

# Gaussian process regression for seismic fragility assessment of building portfolios

Roberto Gentile<sup>1,2</sup>, Carmine Galasso<sup>2,3</sup>

<sup>1</sup>Institute for Risk and Disaster Reduction, University College London, London, United Kingdom

5 <sup>2</sup>Department of Civil, Environmental and Geomatic Engineering, University College London, London, United Kingdom;

<sup>3</sup>Scuola Universitaria Superiore (IUSS) Pavia, Pavia, Italy

*Correspondence to:* Roberto Gentile (r.gentile@ucl.ac.uk)

**Abstract.** Seismic fragility assessment of building portfolios is often based on the analysis of “average” building models representative of structural types (or building classes), thus neglecting building-to-building variability within a structural type. This paper proposes the use of gaussian process (GP) regressions to develop flexible and accurate metamodels explicitly mapping building-class attributes to the seismic fragility parameters. The proposed metamodels can enable analysts to account for building-to-building variability in simulation-based seismic risk assessment of building portfolios. Unlike other commonly-used metamodels, GP regressions do not require the a-priori definition of a prediction function and they quantify the uncertainty on the predictions in a refined and explicit fashion. The proposed method is demonstrated for a portfolio of seismically-deficient reinforced concrete school buildings with construction details typical of some developing countries. Based on the available information about the building attributes (geometry, materials, detailing), building realisations are generated based on two alternative approaches, which are critically compared: design of experiment and Monte Carlo sampling. Cloud-based time-history analysis for each building realisation is performed using unscaled real ground-motion records; fragility relationships are derived for four structure-specific damage states. A GP regression is then developed for each considered fragility parameter (i.e. median and dispersion). To further increase the tractability of the methodology, alternative metamodels are defined based on numerical non-linear static pushover analyses or analytical “by-hand” pushover analyses, through the Simple Lateral Mechanism Analysis (SLaMA) method. The results show that, for the considered portfolio, the fitted GP regressions have a high predictive power in surrogating the modelled fragility, demonstrating the feasibility of the approach in practice. It is also shown that the choice of the sampling technique could be based on the input data availability, rather than on the expected computational burden. Finally, the use of simplified methods for response analysis shows acceptable error levels with respect to the full time-history analysis results. Such simplified methods can be promising alternatives to generate large training datasets for the proposed GP regressions. This increases the potential of training metamodels in practical portfolio risk assessment applications, in which a high number of building types, each characterised by a large number of attributes, is generally involved.

10  
15  
20  
25  
30

## 1. Introduction and motivations

Seismic fragility is quantitatively expressed as the conditional probability that a structure will reach or exceed a specified level of damage (or damage state, DS) for a given value of a considered earthquake-induced ground-motion intensity measure (IM). Fragility relationships describe such a conditional probability for increasing values of the ground-motion IM, taking the form of cumulative distribution functions (CDFs). Typically, a lognormal model characterised by two parameters - median and dispersion ("logarithmic standard deviation") – is used for this purpose. These relationships are a key ingredient in any seismic risk assessment exercise, particularly in the case of large building portfolios. Fragility curves can be obtained 1) empirically, i.e. using observed (post-event) damage data; 2) through expert opinion; or 3) simulating the structural response through mechanics-based numerical models and analysis types of different complexity and computational demand.

Quantifying the potential impact of earthquakes on portfolios of properties located in seismically-prone regions is of primary interest to property owners, (re)insurance companies, capital lending institutions, and local government agencies, among others. Each is likely to have a different viewpoint and different requirements. For instance, building owners and corporate risk managers who may deal with a single building or a handful of buildings located at different site. In this case, performance-based assessment approaches (e.g. the performance-based earthquake engineering framework [1], and its practical implementation in the Federal Emergency Management Agency P-58 guidelines [2]) as opposed to building-level fragility models would be more appropriate and should be the preferred option. At the other end of the spectrum, it is not uncommon for primary insurance companies to perform fragility/loss analyses for large portfolios containing as many as a million properties. In this case, it is generally unfeasible to consider the building-to-building variability by running refined analyses (using refined computational models) for each individual structure within the portfolio.

Therefore, various structural types (or building classes) are often defined through a combination of a few attributes (e.g. lateral-loads resisting systems and their materials, height/number of storeys, age of construction/major retrofitting), assigning specified fragility and vulnerability (i.e. likelihood of loss levels vs IM levels) relationships to each structural type. Such relationships, usually derived for an “average” archetype building of the class (or an “index” building), are adopted for all the buildings within the class.

Approaches accounting for epistemic uncertainties due to structure-specific modelling uncertainties, using various statistical models, are available [3,4], particularly for collapse fragility assessment; very few studies deal with the uncertainties in seismic fragility/vulnerability modelling of building classes for portfolio risk assessment [5,6]. Building-class fragilities, including various epistemic uncertainties can also be derived, for instance by using the robust fragility approach [7]. Once the mean and confidence interval the class fragility curves are estimated, single fragility curves can be simulated (e.g. [8]). With this method, however, the one-to-one correspondence between the building characteristics and the simulated fragility curves is lost. To explicitly include building-to-building variability in practical seismic risk assessment applications (which is the focus of this study), a mapping between the building input attributes (e.g. geometry, material properties) and the seismic fragility parameters may be desirable. Metamodels can be effectively adopted to this aim, by using a small sample of building realisations within

the class for the metamodel training/validation and keeping a mechanics-based approach. A metamodel is “a model of a model”: it defines a relationship between a given set of inputs (e.g. building geometry and materials in this study) and outputs (e.g. fragility parameters in this study). This is obtained analysing a relatively-small set of samples (e.g. structures belonging to the same building class) and fitting a prediction function to the outputs, thus replacing (or surrogating) the real, computationally-expensive numerical model.

This allows one to have a computationally-efficient tool capturing the complex and implicit relationships between model inputs and outputs, such as seismic fragility parameters as a function of structural geometry, materials, detailing, etc. In general, metamodels can consider both structure-specific attributes (materials, modelling, etc.) and their variability as well as class-specific ones (geometry, detailing). However, commonly-adopted metamodels suffer some drawbacks, such as: 1) the user should determine a functional form (i.e. a prediction function) for the fitting, potentially generating substantial misfit or, conversely, physically-unsound overfitting; 2) they often rely on simplifying assumption (e.g. homoscedasticity, i.e. the variance of the error term is independent of the value of the input variables). One may argue that controlling the functional form for the fitting might help in reflecting the mechanics underlying the problem under investigation. Although such mechanics can be easily captured for structure-specific variables, such process might not be straightforward when dealing with class-specific attributes, or when several interacting input parameters are involved.

To address the above issues, the use of Gaussian Process (GP) regression [9] is proposed in this study to develop a flexible and accurate metamodel for the seismic fragility parameters (i.e. median and dispersion) of building realisations within a given building class. The proposed approach is demonstrated for seismically-deficient reinforced concrete (RC) school buildings with construction details typical of developing countries (such as Philippines and Indonesia), for which data is available with regard to the statistical distributions of the building-class input attributes. Two different techniques are independently adopted to sample realisations within the analysed building classes: a Design of Experiment (DoE) and a plain Monte Carlo sampling (MCS). This is done to simulate alternative scenarios in which the probability distributions of the building-specific input attributed are available (MCS) or unavailable (DoE), totally or partially.

For each building realisation, a fragility curve is defined based on cloud-based non-linear time-history analysis employing 150 unscaled real (i.e. recorded) ground motions. A GP regression is fitted for each resulting fragility parameter. Through an illustrative application, the advantages and shortcomings of both sampling techniques are critically discussed, and guidance is provided for the selection of the sampling method and the number of considered samples.

It is worth noting that significant computational effort may still be required to train such metamodels for a single structural type, i.e. thousands of time-history analyses are needed. This means that for practical applications, in which a high number of building types is involved, metamodels trained based on non-linear time-history analysis results may still be too computationally expensive. Therefore, to further increase the tractability of the proposed approach, two simplified-yet-accurate analysis methods are proposed to increase the potential of metamodels in efficiently considering a high number of building

classes and building attributes within a class. The detailed validation of simplified response analysis methods is outside the scope of this study, although this is currently under investigation by the authors, showing promising results [10].

Non-linear static force-displacement curves are alternatively derived through the analytical approach named Simple Lateral Mechanism Analysis (SLaMA; [11–15]), or through a standard numerical pushover (with invariant force profile). The Capacity Spectrum Method (CSM; [16]), adopting the same set of recorded ground motions, is applied using such curves to derive the seismic response in terms of inter-storey drift, selected here as engineering demand parameter (EDP). Therefore, two alternative sets of fragility functions are derived, together with the corresponding GP regressions. The results from this study indicate that such an approach might be feasible in practice.

