation sets, to evaluate the network accuracy in cases where smaller data sets are available. Indeed, the seismic fragility analysis and the estimation of confidence bounds based on recorded ground motions is usually performed with a number of analyses that is smaller than or comparable to 168 (for critical infrastructures), also due to limitations in the number of suitable recorded ground motions available. In Table 2, the total number N of input/output data adopted, and the number of those used in the training (N_{train}), validation (N_{val}) and test (N_{test}) are reported, with respect to the four cases analyzed. Case 1 considers all the $N = 168$ data available, so it is composed by the largest training and validation sets ($N_{train} = 118$, $N_{val} = 25$); cases 2, 3 and 4 halve the training and validation sets of their previous case, e.g., the training and validation sets of case 2 are the rounded half of those of case 1 (i.e., $N_{train} = 59$, $N_{val} = 13$). The size of the test data set is kept constant for all the four cases ($N_{test} = 25$ data), in order to perform a comparison of the network accuracy. All the networks have been trained with a

number h of hidden nodes optimized with respect to the training and validation procedures for each network (Table 2).

Table 2: Number of patterns in the training (N_{train}), validation (N_{val}), and test (N_{test}) data sets for the four different cases evaluated; N is the total number of data and h is the optimal number of hidden neurons adopted to train and validate a single network.

	Case 1	Case 2	Case 3	Case 4
N_{train}	118	59	30	15
N_{val}	25	13	6	3
N_{test}	25	25	25	25
N	168	97	61	43
h	5	2	2	2

As a first general analysis to evaluate the accuracy of the regression estimate of $\mu_\delta(Y_n)$, we have computed the following performance indicator that represents the average standard deviation $\bar{\sigma}_{Boot_{\bar{\delta}}}$ on the mean value:

$$\bar{\sigma}_{Boot_{\bar{\delta}}} = \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \sqrt{\frac{\sigma_{Boot}^2(Y_n)}{B}}, \tag{13}$$

where $\sigma_{Boot}^2(Y_n)$ is the bootstrap variance given in equation (8) and B is the number of bootstrapped networks considered ($B = 100$ in this study). The results are reported in Figure 6 with respect to the test data set, i.e., $Y_n = 1, \ldots, N_{test}$, for the four cases of interest of Table 2, where N_{train} takes values equal to 118, 59, 30 and 15 data. For each case, several (i.e., 50) simulations (represented by the dots in the Figure) are run in order to account for the variability associated with the (randomly sampled) initial training and validation data sets, and to the random sampling with replacement implied by the bootstrap. Notice that in the reference case where $N_{train} = 118$, the training and validation data sets are the same for all the 50 simulations (since they correspond to the whole training data set); then, the variability in these results is due only to the bootstrap sampling. It can be seen that the lower the size of the training and validation data sets, the lower the network accuracy: the average of the $\bar{\sigma}_{Boot_{\bar{\delta}}}$ values (depicted by the squares in the Figure) increases from 0.0111 in case 1 ($N_{train} = 118$) to 0.0233 in case 4 ($N_{train} = 15$).
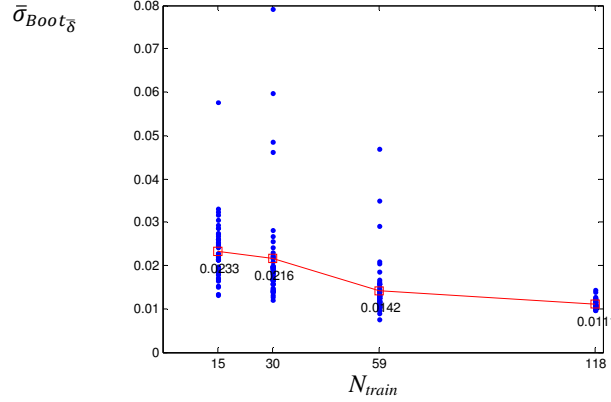
Figure 6: Dots: $\bar{\sigma}_{Boot_{\bar{\delta}}}$ computed on the test data set with respect to different sizes of the training data set (N_{train}). Squares: average of $\bar{\sigma}_{Boot_{\bar{\delta}}}$ for a given size of the training data set.

Some of the $\bar{\sigma}_{Boot_{\bar{\delta}}}$ values in Figure 6 are far from their average. From the analysis of the corresponding box plot (Figure 7), these points (represented by a plus sign) are outliers since they are outside the box and the whiskers (in this case, the edges of the box represent the 25th and 75th percentiles and those of the whiskers the 99.3 coverage, if the data are normally distributed).

The presence of outliers can be explained by the fact that some of the bootstrapped networks may not be well trained because of: (i) "unlucky" configurations of the initial training and validation data sets due to random sampling with replacement implied by the bootstrap (for example, a too large number of repeated samples may be contained in the design of experiments), or (ii) inefficient calibration of the ANN weights during training (for example, the back-propagation algorithm may get stuck in a local optimum). In this case, some biased network components of the bootstrap ensemble may not be able to estimate the model output with a satisfactory level of accuracy and may significantly degrade the estimate of the average of the bootstrapped regression functions $\bar{\delta}_{Boot}$, thus providing an overestimate of the variance σ^2_{Boot}. As suggested in [Zio, 2006], a possible way to tackle this issue is to actually train a large number of bootstrap networks and then retain only those which are regarded as well trained. In this light, for illustration purposes, we remove the outliers from Figure 6 and we compute again the average of the remaining $\bar{\sigma}_{Boot_{\bar{\delta}}}$ values (squares in Figure 8) with the sole aim of highlighting the general "trend" of the ANN accuracy with respect to the size of available training data set. It can be seen that without outliers it is more evident that the smaller the size of the training data set, the larger the value of the average of $\bar{\sigma}_{Boot_{\bar{\delta}}}$, so the smaller the network accuracy.