The paper is organised as follows. Section 2 reviews the commonly adopted metamodeling techniques, highlighting their strengths and limitations. Section 3 describes the methodology adopted for this study. Sections 4 and 5 describe the considered case study and the results obtained through the proposed approach. Section 6 provides the conclusions, highlighting future research questions.

## 2. Commonly adopted metamodeling techniques

This section briefly reviews some of the commonly-adopted metamodeling techniques, briefly compared in Table 1. A thorough examination of all the available approaches in the literature is outside of scope of this paper. The desired features of the metamodel (in relation to fragility modelling) are described, highlighting the ones leading to the choice of the GP regressions.

**Table 1 Characteristics of the considered metamodeling techniques.**

	Functional form	Error term	Notes
Response surface method	Assumed	Generally homoscedastic	2 <sup>nd</sup> order polynomials are generally used
High dimensional model representation	Assumed	Generally homoscedastic	Higher-level polynomials are used
Multivariate adaptive regression spline	Not assumed	Heteroscedastic	A finite set of candidate basis functions is assigned
Support vector machines	Not assumed	Heteroscedastic	Inherently non-transparent due to the hidden layer
Gaussian process	Not assumed	Heteroscedastic	

The Response Surface Method (RSM; [17]) models the relationship between several explanatory (input) variables and one or more output variables. It consists of first calculating one or more response variables (typically for a DoE) that depend on a suite of input variables. Hence, an assumed functional form (usually a polynomial one) is fitted to the dataset, for instance through a least squares approach. Such method often assumes homoscedasticity and normality of the error terms. RSM has been effectively used in earthquake engineering applications, e.g. [4,5,18–20].

High-dimensional model representation [21] is based on a hierarchical expansion of functions that represent a response surface. A first order polynomial function represents the effect of each explanatory variable acting independently, a second order one represents the coupled effects of any pair of variables, etc. The user can choose the required level of refinement by truncating the expansion. Such approach is more convenient than the RSM when the number of inputs is particularly high. Although such

method provides better accuracy than the RSM [22], an assumed functional form is needed also in this case for the adopted prediction function. Such methods have been adopted in earthquake engineering (e.g. [23]).

Multivariate adaptive regression splines are non-parametric regression techniques based on a weighted sum of a number of basis functions. Such functions are fitted to the observed data in different portions of the input domain [24]. It can be seen as an extension of linear models that “automatically” captures non-linearities and interactions between variables. Such method provides higher flexibility than RSM, given that the hypotheses of homoscedasticity and Normal distribution of errors can be removed. However, a finite set of candidate basis functions should be assigned a priori.

Radial basis function networks [25] are artificial neural networks composed of three layers of neurons: input layer, hidden layer, output layer. The method is based on the non-linear mapping of the input layer to the “hidden layer” using non-linear functions, and a subsequent linear mapping to the output layer. In particular, the non-linear mapping is based on radial basis functions which usually have the form of Gaussian or spline functions. An example of using artificial neural networks for seismic fragility estimations can be found in [26]. Although such method generally provides particularly good fits, it is inherently non-transparent, similar to any artificial neural network approach.

Support vector machines [27] are a class of statistical methods based on the mapping of the input data to a higher-dimensional space (so called feature space) and performing linear regression in this space. The initial mapping is done using non-linear kernel functions that can also allow one to explicitly consider the correlation among different input vectors. The linear regression in the feature space is based on the minimization of a loss function related to the observed vs predicted output. Such method has been used for seismic fragility estimation, for example in [28]. One of the main disadvantages of this method is that the parameters of the kernel functions should be specified, rather than learned from the data.

GP regressions [9] are one of the possible methods to overcome the above-mentioned downsides of commonly adopted metamodeling techniques and, for this reason, they are selected in this paper. A GP regression is a non-parametric statistical method identifying a multivariate Gaussian distribution over the possible functions fitting a dataset of observed data. The mathematical formulation of a GP regression is briefly described in Section 3.5, with particular reference to the specific application in this study. As an example, a GP approach was adopted in [29] to fit probabilistic seismic demand models (EDP vs IM relationships) in the context of seismic fragility analysis.

### 3. Methodology

This study aims 1) to show the feasibility of using GP regressions to explicitly consider building-to-building variability in seismic fragility assessment of building portfolios; 2) to provide practical guidance on how to develop (i.e. train and validate) GP regressions for case-study applications. The selected parameters to be surrogated are the median and the dispersion (i.e. logarithmic standard deviation) of the fragility functions defined for four different damage states (Figure 1). The methodology involves: 1) defining an archetype building representative of each analysed class; 2) based on the available statistics (for

instance, as in this study, based on a large rapid-visual-surveying exercise for a given building portfolio) of the geometric and mechanical properties of the buildings within each class, generating representative building samples; 3) deriving fragility curves for each building realisation within the class; 4) fitting one GP regression model for each considered parameter of the fragilities; 5) testing the goodness-of-fit of the developed GP regression, also checking if the independently-fit models produce consistent sets of fragility functions for every possible combination of the input parameters.

The two adopted sampling strategies are described in Section 3.1, while the description of the considered case-study portfolio is provided in Section 4. Sections 3.2 and 3.3 describe the three increasingly-refined analysis methods that are independently adopted for the seismic response analysis, while Section 3.4 and 3.5 discuss the strategies to fit fragility curves and GP regressions, respectively.

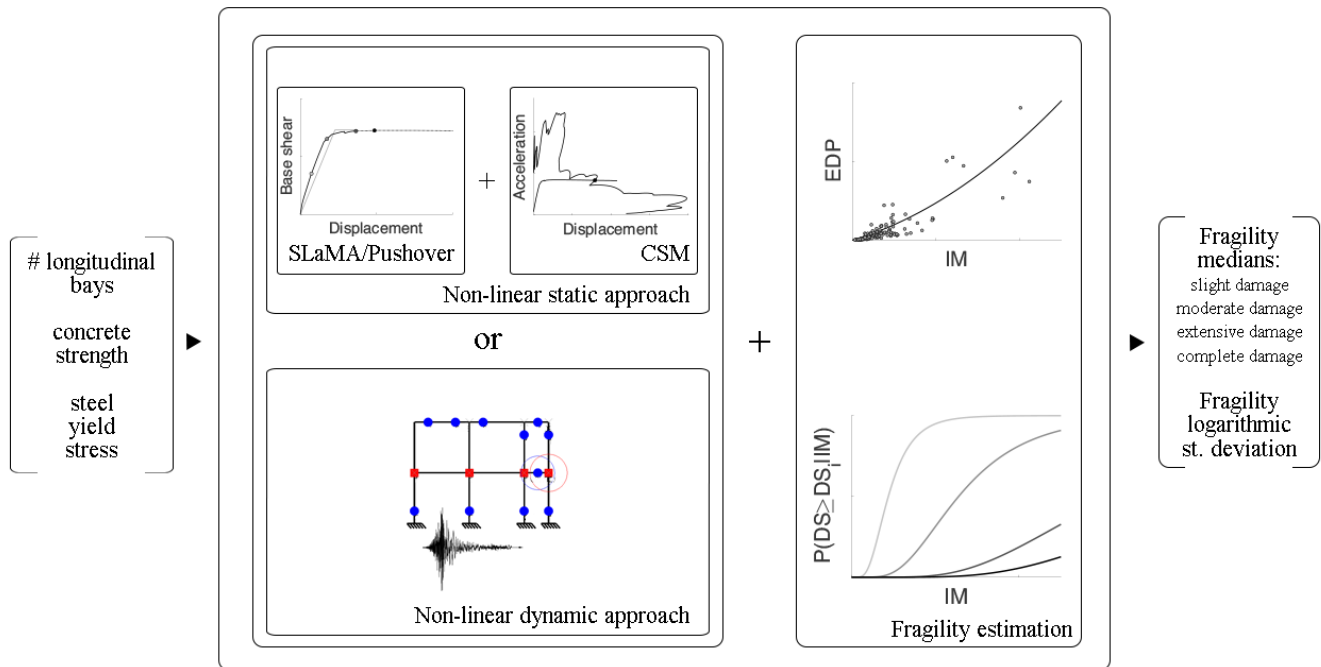


Figure 1 Description of the metamodeling approach.