For completeness, the average standard deviation values $\bar{\sigma}_{Boot_{\bar{\delta}}}$ computed on the training data set (i.e., by replacing N_{test} with N_{train} in equation (13)) is reported in Figure 9. The same considerations can be drawn on the average trend: the smaller the training data set, the higher the standard deviation.
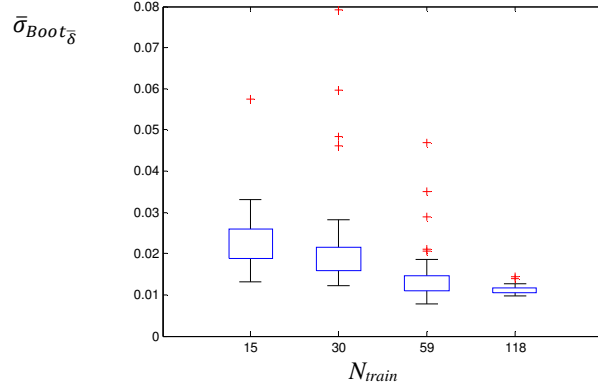
23

Figure 7: Box plot of $\bar{\sigma}_{Boot_{\bar{\delta}}}$ computed on the test data set with respect to different sizes of the training data set (N_{train}).
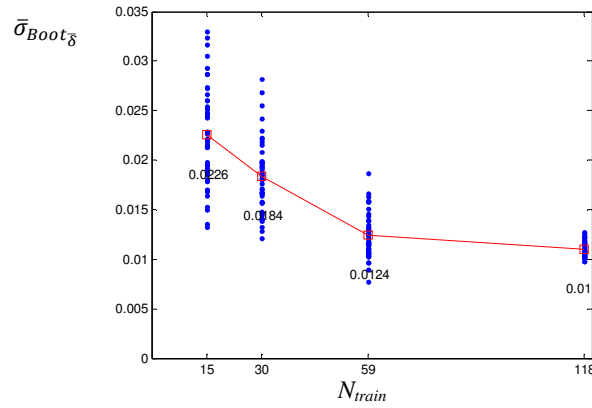

Figure 8: Dots: $\bar{\sigma}_{Boot_{\bar{\delta}}}$ computed on the test data set with respect to different sizes of the training data set (N_{train}). Squares: average of $\bar{\sigma}_{Boot_{\bar{\delta}}}$ for a given size of the training data set. The outliers highlighted in Figure 6 have been removed.
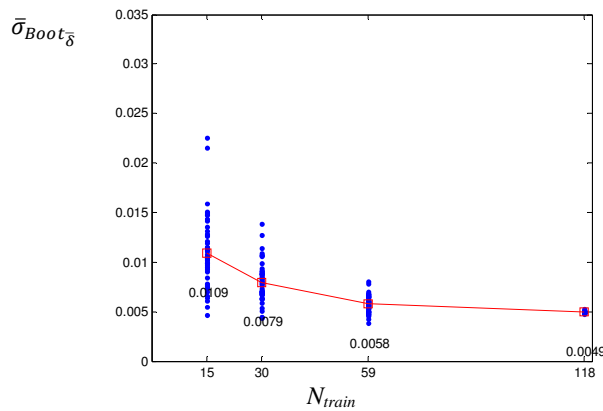

Figure 9: Dots: $\bar{\sigma}_{Boot_{\bar{\delta}}}$ computed on the training data set with respect to different sizes of the training data set, (N_{train}). Squares: average of $\bar{\sigma}_{Boot_{\bar{\delta}}}$ for a given size of the training data set.

The 50% confidence intervals $\left[\underline{\delta}^{50\%}(Y_n), \ \bar{\delta}^{50\%}(Y_n) \right]$ for the quantity $\delta_{FEM}(Y_n)$ on the test data set, $n = 1, \ldots, N_{test} = 25$, obtained by the bootstrap approach (Section 5) are illustrated by the box plots of

24

Figures 10 – 13, with respect to the four cases of Table 2, where the edges of the box represent the 25th and 75th percentiles and those of the whiskers the 99.3 coverage if the data are normally distributed. The estimates of the best ANN, δ_{ANN}, (light dots) and of the bootstrap ensemble of ANNs, $\bar{\delta}_{Boot}$ (plus signs), are reported together with the true output, δ_{FEM} (dark dots).

Figure 10: Confidence intervals of the quantity $\delta_{FEM}(Y_n)$ (dark dots) given by the bootstrap approach with respect to the test data set for the case 1 of Table 2. The estimates of the best ANN, δ_{ANN}, (light dots) and of the bootstrap ensemble of ANNs, $\bar{\delta}_{Boot}$, (plus signs) are also reported.
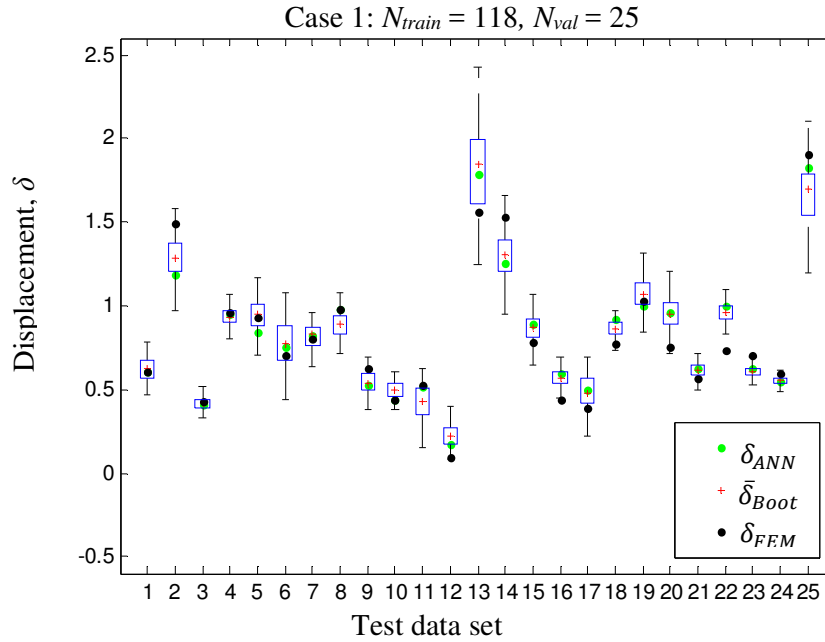
Case 2: N_{train} = 59, N_{val} = 13

Figure 11: Confidence intervals of the quantity $\delta_{FEM}(Y_n)$ (dark dots) given by the bootstrap approach with respect to the test data set for the case 2 of Table 2. The estimates of the best ANN, δ_{ANN}, (light dots) and of the bootstrap ensemble of ANNs, $\bar{\delta}_{Boot}$, (plus signs) are also reported.

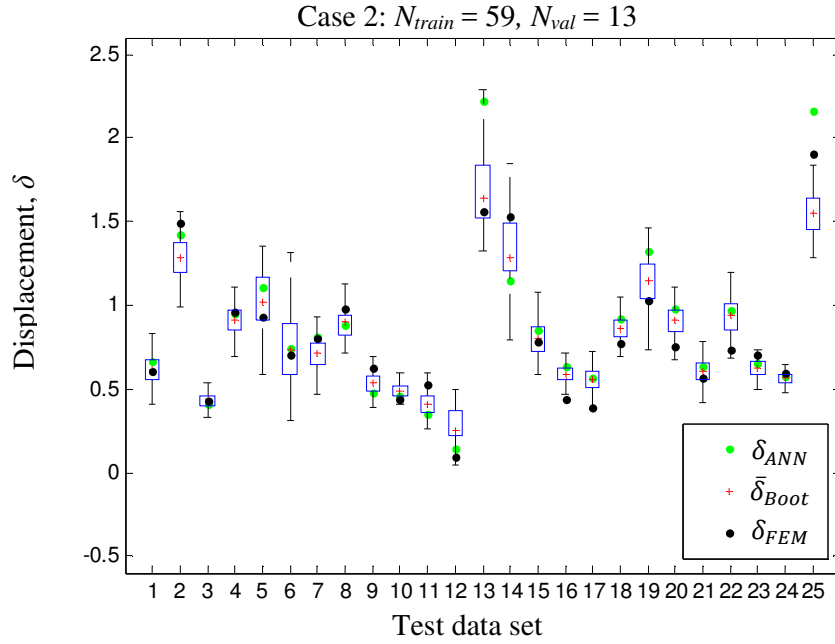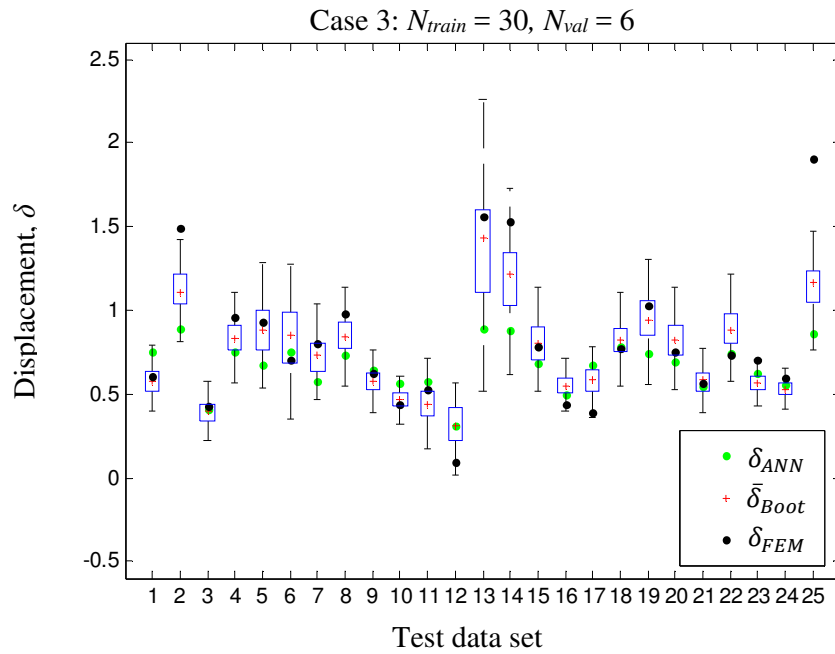Case 3: N_{train} = 30, N_{val} = 6

Figure 12: Confidence intervals of the quantity $\delta_{FEM}(Y_n)$ (dark dots) given by the bootstrap approach with respect to the test data set for the case 3 of Table 2. The estimates of the best ANN, δ_{ANN}, (light dots) and of the bootstrap ensemble of ANNs, $\bar{\delta}_{Boot}$, (plus signs) are also reported.
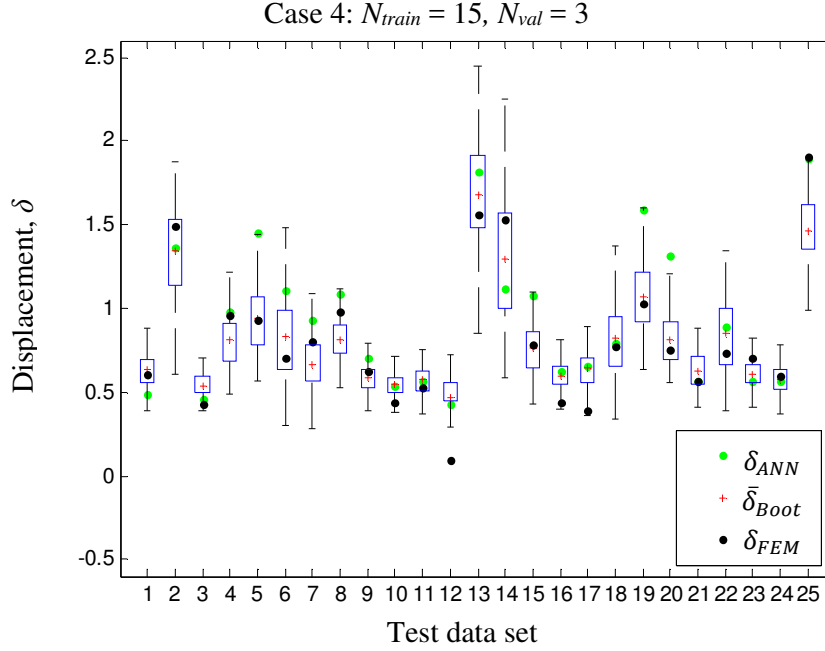
26