### 3.1. Adopted sampling techniques

A structural type (or building class) is generally represented by an “average” archetype (or index) building. A given building within the class corresponds to a realisation of a vector of  $n$  selected random variables,  $\mathbf{x} = \{x_1, x_2, \dots, x_n\}^T$ , which may include geometric characteristics, materials properties, structural details, etc. For simplicity, the underlying variables  $x_i$  are assumed to be statistically independent; a reasonable assumption in most of the cases. Two different sampling techniques are

adopted herein: a DoE and a plain MCS. The methodological aspects of both techniques are discussed in this section, while their advantages and shortcomings are discussed in Section 5.3, considering the results obtained for the illustrative application. The adopted DoE is based on a  $3^n$  factorial sampling, where  $n$  is the number of input variables, considering mean plus or minus one standard deviation for each of them. This results in a relatively small sample size (e.g. 27 realisations considering three input variables), although this somehow requires that the underlying parameters  $x_i$  are normally distributed, or at least narrowly-distributed around the mean.

On the other hand, the adopted MCS explicitly uses the CDFs available for the considered input data; each CDF may be either analytical (i.e. a parametric model can be used) or empirical (i.e. non-parametric).  $N$  random realisations of the vector  $\mathbf{x}$  are defined based on the available CDFs. Such realisations will define the case-study buildings for which the fragility parameters are calculated. Since the derived fragility parameters are used to fit the GP regressions, it is not generally straightforward to a priori estimate an appropriate value for  $N$ . Therefore, a trial-and-error approach is discussed in Section 5.3.

### 3.2. Numerical modelling: time-history and pushover analyses

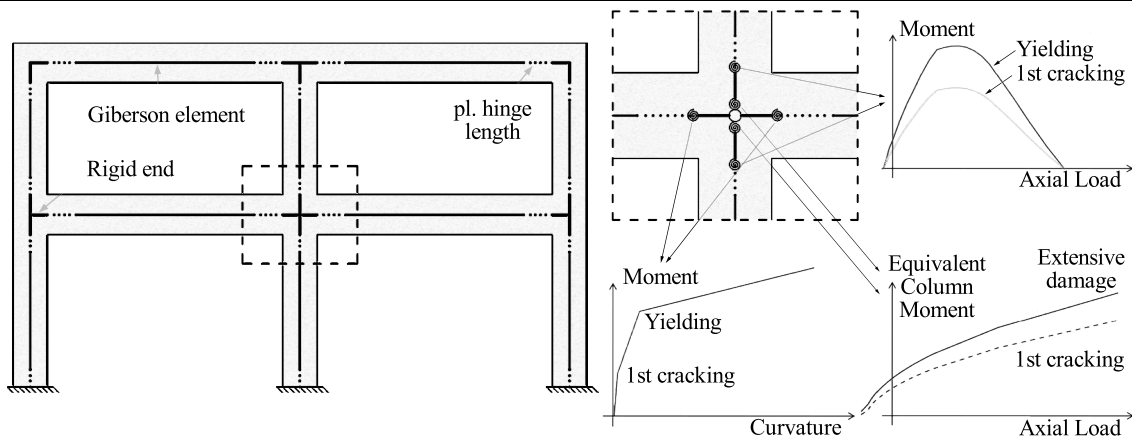
For each building realisation, refined numerical time-history analyses are carried out with the finite element modelling software Ruaumoko [30]. The adopted modelling strategy, previously validated on experimental results [31], is based on a lumped plasticity approach and is resumed in Table 2 and Figure 2. Floor diaphragms are modelled as rigid in their plane, and fully fixed boundary conditions are considered at the base. P-Delta effects are not modelled, given that the case-study buildings in the illustrative application are two-storey RC frames, for which P-Delta effects are negligible.

The flexural capacity of the RC members is derived using moment-curvature analysis. The flexural response is checked against other failure mechanisms that may significantly modify the member lateral response. The capacity model of the members included in the analysis is modified (if appropriate) to include those failure mechanisms accordingly. Flange effect is considered by increasing of 50% the negative moment capacity of the beams [11]. Lap splice failure, although not relevant for the illustrative application of this study, is checked according to the procedure described in Priestley et al. [32]. Shear capacity is checked [33], and the flexural response is modified, if required, according to the shear-drift law. The deformation related to bar buckling is calculated [34] and, if smaller than the flexural capacity, it is considered as ultimate point of the member response. The revised Takeda hysteretic model [35] is used for beams and columns, with the columns having a thinner loop. The hysteretic behaviour of the beam-column joints is modelled using the Modified Sina model [35], which is able to capture their pinching behaviour.

**Table 2 Details of the numerical modelling strategy.**

	<b>Modelling strategy</b>	<b>Characterisation</b>	<b>Notes</b>
<b>Beams</b>	Mono-dimensional Giberson elements [36] characterised with Moment-Curvature (end sections)	Moment-Curvature analysis and 50% increase in negative moment capacity (flange effect, [11])	CUMBIA software [37,38] for Moment-Curvature
<b>Columns</b>	Mono-dimensional Giberson elements [36] characterised with Moment-Axial load (end sections)	Moment-Axial load interaction analysis	CUMBIA software for Moment-Axial load

<b>Joints</b>	Non-linear moment-rotation springs connected with the rigid ends in the beam-column intersections	Equivalent Column Moment vs Axial load curve [11] for the non-linear springs	Drift limits based on experimental tests [11]
---------------	---	--	---



**Figure 2 Numerical modelling strategy** [12].

Cloud-based non-linear time-history analysis [39] is carried out for each building realisation. This consists of a series of non-linear time-history analyses adopting a large number of unscaled ground motion records. In this case, the SIMBAD database (Selected Input Motions for displacement-Based Assessment and Design, [40]) is adopted for this analysis. SIMBAD includes a 467 tri-axial accelerograms, generated by 130 worldwide seismic events (shallow crustal earthquakes with moment magnitudes ranging from 5 to 7.3 and epicentral distances up to 35 km). A subset of 150 records is considered here to provide a statistically significant number of strong-motion records of engineering relevance for the applications presented in this paper. As in [41], these records are arbitrarily selected by first ranking the 467 records in terms of their peak ground acceleration (PGA) values (by using the geometric mean of the two horizontal components) and then keeping the component with the largest peak ground acceleration. Such approach is compatible with the cloud analysis [42], and therefore it does not require a site-specific record selection. This is consistent with regional-scale seismic risk assessment of building portfolios.

Pushover analyses are also carried out, to have an intermediate level of refinement (and computational cost) in the seismic response estimation. The above-mentioned two-dimensional models are adopted to perform such analysis in displacement-control protocol, and applying a uniform force profile. Using the pushover curve, the CSM [43] is carried out to calculate the structural response. Specifically, the equivalent viscous damping formulation provided in [44] is used here to carry out the CSM, using the same 150 natural ground motion records described above. By adopting the displacement shapes provided in [12], the maximum inter-storey drift is calculated for each ground motion record.

### 3.3. Analytical modelling: Simple Lateral Mechanism Analysis

Among the selected approaches for estimating the seismic response of a given building realisation, SLaMA is the one with the lowest level of refinement (and computational cost). SLaMA [11–15] allows one to estimate both the plastic mechanism and



the capacity curve (i.e. a force-displacement curve) of RC frame, wall and dual-system buildings by using a “by-hand” procedure (i.e. using an electronic spreadsheet).

This is based on the calculation of the hierarchy of strength at sub-system level (beam-column joint sub-assemblies for frame structures) and the use of equilibrium and compatibility principles to “assemble” together the local results to obtain the global capacity curve. SLAMA enables the identification of potential structural weaknesses in the lateral resisting mechanism and to test the reliability of numerical computer models in capturing the behaviour of a structure. It is worth mentioning that each beam and column in the system is characterised considering many possible failure mechanisms (i.e. flexure, bar buckling, lap-splice failure, shear). The weakest mechanism will govern the overall member structural behaviour. The characterisation of the structural members is consistent to the approach adopted for pushover and time-history analyses (Section 3.2).

Using the SLAMA force-displacement curve, the CSM is carried out with the same suite of 150 natural ground motions described in Section 3.3. In this case, the displacement shapes provided in [12] are adopted.