Figure 13: Confidence intervals of the quantity $\delta_{FEM}(\mathbf{Y}_n)$ (dark dots) given by the bootstrap approach with respect to the test data set for the case 4 of Table 2. The estimates of the best ANN, δ_{ANN}, (light dots) and of the bootstrap ensemble of ANNs, $\bar{\delta}_{Boot}$, (plus signs) are also reported.

In Table 3, the rankings of the standard deviation values $\sigma_{Boot}(\mathbf{Y}_n)$, $n = 1, \ldots, N_{test}$, computed from equation (8) for the model outputs $\delta_{Boot_b}(\mathbf{Y}_n)$, $b = 1, \ldots, 100$, $n = 1, \ldots, N_{test}$, are reported with respect to the different amount of data available for the training and validation processes. In general, the smaller the number of input/output patterns used to train (and validate) the network, the larger the probability to find standard deviation values, $\sigma_{Boot}(\mathbf{Y}_n)$, $n = 1, \ldots, N_{test}$, higher than a given threshold. For example, with respect to a threshold equal to 0.1, this probability is equal to $13/25 \sim 0.5$ in case 1 and doubles up to $25/25 = 1$ in case 4.

From Table 3 it can be seen also that the highest values of $\sigma_{Boot}(\mathbf{Y}_n)$ (at the top of the ranking) are registered in correspondence of the inputs $\mathbf{Y}_n$, $n = 6, 13, 14, 25$, for all the four cases considered. Thus, the outputs associated with these inputs are affected by high uncertainty. This can be explained by looking to Figure 14, where the N_{test} input/output test patterns (circle) are illustrated together with the N input/output data (dots) with respect to IM y_1 (I_{Arias}). The regression model may not fit well the test points 6, 13, 14 and 25 since they are at the border of the data cloud and, in addition, they are (except for test point 6) at the top of the Figure characterized by a plastic force-displacement behavior that is more difficult to predict.

Case 1		Case 2		Case 3		Case 4	
n	σ_{Boot}	n	σ_{Boot}	n	σ_{Boot}	n	σ_{Boot}
13	0.40	13	0.55	13	0.43	13	0.46
25	0.29	14	0.47	14	0.32	14	0.41
6	0.19	6	0.22	25	0.24	6	0.34
14	0.19	12	0.22	6	0.22	12	0.31
17	0.15	5	0.18	2	0.18	25	0.31
2	0.15	19	0.15	5	0.18	2	0.30
7	0.14	15	0.13	19	0.17	22	0.26
11	0.12	25	0.13	12	0.16	7	0.24
5	0.12	8	0.13	15	0.15	18	0.24
19	0.11	2	0.12	8	0.15	5	0.23
15	0.11	7	0.12	20	0.15	19	0.22
20	0.11	22	0.12	22	0.14	4	0.19
9	0.10	20	0.12	11	0.14	15	0.19
8	0.09	11	0.11	4	0.14	20	0.17
12	0.08	4	0.10	18	0.14	21	0.16
1	0.08	17	0.09	7	0.13	1	0.15
22	0.07	18	0.08	21	0.11	17	0.14
18	0.06	1	0.08	1	0.10	8	0.13
16	0.06	21	0.07	17	0.09	11	0.12
10	0.06	9	0.07	3	0.08	23	0.12
4	0.06	16	0.06	9	0.07	10	0.11
21	0.05	3	0.05	10	0.07	24	0.11
3	0.04	23	0.05	16	0.06	3	0.11
23	0.03	10	0.05	23	0.06	16	0.10
24	0.03	24	0.04	24	0.06	9	0.10

Figure 14: Input/output patterns with respect to the logarithm of the intensity measure y_1 (I_{Arias}).

To estimate the quality of the 50% confidence intervals, $\left[\underline{\delta}^{50\%}(\mathbf{Y}_n), \ \bar{\delta}^{50\%}(\mathbf{Y}_n)\right]$, for the quantity $\delta_{FEM}(\mathbf{Y}_n)$ shown in Figures 10 – 13, the CP (CP) with respect to the output given by the FEM and the

NMW (*NMW*) (eqs. 9 and 10, respectively) are illustrated in Figure 15 for the four cases under analysis. Notice that to compare the estimates of the NMW, the normalization has been performed with respect to the max and min values of the test data set, i.e., in equation 10, $\delta_{max} = \max_{n}(\delta_{FEM}(\boldsymbol{Y}_n))$ and $\delta_{min} = \min_{n}(\delta_{FEM}(\boldsymbol{Y}_n))$, $n = 1, \ldots, N_{test}$, that is the same for all the cases analyzed.

As expected, the smaller the number of data used to train and validate the network, the higher both the *CP* and the *NMW* of the intervals. The complement to 1 of *CP* has been plotted on the horizontal axis to highlight the conflict between these two measures (that are in fact on a Pareto frontier). Actually, an estimate is considered satisfactory if the corresponding confidence interval is characterized by high coverage and small width.