### 3.4. Derivation of fragility relationships

Maximum inter-storey drift values derived with the three increasingly-refined analysis methods are independently adopted to derive fragility curves. The geometric mean of the 5%-damped pseudo-spectral accelerations in a range of periods (*AvgSA*) is the adopted IM. The period range considered in this study is discussed in Section 5.1, for the specific case-study application. It has been demonstrated that using *AvgSA* leads to an increased IM efficiency and sufficiency/relative sufficiency (e.g. [45,46], among many others) with respect to conventional ground-motion IMs. For this study, four DSs are considered, which are defined in detail in Section 4, with reference to the case-study application. Fragility curves are derived according to Eq. 1.

$$P(EDP \geq EDP_{DS}|IM) = P(EDP \geq EDP_{DS}|IM, NoC)(1 - P(C|IM)) + P(C|IM) \quad 1$$

Consistent with [47], the obtained IM-EDP pairs are partitioned in two sets: the “Collapse (C)” and “Non-Collapse (NoC)” cases. Collapse herein corresponds to a global dynamic instability (i.e. non convergence) of the numerical analysis, likely corresponding to a plastic mechanism (i.e. the structure is under-determined) or exceeding 10% maximum inter-storey drift. The total probability theorem is adopted to consider both “C” and “NoC” cases. In Eq. 1,  $P(EDP \geq EDP_{DS}|IM, NoC)$  is the conditional probability that the EDP threshold ( $EDP_{DS}$ ) is exceeded given that collapse does not occur, and  $P(C|IM)$  is the probability of collapse. It is implicitly assumed that ( $EDP_{DS}$ ) is exceeded for collapse cases, i.e.  $P(EDP \geq EDP_{DS}|IM, C) = 1$ .

The linear least square method is applied on the “NoC” pairs in order to estimate the conditional mean and standard deviation of EDP given IM and derive the commonly-used power-law model  $EDP = aIM^b$  [39], where  $a$  and  $b$  are the parameters of the regression. The derived probabilistic seismic demand model is used to define the median and logarithmic standard deviation of the lognormal distribution representing  $P(EDP \geq EDP_{DS}|IM, NoC)$  for each DS.

The probability of collapse  $P(C|IM)$  can be represented by a generalised regression model with a “logit” link function (logistic regression), which is appropriate for cases in which the response variable is binary (in this case, “collapse” or “no collapse”). More details on the mathematical formulation can be found in [47]. Although allowing for an approximation [47], it is decided to model the resulting fragility formulation, Eq. 1, with a lognormal distribution. The parameters of such distribution are the median IM ( $\mu$ ), equal to the original median in Eq. 1 and the logarithmic standard deviation ( $\beta$ ) calculated according to Eq. 2, in which  $IM_{16/84}$  are the 16<sup>th</sup> and 84<sup>th</sup> percentiles of the original distribution.

$$\beta = \frac{\ln(IM_{84}) - \ln(IM_{16})}{2} \quad 2$$

### 3.5. Gaussian process regressions

This study involves the definition of building realisations, each corresponding to a vector  $\mathbf{x}$  of building attributes. In this study, the vector  $\mathbf{x}$  includes the number of longitudinal bays ( $N_{bays,x}$ ), the concrete compressive strength ( $f_c$ ), and the steel yield stress ( $f_y$ ), i.e.,  $\mathbf{x} = \{N_{bays,x}, f_c, f_y\}^T$ . The choice of these specific building attributes is discussed in details in Section 4, with reference to the considered case-study building portfolio. In general terms, the vector  $\mathbf{x}$  may include any geometric characteristics, materials properties and/or structural details relevant to the specific application. For each building realisation, the median and dispersion of the fragility relationships (for a given DS) are calculated through numerical structural analysis, as discussed above. Each fragility parameter is denoted as  $y$ .

From a general, high-level perspective, the process to derive fragility relationships for each building realisation can be considered as a “black box” (or a “simulator”) which takes a vector of unique inputs ( $\mathbf{x}$ ) and produces an output/target ( $y$ ) described by the relationship  $y = f(\mathbf{x})$ . This simple equation form means that the simulator can encompass a wide range of specific analysis procedures and assumptions (various types of structural response analysis, various ground-motion record sets, various fragility fitting methods, etc.). Metamodels, and specifically GP regression in this study, aim to mimic the behaviour of the complex system (the simulator) by predicting the value of the target variable(s) given a combination of the input variables. In other words, a statistical model is adopted to surrogate the relation  $f(\mathbf{x})$  between inputs and output such that it is possible to make predictions for any input vector while remaining computationally cheap to evaluate.

The parameters and mathematical formulation defining the relationship between input and output variables are not known in advance. They are estimated from a smaller number of intelligently-chosen observations (i.e. a training set), containing pairs of input (covariates) and output data. In the specific application discussed here, a training dataset  $(X, \mathbf{y})$  can be defined, where the matrix  $X$  collects the training input vectors for all the considered building realisations, i.e.  $\mathbf{x} = \{N_{bays,x}, f_c, f_y\}^T$ , and the vector  $\mathbf{y}$  includes the corresponding fragility parameter ( $\mu$  or  $\beta$ ) for a given DS for each input vector.

This section presents a (non-exhaustive) overview of the definition and training of GP regressions. Interested readers are referred to literature books (e.g. [48]) for a comprehensive description of the subject. In a GP regression,  $f(\mathbf{x})$  is regarded as

a realised value of a gaussian process. A GP is a generalisation of the Gaussian probability distribution model, describing the distribution of functions  $f(\mathbf{x})$ , rather than scalars or vectors. A GP is fully specified by its mean function  $m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$  and its covariance function  $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]$ , where  $\mathbf{x}$  and  $\mathbf{x}'$  are two different input vectors:

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad 3$$

- The starting point to fit a GP regression is the definition of a prior distribution (in a Bayesian framework) for all the possible functions  $f(\mathbf{x})$ , which reflects the starting knowledge about the output before having any data. This is done assigning some properties of the mean and the covariance functions (e.g. smoothness) in Eq. 3. Therefore, higher probabilities are assigned to specific functions, for example because they are smoother than others, reflecting the physics of the problem under investigation. The prior distribution is converted into a posterior distribution (over functions) based on the observed data and such a posterior (or predictive) distribution is used for predictions.
- Generally, a zero mean function is assumed for the prior distribution (for simplicity). This does not represent a limitation, since the mean of the posterior process is not confined to be zero [48]. The properties of the output function  $f(\mathbf{x})$  - with particular reference to its smoothness - are governed by the covariance function (or kernel), which captures the correlation among different input vectors and reflects it in the output. For instance, for the specific application in this study, the fragility parameters are expected to be similar for two buildings having similar attributes.
- The structure of the covariance function is selected by the GP user and should reflect the expected behaviour of the output. A popular choice of covariance function is the squared exponential covariance – also adopted in this study – since it reflects the “stability” of the involved physical quantities (i.e. a small perturbation of the input geometry or materials produces small changes in the considered output fragility parameter):

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_i \frac{(x_i - x'_i)^2}{\sigma_i^2}\right). \quad 4$$

- The parameters of the covariance function are called hyperparameters:  $\sigma_i^2$  and  $\sigma_f^2$ , in this case. The  $\sigma_i^2$  parameters are related to the length scale of the output in each input dimension  $i$ . Given the distance between two input attributes in the dimension  $i$  ( $x_i - x'_i$ ), the length scale describes the rate of decay of the correlation between output values at increasingly-separated values of the input (i.e. for small  $\sigma_i^2$  values, the output fragility parameter can change quickly with respect to the building attributes).  $\sigma_f^2$  is the signal variance, which is the variability of the output for a given input (i.e. the covariance function is equal to  $\sigma_f^2$  when the distance between  $\mathbf{x}$  and  $\mathbf{x}'$  is zero). For this specific problem, the output variability is introduced since three scalar input parameters are used as a proxy of a refined numerical model of the structure and of the whole fragility derivation process.

The estimation of the hyperparameters is typically achieved by maximising the likelihood,  $p(\mathbf{y}|X, \boldsymbol{\theta})$ , which is the probability of predicting the training data  $\mathbf{y}$ , given the training input dataset  $X$  and a suite of hyperparameters  $\boldsymbol{\theta} = \{\sigma_i, \sigma_f\}$ . In this study, this is carried out adopting a quasi-Newton numerical optimisation algorithm [49] in Matlab [50].

- 5 Given the hyperparameters, the predictive distribution introduced above is used to estimate the output for a given test dataset, i.e. a collection of arbitrarily-selected input vectors (not included in the training dataset) for which the output (a fragility parameter for a given DS) is wanted. Based on the GP properties, the predictive distribution is a  $n + d$  dimensional joint Gaussian distribution, where  $n$  is the number of points in the training dataset and  $d$  is the number of points in the test dataset. This distribution has mean ( $\bar{\mathbf{y}}_*$ ) and covariance ( $\text{cov}(\mathbf{y}_*)$ ) defined as:

$$\mathbf{y}_*|X, \mathbf{y}, X_* \sim \mathcal{N}(\bar{\mathbf{y}}_*, \text{cov}(\mathbf{y}_*)) \quad 5$$

$$\bar{\mathbf{y}}_* = K(X_*, X)K(X, X)^{-1}\mathbf{y} \quad 6$$

$$\text{cov}(\mathbf{y}_*) = K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*). \quad 7$$

- 10 The test dataset is  $(\mathbf{y}_*, X_*)$ , where  $X_*$  is the matrix of the collected input vectors  $\mathbf{x}_*$  (arbitrarily selected), each corresponding to one component of the vector  $\mathbf{y}_*$  of the surrogated output, and  $(X, \mathbf{y})$  is the training dataset.  $K$  is the covariance matrix, obtained applying Eq. 4 to the relevant combinations of the inputs. It is worth mentioning that no noise is introduced in the surrogate model since the fragility parameters in the training set are derived with deterministic numerical simulations. Equations 6 and 7 can be directly used for predictions using matrix operations.
- 15 In summary, eight parameters (median and dispersion for each DS) defining seismic fragility for a given building realisation are available based on the procedure described in Sections 3.1 to 3.4. The same parameters are calculated based on three increasingly-refined approaches (SLaMA, pushover, time-history analysis) for the seismic response estimation, and for each building class. One GP regression is fitted for each of those parameters, leading to a total of  $8 \times 3 \times 2 = 48$  GP regressions.

#### 4. Illustrative application

- 20 The case-study buildings selected here represent seismically-deficient RC school buildings, defined based on a large data-collection exercises performed by the authors [51], involving rapid visual surveys for over 200 school buildings. Specifically, a rapid visual survey was carried out for each building to collect administrative (i.e. location, year of construction/retrofit, number of students, teachers), geometric (i.e. member dimensions) and mechanical data (i.e. structural details, nominal material properties, if known). The analysis of the collected data, coupled with simulated design according to the relevant building code (give the year of construction), allows one to firstly define building classes/archetypes based on the most
- 25

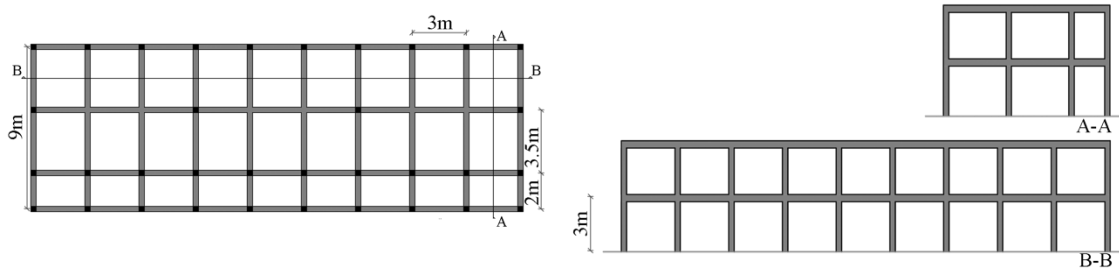
common lateral load resisting system, average geometrical and mechanical features and structural details. Then, the distributions of the collected data are used to generate realisations of the buildings (through DoE or MCS) within each class, selecting the most relevant varying attributes and their samples.

#### 4.1. Archetype building definition

5 The archetype building geometry (Figure 3) refers to a two-storey rectangular-plan RC frame, representing approximately 80% of the surveyed schools. Based on the collected geometrical data, the parameter with a significant variation is the number of longitudinal bays, while other parameters such as the number of storeys, the length of the transverse bays, the dimension of the beams/columns have negligible variability within the surveyed sample.

The knowledge on the material properties is comparatively lower with respect to geometry, given the visual nature of the data  
 10 collection. Indeed, average values related to Southeast Asian statistics (e.g. [52]) for the concrete cylindrical strength and steel yield stress are used (average values of 24MPa and 400MPa, respectively). Moreover, coefficients of variation (CoV) respectively equal to 18% and 5% are selected based on literature studies [53,54]. This reflects the assumption that all the buildings in the considered portfolio are characterised by the same nominal value of material properties. Hence, the building-to-building variability coincides with the within-building variability. It is worth mentioning that both the analytical and  
 15 numerical models adopted for the study are significantly affected by such quantities, since those affect the flexural and shear characterisation of beams, columns and beam-column joints.

With respect to the level of structural detailing, two building classes (sharing the same archetype building geometry) are defined. Specifically, a simulated design approach is employed to reflect two different nominal seismic performances. This results in two index buildings designed according to the Uniform Building Code, UBC [55], and the American Society of Civil  
 20 Engineers (ASCE) 7-10 standards [56], respectively. In fact, building codes in several developing countries typically refer (and are fully consistent with) to the UBC and/or the United States codes [51]. However, some of the provisions in such codes (e.g. stirrups in the joints) have not been fully implemented in the simulated design to conservatively take into account the potential lack of code enforcement (often observed during the field survey). The resulting detailing (Table 3) leads to the definition of a *pre-Code* and a *low Code* design level/building classes, consistently with the HAZUS MH4, HAZard United  
 25 States [57], definitions.



**Figure 3 Geometry of the archetype building.**

**Table 3 Detailing of the members in the two variations of the archetype building.**

	Typical beam	Typical column	Typical joint
Pre-Code	3 $\phi$ 16 top layer	3 $\phi$ 16 top layer	No stirrups
	3 $\phi$ 16 bottom layer	3 $\phi$ 16 bottom layer	
	$\phi$ 10@150mm hooks	$\phi$ 10@200mm hooks	
Low Code	3 $\phi$ 16 top layer	3 $\phi$ 16 top layer	No stirrups
	3 $\phi$ 16 bottom layer	3 $\phi$ 16 bottom layer	
	$\phi$ 10@150mm hooks	$\phi$ 10@100mm hooks	

## 4.2. Sampling definition

According to the available information for each considered building class (from the rapid visual survey), the distributions of the geometric parameters related to the longitudinal direction of the building show a higher variability than those related to the transverse direction. For this reason, this study focuses on the longitudinal direction of each archetype building, for which an accurate mapping of the input parameters to the fragility parameters is expected to have the highest impact on any portfolio risk assessment. Both the *pre-code* and *low code* building classes are investigated and, for each, the three considered parameters for the sampling are: 1) number of longitudinal bays,  $N_{bays,x}$ ; 2) concrete cylindrical strength,  $f_c$ ; and 3) steel yield stress,  $f_y$ . As discussed above, two sampling techniques, MCS and DoE, are independently used to simulate alternative scenarios in which information regarding the statistical distribution of the building input parameters is respectively (partially) unavailable or available.

It is worth stressing that the statistical distribution of the input parameters could affect, although only indirectly, the GP predictive power, but this would depend on the specific problem (in particular, on the sensitivity of the structural response to perturbations of the input). Indeed, if the structural response is particularly “flat” and smooth with respect to the inputs, any input distribution (for a given range) would likely require a similar number of samples to provide a similar accuracy of the GP regression. For a very “sensitive” structural response, conversely, a complex input distribution may require a higher number of samples to obtain the same accuracy. However, this is much more affected by the output rather than the input. When adopting GP regression, the focus is the input-output map, rather than the input distribution [48]. Therefore, the desired level of accuracy (using a quantitative metric) should drive the dimensioning of the sample set.

The definition of the DoE sampling is firstly shown in Table 4. By rounding the mean and standard deviation of the measured values of the number of bays, case studies with 6, 9 and 12 bays are considered. On the other hand, no direct information about the distributions of both concrete cylindrical strength and steel yield stress are available based on the rapid visual survey. Therefore, the nominal material properties, and the related statistics, to be used in the study are based on the year of construction (known from the rapid visual survey) and the available Southeast Asian data [52]. On this basis, the same nominal values are selected for the two building classes, and the distributions of both concrete cylindrical strength and steel yield stress are modelled by means of Normal distributions with mean and standard deviations based on the above-mentioned references. For

each considered building class, 27 building realisations are adopted, each analysed using 150 ground motions (i.e. 4050 seismic response analyses are performed).