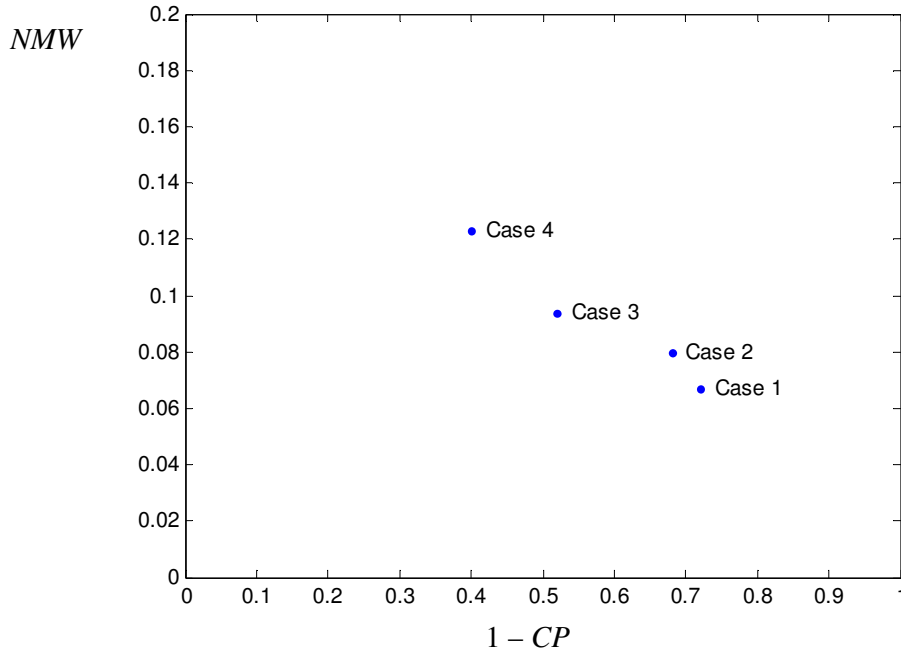|	**Case 1**	**Case 2**	**Case 3**	**Case 4**
N_{train}	118	59	30	15
N_{val}	25	13	6	3
CP	0.280	0.320	0.480	0.600
NMW	0.067	0.080	0.094	0.123

Figure 15: Coverage probability (CP) with respect to the true output given by the FEM and Normalized Mean Width (NMW).

A comparison between the estimates given by the bootstrap ensemble of ANNs and the best ANN computed is given by the computation of the RMSEs, i.e., by *RMSE_Boot* (eq. 11) and *RMSE_ANN* (eq. 12), respectively, as reported in Table 4 for the four cases of interest. As expected, the smaller the training data set, the higher the RMSEs. In addition, *RMSE_Boot* is always lower than *RMSE_ANN*. This confirms the findings reported in the literature with respect to the superior estimation accuracy of the

bootstrap aggregation procedure. In particular, the smaller the training data set, the higher the accuracy of the bootstrap ensemble compared to that of the best ANN: the percentage difference between the $RMSE_{Boot}$ and $RMSE_{ANN}$ is around 5% for case 1 and 38% for case 4.

Table 4: Percentage of the Root Mean Square Errors (RMSEs) of the best ANN ($RMSE_{ANN}$) and of the bootstrap ensemble of ANNs ($RMSE_{Boot}$) with respect to the true output given by the FEM.

	Case 1	Case 2	Case 3	Case 4
N_{train}	118	59	30	15
N_{val}	25	13	6	3
$RMSE_{ANN}$ (%)	13.6	20.7	33.6	26.4
$RMSE_{Boot}$ (%)	12.9	13.7	20.5	16.4

7.2.2. Fragility curves

As recalled in Section 2, the fragility curve is the probability of exceeding a damage threshold of interest, δ^*, for a given ground motion level. For illustration purposes, we assume a damage threshold δ^* equal to 0.63 cm (that corresponds to a slight damage of the structure) and we build the fragility curve on the test data set, where $N_{test} = 25$. The following evaluations are carried out with respect to an ANN trained with $N_{train} = 118$ data (case 1 of Table 2).

The α and β parameters estimated by the maximum likelihood method with respect to the ANN output (i.e., the displacement δ_{ANN}) assume values equal to 1.02 and 1.17, respectively. The corresponding estimates of the 95% confidence intervals are $\left[\underline{\alpha}^{95\%}, \ \bar{\alpha}^{95\%}\right] = [0.32, 1.73]$ and $\left[\underline{\beta}^{95\%}, \ \bar{\beta}^{95\%}\right] = [0.88, 1.96]$, respectively. In Figure 16, the fragility curve F (solid line) is illustrated together with its lower, $\underline{F}$ (dotted line), and upper, $\bar{F}$ (dashed line), bounds: these bounds represent the epistemic uncertainty on the estimates of the α and β parameters due to the paucity of data ($N_{test} = 25$); actually, such bounds are intended to envelop all the fragility curves that can be generated by all the possible combinations of α and β values ranging in their corresponding CIs. It can be noticed a discontinuity when the fragility is equal to 0.5. This is due to the properties of the lognormal distributions: the CDFs characterized by the same value of α and different values of β intersect in 0.5. In Figure 17, the intersections of the CDFs obtained by combining the highest and lowest values of the confidence intervals of α and β are illustrated. It can be seen that the upper bound is given by α equal to the minimum value ($\alpha = \underline{\alpha}^{95\%}$) and β equal to the maximum value ($\beta = \bar{\beta}^{95\%}$) if the fragility is lower than 0.5, and to the minimum value ($\beta = \underline{\beta}^{95\%}$) otherwise; vice versa, the lower bound is identified by α equal to the maximum value ($\alpha = \bar{\alpha}^{95\%}$) and β equal to the minimum value ($\beta = \underline{\beta}^{95\%}$) if the fragility is lower than 0.5, and to the maximum value ($\beta = \bar{\beta}^{95\%}$) otherwise.
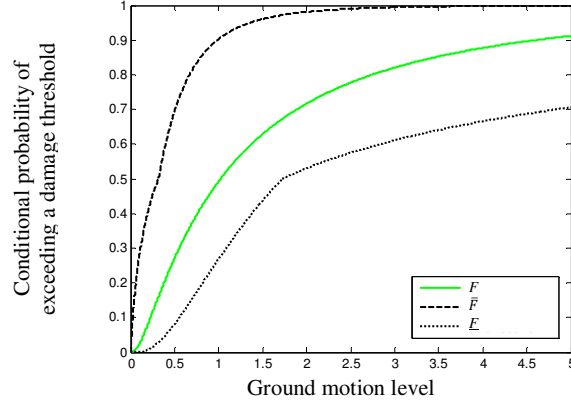
Figure 16: Estimated fragility curve F (solid line) with the corresponding upper, $\bar{F}$ (dashed line), and lower, $\underline{F}$ (dotted line), bounds representing the epistemic uncertainty due to the data.
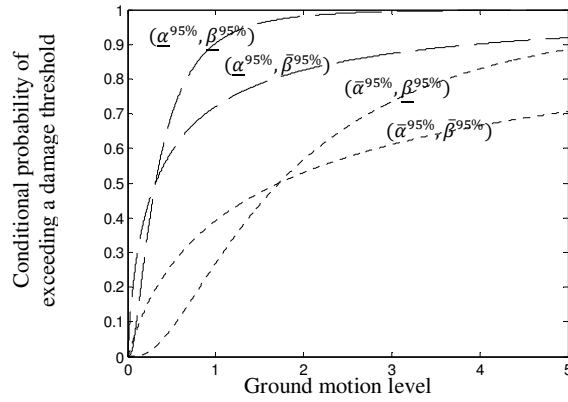