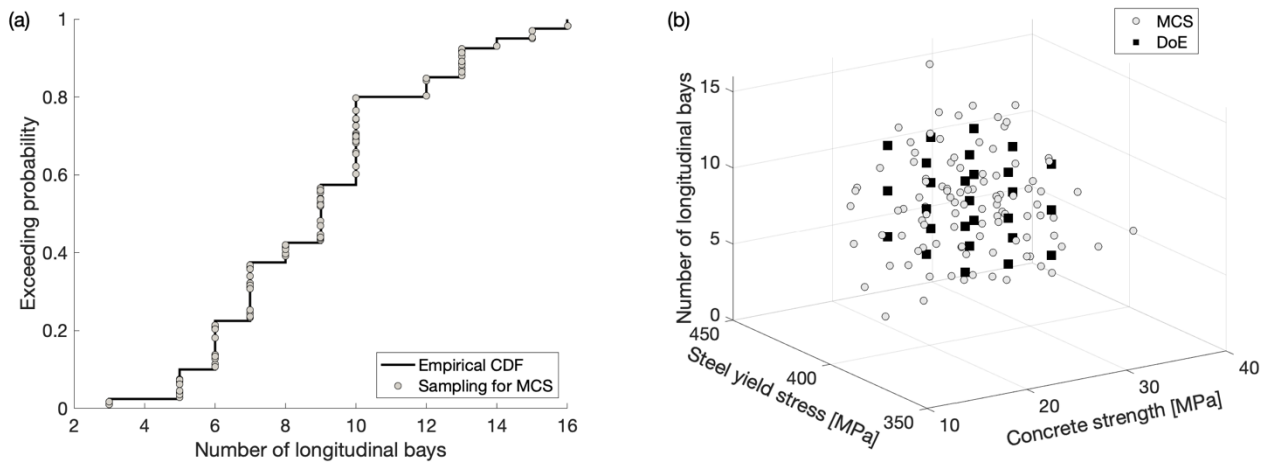
**Table 4 Sampling based on Design of Experiment.**

Variable		Samples		
		Mean – s. deviation	Mean	Mean + s. deviation
Number of longitudinal bays	$N_{bays,x}$ [-]	6.0	9.0	12.0
Concrete cylindrical strength	$f_c$ [MPa]	19.7	24.0	28.3
Steel yield stress,	$f_y$ [MPa]	380.0	400.0	420.0

5

Figure 4 summarises the sampling results based on MCS for the two categories of structural details (i.e., the two building classes). 100 random samples are independently extracted for the three input parameters. As an example, Figure 4a shows the empirical CDF for the number of longitudinal bays, and the adopted values for the sampling. The same concept is applied adopting the analytical CDF related to the material properties. This leads to the sampled vectors represented in a three-dimensional space in Figure 4b, which is compared to the DoE. A total of 200 building realisations are analysed, considering the two building classes, leading to 30,000 seismic response analyses for each of the three refinement levels of the analyses. The MCS sampling can more-accurately represent any distribution of the input parameters, provided that a sufficiently large number of samples is drawn. If the distributions of the input parameters are available, this may increase the efficiency of the GP regression: the “resolution” of the metamodel can be higher (i.e. more training samples) for the most frequent combinations of the input parameters.

15



**Figure 4 a) 100 realisations (MCS) for the number of longitudinal bays. b) DoE (27 samples) vs MCS (100 samples) comparison (valid for both building classes).**

## 5. Results and discussion

### 5.1. Estimated database of fragility functions

Fragility analysis is carried out for each building realisation, based on both DoE and MCS sampling techniques. Figure 5 shows all the results based on time-history analysis. The period range for the definition of  $AvgSA$  is defined as  $(T_{1,min}; 1.5T_{1,max})$ , where  $T_{1,min} = 0.38s$  is the minimum first-mode period for the entire database while  $T_{1,max} = 0.53s$  is the maximum. This allows one to obtain consistent results (i.e. using the same ground-motion IM for all the building realisation), with a minimum loss in terms of IM efficiency. Fragility curves are calculated for four different DS (Slight Damage, Moderate Damage, Extensive Damage, Complete Damage), which are based on the HAZUS definitions [57]. The corresponding drift limits, [0.25, 0.6, 1.5, 2] %, are assumed as representative of the entire building class. Those are quantified using the pushover analysis results (also shown in Figure 6a).

The results clearly show the effect of the building-to-building variability on the seismic fragility. As expected, such effect is much more evident in the most severe damage conditions (DS3, DS4), rather than in the elastic or near-elastic conditions. It is also evident that, for this particular case, the structural detail type has a minor effect on the fragilities. On average, the DS4 fragility median (time-history analysis) is equal to 1.23g for the *pre-code* and 1.31g for the *low code* building classes, respectively. Regardless of the improved structural details in beams and columns (*low code* vs *pre-code*), the plastic mechanism expected for both building classes (also shown in Figure 6a) is governed by failure in the beam-column joints at the first storey. This explains such minor differences among the fragilities of the two classes.

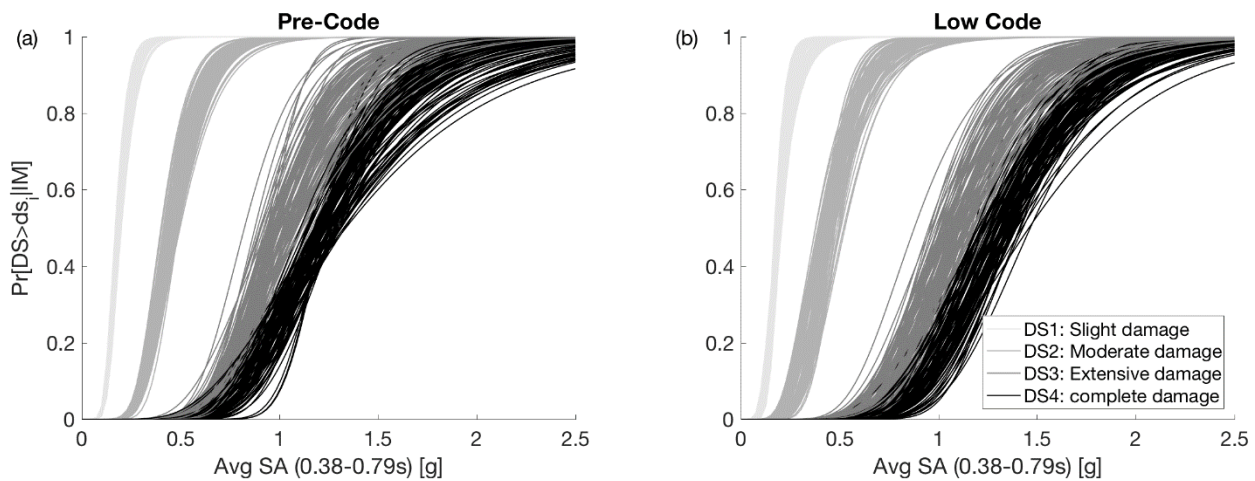


Figure 5 Fragility curves based on time-history analysis of the whole dataset: a) Pre-code building class; b) Low code building class.

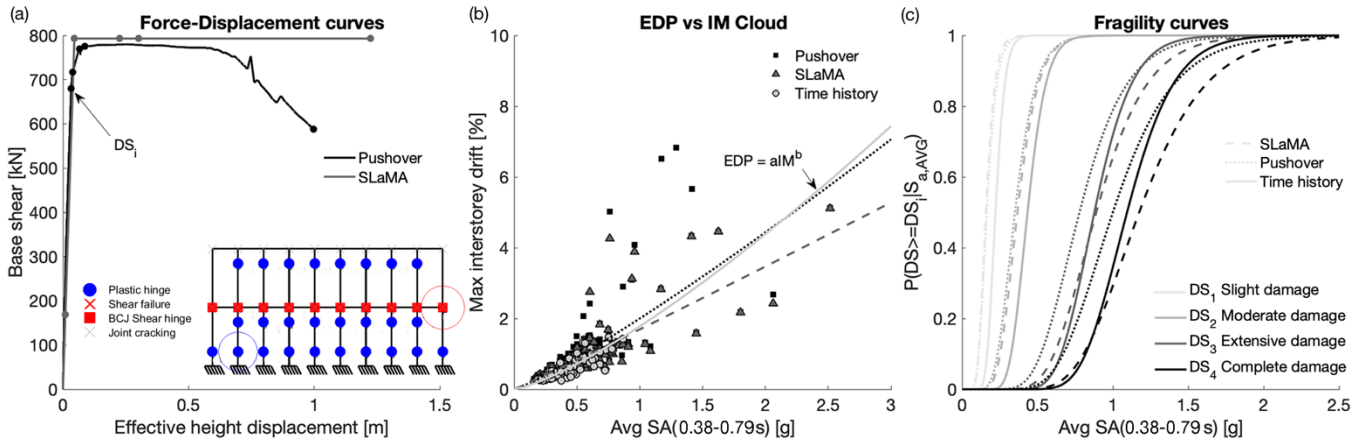


The fragility curves computed with Eq. 1 have slightly different values of the dispersion for different damage states. Indeed, although the non-collapse fragility curves,  $P(EDP \geq EDP_{DS}|IM, NoC)$ , share the same dispersion regardless of the damage state (because of the considered cloud approach), the final value of this parameter is generally smaller for higher DSs (due to the higher number of collapses for high levels of the ground-shaking intensity). Different values of the dispersion cause intersections in the tails of such fragilities, which imply, for some values of the IM, a negative probability of attaining of one or more DSs. This effect is more emphasised when the difference in two fragility dispersions is higher. This issue is practically negligible for DS1, DS2 and DS3 fragilities, which have particularly similar values of the dispersion. Within the entire training dataset, the dispersion of the DS4 fragility is smaller than the DS3 one due to the higher number of collapses (0.093 maximum difference). Considering that, on average, the DS4 median fragility is 18% higher than the DS3 one, such a difference in beta is causing crossings in those fragility curves only for high values of the intensity measure, which would correspond unphysically-large (with respect to a given hazard curve) ground motions. Therefore, no intersection of the fragility relationships would be observed within the range of the considered hazard levels (in terms of the selected IM), thus not affecting the results of a building-portfolio seismic risk assessment.