Figure 17: Intersections of the fragility curves in correspondence of the combinations of the highest and lowest values of the parameters α and β.

In order to include also the model uncertainty due to ANN regression, the bootstrap approach of Section 5 is carried out. For each bootstrapped regression models b, the outputs $\delta_{Boot_b}(Y_n)$, $b \in \{1, \dots, B\}$, $n = 1, \dots, N_{test}$, are computed. Then, a fragility curve F and its lower $\underline{F}$ and upper $\bar{F}$ bounds are estimated, as illustrated in Sections 2 and 6. Repeating this procedure for all the B bootstrapped regression models, a family of fragility curves F_b, $b = 1, \dots, B$, is obtained (Figure 18, left) and the corresponding families of the lower $\underline{F}_b$ and upper $\bar{F}_b$ bounds, $b = 1, \dots, B$, are determined (Figure 18, right). These results are combined in Figure 19 where the solid lines envelope all the possible fragility curves F_b, $b = 1, \dots, B$, and represent the epistemic uncertainty due to the ANN model; the dotted and dashed lines represent the extreme lower ($\underline{\underline{F}}$) and upper ($\bar{\bar{F}}$) bounds of all the fragility curves, and they account both for the uncertainties due to the ANN model and to the data.
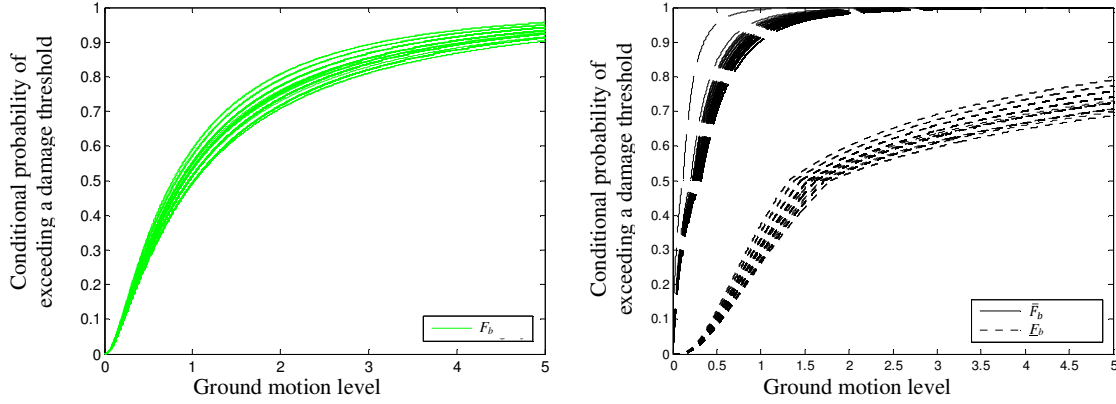
Figure 18: Left: fragility curves F_b, b = 1, ..., B, estimated from the B bootstrapped networks; right: corresponding upper, $\bar{F}_b$, (dashed lines) and lower , $\underline{F}_b$,(dotted lined) bounds of the fragility curves, b = 1, ..., B.
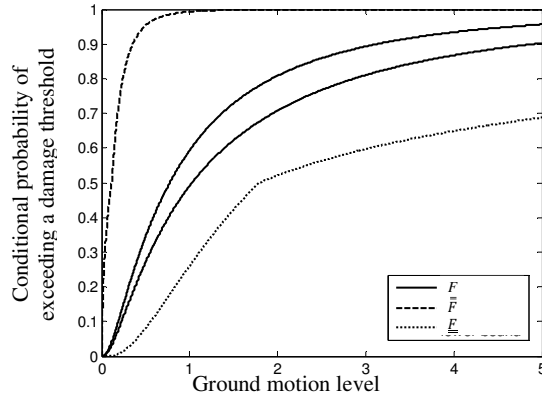

Figure 19: Uncertainty due to the network and the data of the fragility curves. Solid lines: upper and lower bounds of the fragility curves F_b, b = 1, ..., B, estimated from the B bootstrapped networks (network uncertainty). Dashed and dotted lines: corresponding upper and lower bounds, $\bar{\bar{F}}$, and $\underline{F}$,respectively (network and data uncertainty).

8. Conclusions

In the context of seismic risk analysis, FEMs are typically employed to simulate the structural response of a system; however, the computational burden associated with this analysis is impracticable, at times. This is due to the presence of large (aleatory and epistemic) uncertainties affecting the system. As a consequence, to explore their wide range of variability, a very large number (e.g., several thousands) of FEM simulations is required for an accurate assessment of the system behavior under different seismic conditions. In addition, long calculations (hours of even days) are necessary for each run of the FEM.