## 5.2. Accuracy of the non-linear static analysis methods

As described in Section 3, three increasingly-refined analysis methods are adopted for each building realisation, both for the DoE and MCS realisations. It is worth stressing that the simplified methods are adopted to increase the feasibility of the GP regression approach for a high number of building classes. If a single family of fragility relationships for each building class is needed (outside the scope of the paper), it could be also possible to combine the results of the three increasingly-refined response analysis methods. For example, this can be done through Bayesian framework for robust seismic fragility assessment based on various model classes, currently under development by the authors [58].

Figure 6 summarises the fragility analysis results for a *low code* building-class realisation. Firstly, it is evident that the analytical approach (SLaMA) provides a particularly-good approximation of the numerical pushover, both in terms of plastic mechanism and force-displacement curve (Figure 6a, which also shows the adopted DS for the fragility analysis). The two approaches diverge for very high displacements, well beyond the ultimate limit state of the frame, where strength degradation (not considered in SLaMA) takes place. Such a minor discrepancy is reflected in Figure 6b, where the EDP vs IM cloud is shown, as predicted according to the CSM adopting the selected 150 ground motions as an input. Indeed, the largest SLaMA vs Pushover discrepancies for maximum inter-storey drift are recorded for high values of the *AvgSA*, which forces high drift demand to the structure. Interestingly, a very good match is observed, overall, between the non-linear static methods and the non-linear dynamic approach, considering the simplified nature of the former methods. For this specific building realisation, the highest error on the estimation of the seismic fragility is registered for DS4: +7.3% and -8.7% with respect to the time-history analysis for the median values based on SLaMA and pushover analysis, respectively (+30.2% and +38.0% for the dispersion).



**Figure 6** Analysis results for the Low-Code, 9-bays,  $f_y=400\text{MPa}$ ,  $f_c=19.7\text{MPa}$  building realisation. a) force-displacement curves, b) EDV vs IM Cloud, c) fragility curves.

The bias of the non-linear static methods (with respect to the non-linear time-history analyses) in the estimation of seismic fragility is systematically quantified for the 254 considered building realisations (100 for the MCS and 27 for the DoE, for each building class). As shown in Figure 7, this is done by means of the static-to-dynamic ratio  $F$  calculated for each fragility parameter (median  $\mu$  and logarithmic standard deviation  $\beta$  for each considered DS, 1:4, for SLaMA, S, and pushover, P). Due to the high similarity between the two building classes, those are considered together in Figure 7.

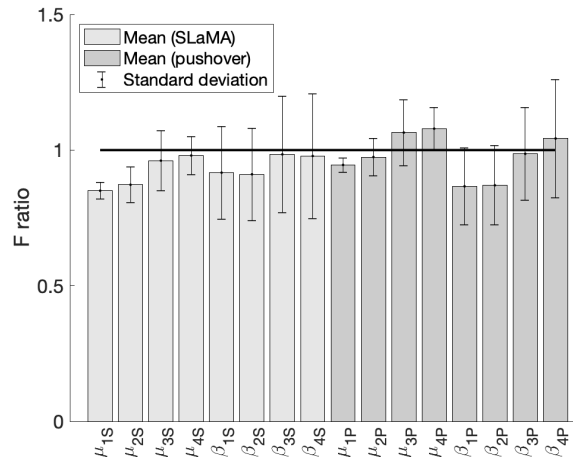
The SLaMA-based fragility relationships are generally slightly conservative: the mean  $F$  ratio ranges between 0.85 and 0.98, while the standard deviation falls within the range [0.03; 0.27]. On the other hand, the pushover-based mean  $F$  ratio is in the range (0.86, 1.08), with standard deviation in the range (0.03, 0.22).

For both methods, the standard deviation registered for the fragility dispersion (maximum 0.27) is considerably higher than the one for the median (maximum 0.12). This is because the non-linear static methods cannot capture any dynamic instability, which can be particularly sensitive to the geometric/mechanical properties of the building. This creates a great variability in the non-linear dynamic-based results, which is not captured in the non-linear static-based ones, therefore leading to a greater standard deviation in the  $F$  ratios for the logarithmic standard deviation.

The bias of the DS1 and DS2 fragilities (minimum 0.85), both SLaMA- and pushover-based, is higher than for DS3 and DS4 (maximum 1.08). This is due to the lower order of magnitude of the former fragility medians (0.1, 0.3) if compared to the latter ones (0.9, 1.2). In absolute term, the error level is comparable for all the considered DSs.

The results clearly indicate that, for these specific building classes, non-linear static methods show error levels within acceptable limits for portfolio-scale analysis. Those are shown here to demonstrate the scalability (i.e. high number of building classes) of the metamodeling techniques combined with simplified response analysis methods. However, the error trends ( $F$  ratios) are not yet generalisable and require a much wider validation, currently being investigated by the authors [10]. In

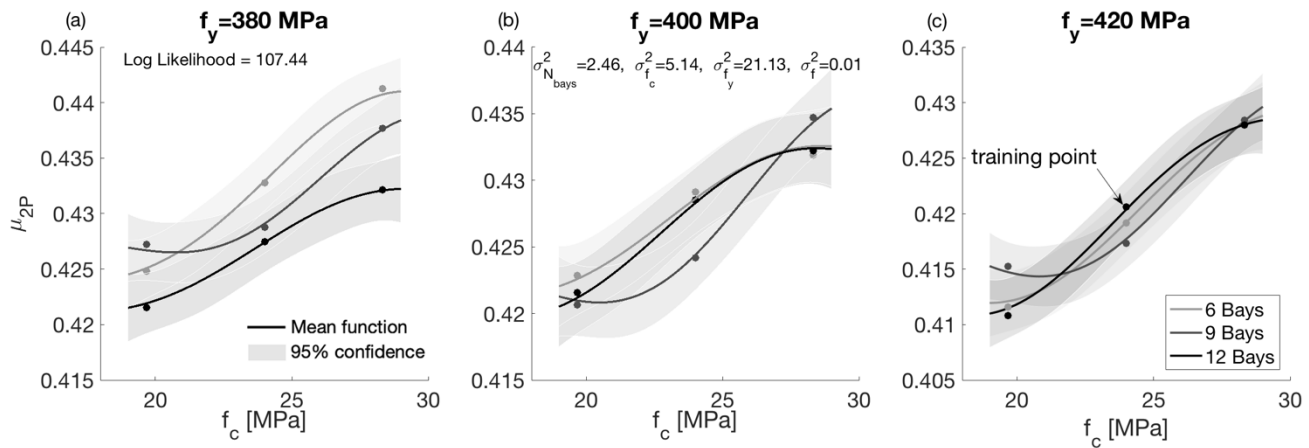
particular, a systematic validation of the CSM method using natural ground motions is needed, together with a validation of the overall fragility estimation for different building typologies.