For these reasons, in [Ferrario et al., 2015] the authors have explored the possibility of replacing the FEM by a fast-running regression model, i.e., the ANN, to reduce the computational costs. The results showed a satisfactory capability of ANNs to approximating the FEM output in less than a second.

In this paper, the authors have extended the previous analysis to account for the accuracy of ANNs. Actually, this issue is essential for the confident and reliable use of ANNs in safety critical systems as NPPs. The bootstrap method has been adopted to estimate confidence intervals on the quantity of interest, since in the scientific literature it has been proven that it provides accurate uncertainty estimates; a drawback of the approach can be found in the computational time that may be long due to the elaborate training algorithm for building the structurally complex neural model.

As case study, we have looked at the structural behavior of a masonry structure under the occurrence of a seismic event. In particular, we have identified the structure fragility curves representing the conditional probability of damage of the structure for any given ground motion level.

ANNs have been built on the basis of sets of data of limited, varying sizes, which represent examples of the nonlinear relationships between five uncertain inputs (representing the seismic characteristics of a single earthquake) and one output (the maximal structural top displacement).

The following analyses have been performed to analyze the uncertainty on the ANN regression function:

- Computation of an indicator of network precision, based on the bootstrap standard deviation. As expected, this analysis has shown a reduction of the network accuracy when the size of the training and validation data sets decreases.

- Determination of the 50% confidence intervals by the bootstrap approach for the output of interest. It can be seen that the number of patterns in the data sets for training and validating the networks is relevant to its performance. In addition, this analysis has led to:
  - the identification of those patterns that are not well-fitted by the ANN (i.e., those characterized by a large confidence interval). More training data close to these patterns should be employed to reduce the error of the regression function.
  - the identification of the Pareto front of the CP and the NMW that allows understanding a satisfactory tradeoff between these conflicting measures with respect to the size of the training and validation data sets.

- Comparison between the accuracy of the bootstrapped ANNs and the best ANN in the bootstrap ensemble of ANNs by the computation of the RMSE. The accuracy of the ensemble of ANNs is always better than that of the best ANN [Zio, 2006], then the bootstrap aggregation procedure provides more accurate estimates of the model output and a better "reproduction" of the behavior of the underlying physical process itself. The smaller the training data sets, the higher the difference between the accuracy of the ensemble and the best ANN.

Finally, the fragility curve of the masonry building and its uncertainties have been estimated for a given damage threshold. First, only the epistemic uncertainty due to the paucity of data in inferring the fragility parameters has been considered. Then, the epistemic uncertainty due to the ANN metamodel has been added to the analysis. As a result, a family of fragility curves accounting for the uncertainty in the ANN and in the data has been obtained.

Bootstrapped ANNs, properly trained and validated to reproduce the behavior of the original, long-running system model code, represent a tool that can be used *in general* for carrying out thorough uncertainty and sensitivity analyses of computationally expensive FEMs for structural risk assessment. The adopted approach represents a fast and effective tool for the quantification of the epistemic uncertainty associated with the estimates of the system response, since it produces a family of (epistemic) probability distributions for the output, without making any underlying assumptions about the model structure (actually, it is a non-parametric method). Moreover, the quantification of epistemic uncertainty by means of probability distributions allows its straightforward propagation through computer codes (which is not always the case when using confidence intervals).
In conclusion, the adopted approach allows a robust quantification of the uncertainty due to the (few) available data, which provides reliable and conservative results that are essential for safety-critical applications and risk-informed decisions (like those of interest to the present paper).

Future work will extend the application of the approach to the KAshiwazaki-Kariwa Research Initiative for Seismic Margin Assessment (KARISMA) benchmark (IAEA 2014) that refers to the real case of a NPP (unit 7 of the site Kashiwazaki Kariwa) affected, in 2007, by a strong earthquake (beyond the design criteria) and for which a rich data set (measurements and analytical methods) is available.

References

Ak, R., Li, Y. F., Vitelli, V., Zio, E., Droguett, E. L., and Jacinto, C. M. C. (2013). "NSGA-II-trained neural network approach to the estimation of prediction intervals of scale deposition rate in oil & gas equipment." *Expert Systems with Applications*, 40(4), 1205-1212.

Apostolakis, G. (1990). "The Concept of Probability in Safety Assessments of Technological Systems." *Science*, 250(4986), 1359-1364.

Aubry, D., Chouvet, D., Modaressi, A., and Modaressi, H. (1986). "GEFDYN: Logiciel d'Analyse de Comportement Mécanique des Sols par Eléments Finis avec Prise en Compte du Couplage Sol-Eau-Air. Manuel Scientifique.", Ecole Centrale Paris, LMSS-Mat.

Aven, T. (2010). "On the Need for Restricting the Probabilistic Analysis in Risk Assessments to Variability." *Risk Analysis*, 30(3), 354-360.

Aven, T., and Zio, E. (2011). "Some considerations on the treatment of uncertainties in risk assessment for practical decision making."*Reliability Engineering & System Safety*, 96(1), 64-74.

Babuska, I., Nobile, F., and Tempone, R. (2010). "A Stochastic Collocation Method for Elliptic Partial Differential Equations with Random Input Data." *Siam Review*, 52(2), 317-355.

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). "Sequential design of computer experiments for the estimation of a probability of failure." *Statistics and Computing*, 22(3), 773-793.

Beer, M., Kougioumtzoglou, A., Patelli, E. (2014). "Emerging concepts and approaches for efficient and realistic uncertainty quantification." *Maintenance and Safety of Aging Infrastructure*: *Structures and Infrastructures Book Series*, Vol. 10, 121-162.

Beer, M. Spanos, P. D. (2009). "A neural network approach for simulating stationary stochastic processes." *Structural Engineering and Mechanics*, 32(1), 71-94.

Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M. (2008). "Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions." *AIAA Journal*, 46(10), 2459-2468.

Bishop, C. M. (1995). *Neural Networks for pattern recognition*, Oxford University Press.

Bucher, C., and Most, T. (2008). "A comparison of approximate response functions in structural reliability analysis." *Probabilistic Engineering Mechanics*, 23(2-3), 154-163.

Cardoso, J. B., de Almeida, J. R., Dias, J. M., and Coelho, P. G. (2008). "Structural reliability analysis using Monte Carlo simulation and neural networks." *Advances in Engineering Software*, 39(6), 505-513.

Chojaczyk, A. A., Teixeira, A. P., Neves, L. C., Cardoso, J. B., Guedes Soares, C. (2015) "Review and application of Artificial Neural Networks models in reliability analysis of steel structures." Structural Safety, 52, Part A, 78-89.

Ciriello, V., Di Federico, V., Riva, M., Cadini, F., De Sanctis, J., Zio, E., and Guadagnini, A. (2013). "Polynomial chaos expansion for global sensitivity analysis applied to a model of radionuclide migration in a randomly heterogeneous aquifer." *Stochastic Environmental Research and Risk Assessment*, 27(4), 945-954.

Cybenko, G. (1989). "Approximation by superpositions of a sigmoidal function." *Mathematics of Control Signals Systems*, 2, 303-314.

Dubourg, V., Sudret, B. (2014) "Meta-model-based importance sampling for reliability sensitivity analysis." Structural Safety, 49, 27-36.

Dybowski, R., and Roberts, S. J. (2001). "Confidence intervals and prediction intervals for feedforward neural networks." Clinical Applications of Artificial Neural Networks, R. Dybowski and V. Gant, eds., Cambridge University Press.

Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife." *Annals of Statistics*, 7, 1-26.

Efron, B., and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, Chapman & Hall/CRC, New York.

Ellingwood, B.R., Kinali, K. (2009) "Quantifying and communicating uncertainty in seismic risk assessment." Structural Safety, 31(2), 179-187

EPRI. (2003). "Seismic Probabilistic Risk Assessment Implementation Guide." *TR-1002989*, Palo Alto, CA.

Ferrario, E., Pedroni, N., Zio, E., and Lopez-Caballero, F. (2015). "Application of metamodel-based techniques for the efficient seismic analysis of structural systems." European Safety and Reliability Conference ESREL, Zürich.

Ferson, S., and Ginzburg, L. R. (1996). "Different methods are needed to propagate ignorance and variability." *Reliability Engineering & System Safety*, 54(2-3), 133-144.

Hurtado, J. E. (2007). "Filtered importance sampling with support vector margin: A powerful method for structural reliability analysis." *Structural Safety*, 29(1), 2-15.

Liel, A. B., Haselton, C. B., Deierlein, G. G., and Baker, J. W. (2009). "Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings." *Structural Safety*, 31(2), 197-211.

Lopez-Caballero, F., Modaressi-Farahmand-Razavi, A., and Tabbakhha, M. (2011). "Effects of soil foundation on the performance of masonry buildings." Proceedings of the 8th International Conference on Structural Dynamics, EURODYN 2011, Leuven, Belgium

NASA. (2010). "Risk-Informed Decision Making Handbook." NASA/SP-2010-576 - Version 1.0.

Pedroni, N., Zio, E., and Apostolakis, G. E. (2010). "Comparison of bootstrapped artificial neural networks and quadratic response surfaces for the estimation of the functional failure probability of a thermal-hydraulic passive system." *Reliability Engineering & System Safety*, 95(4), 386-395.

Porter, K. (2015). "A Beginner's Guide to Fragility, Vulnerability, and Risk." Denver CO USA.

Rivals, I., and Personnaz, L. (1998). "Construction Of Confidence Intervals In Neural Modeling Using A Linear Taylor Expansion." Proceedings of the International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). "Learning internal representations by error back-propagation." Parallel distributed processing: exploration in the microstructure of cognition (vol. 1), D. E. Rumelhart and J. L. McClelland, eds., Cambridge (MA): MIT Press.

Saez, E., Lopez-Caballero, F., and Modaressi-Farahmand-Razavi, A. (2011). "Effect of the inelastic dynamic soil-structure interaction on the seismic vulnerability assessment." *Structural Safety*, 33(1), 51-63.

Shinozuka, M., Feng, M. Q., Lee, J., and Naganuma, T. (2000). "Statistical analysis of fragility curves." *Journal of Engineering Mechanics-Asce*, 126(12), 1224-1231.

Sudret, B., and Mai, C. V. (2015). "Computing derivative-based global sensitivity measures using polynomial chaos expansions." *Reliability Engineering & System Safety*, 134, 241-250.

Ulrich, T., Negulescu, C., Douglas, J. (2014). "Fragility curves for risk-targeted seismic design maps." Bulletin of Earthquake Engineering, Springer Verlag, 12 (4), pp.1479-1491.

USNRC. (2009). "Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making." NUREG-1855. US Nuclear Regulatory Commission, Washington, DC.

Villemonteix, J., Vazquez, E., and Walter, E. (2009). "An informational approach to the global optimization of expensive-to-evaluate functions." *Journal of Global Optimization*, 44(4), 509-534.

Zhang, L. G., Lu, Z. Z., and Wang, P. (2015). "Efficient structural reliability analysis method based on advanced Kriging model." *Applied Mathematical Modelling*, 39(2), 781-793.

Zio, E. (2006). "A study of the bootstrap method for estimating the accuracy of artificial neural networks in predicting nuclear transient processes." *Ieee Transactions on Nuclear Science*, 53(3), 1460-1478.

Zio, E., and Pedroni, N. (2011). "How to effectively compute the reliability of a thermal-hydraulic nuclear passive system." *Nuclear Engineering and Design*, 241(1), 310-327.