5 **Figure 7 Static-to-dynamic ratio for the fragility parameters (pre-code and low code building classes are considered together). P: pushover vs time history; S: SLaMA vs time history.**

### 5.3. Predictive power of the GP regressions

One GP regression is fitted for each fragility parameter ( $\mu_{1-4}, \beta_{1-4}$ ), for each analysis method (SLaMA, pushover, time-history analysis; indicated by the subscripts S, P, T), for each sampling technique (DoE, MCS), and for each building class. As an example, Figure 8 shows the DoE-based GP that surrogates the pushover-based, DS2 median fragility of the *pre-code* building class  $\mu_{2P}$ . The four-dimensional function is “sliced” for three different values of the steel yield stress ( $f_y$ ), represented in each panel of the figure, and further sliced for three different values of the number of bays ( $N_{bays,x}$ ), represented by each line in the panels. It is evident that, after the maximisation of the likelihood function, the mean function of the GP can interpolate the observed data. The shaded areas in Figure 8 represent the 95% confidence bounds of the predictions. Although slightly higher uncertainty is expected for predictions at intermediate values between two observed points in the DoE, the highest observed 95% confidence range is approximately equal to  $\pm 1\%$  with respect to the mean. The central panel of this figure also shows the adopted hyperparameters for the kernel (length scale in each input dimension,  $\sigma_{N_{bays}}^2, \sigma_{f_c}^2, \sigma_{f_y}^2$ , and signal variance  $\sigma_f^2$ ).



**Figure 8 Pre-code GP regression (DoE-based) for the DS2 fragility median based on pushover. Steel yield stress: a) 380MPa, b) 400MPa, c) 420MPa.**

As mentioned in Section 5.1, if fragility relationships for two different (adjacent) DSs are characterised by particularly different dispersion values, fragility intersections occur. Although this is not a concern within the training set, larger differences in the fragility dispersions may occur when making predictions with the GPs outside the training set (depending on the hyperparameters of the fitted GPs). At the expense of using a considerably-more complicated model, this issue could be solved with a multi-output GP regression [59] as opposite to independent GPs for each fragility median and dispersion. By carefully specifying the structure of the cross-covariance function (i.e. performing kernel design), together with each single covariance function, it is possible to define constraints between some hyperparameters, in such a way that the dispersions of the fragility curves dispersions among different GPs are effectively constrained. In the opinion of the authors, this more refined model may produce negligible improvements in the accuracy of the risk estimates, since fragility intersections occurring within the range of the considered hazard levels are reasonably unlikely. On the other hand, due to its higher complexity, the multi-output GP approach may be less attractive for practical applications.

As a practical measure to control such issue, it is instead proposed to make predictions with the calibrated single-output GPs for selected combinations of the input parameters, and compute the difference between the DS3 and DS4 fragility dispersions. The input combinations should be selected uniformly at random within the considered range of the building input parameters. If for a particular input vector such difference is higher than a reasonable threshold based on engineering judgement (0.15 in this case), an additional sample should be added in the training dataset before re-fitting the GPs.

Specific validation datasets are defined for the GP regressions. For each building class, the validation dataset adopted for the MCS-based GP regressions is the training dataset adopted for the DoE-based ones (27 samples for each GP). On the other hand, the validation dataset for the DoE-based GP regressions is composed of 30 samples extracted to the MCS-based sampling, whose input values fall within the limits of the DoE itself (i.e. the fitted GPs are not meant to be used for extrapolation). The predictive power of all the fitted GPs is systematically evaluated by means of two metrics (Figure 9): the

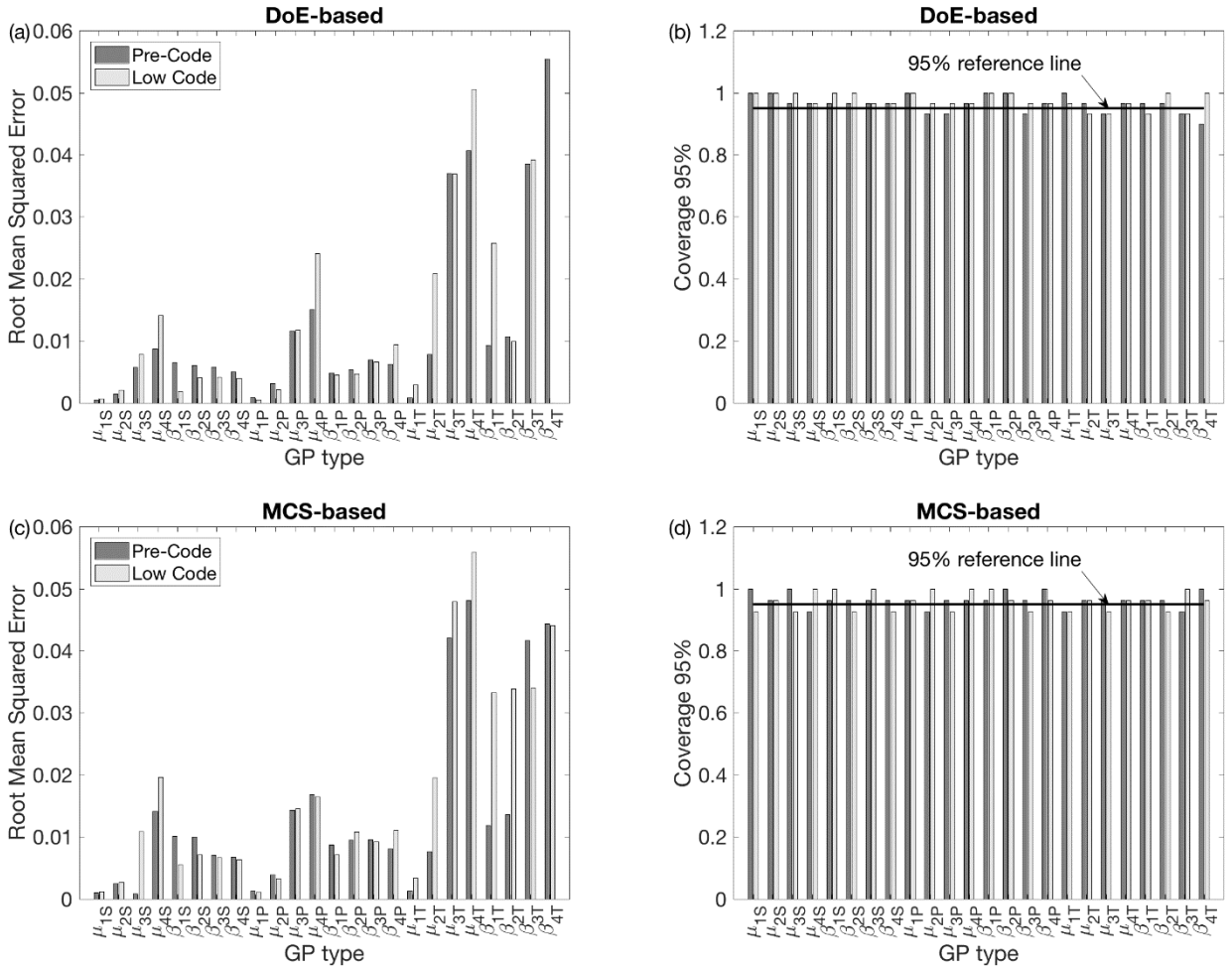
surrogated-vs-computed root mean squared error (RMSE); and the coverage probability of the 95% prediction intervals (95% coverage) of the fitting. The latter statistics is defined as the proportion of the confidence intervals of the surrogated model that contains the modelled data. The numerical values of such quantities, for all the fitted GPs, is given in Table A 1 and Table A 2 in Appendix A. Such choices allow for a fair comparison between the DoE- and MCS-based GP regressions, which is the ground for a critical comparison of advantages and shortcomings of the two approaches.

Figure 9 clearly shows that, regardless of the adopted sampling technique, the fitted GPs have a high predictive power in surrogating the computed results. Indeed, the RMSE (in the units of g for the fragility median) is smaller than 0.06 for all the investigated parameters, regardless of the structural-detail category and the sampling technique. Moreover, the calculated 95% coverage probability falls in the range (92.6%-100%), which is reasonably close to the theoretical value of 95%.

The RMSE related to the MCS-based GP regressions shows similar trends but slightly-higher values if compared to the DoE-based GPs results, even considering the larger training dataset (100 vs 27 building realisations). No remarkable differences are observed between the trends for the *pre-code* and *low code* building classes. As explained in Section 5.2, this reflects the small differences in the expected plastic mechanisms for the two categories, both governed by the joint panels.

The RMSE related to the non-linear static-based GPs is approximately two to three times smaller than for the time-history analysis-based GPs. This is due to the observed cases of dynamic instability, which increase the variability of the time-history analysis-based training dataset, resulting in a higher noise variance for the corresponding GP regressions. Such higher noise variance avoids having particularly-low values of the signal variances, effectively avoiding overfitting. Being an absolute quantity, the RMSE of the fragility medians shows an increasing trend with the considered DS, while it is rather constant for the logarithmic standard deviations. One exception to this trend is represented by the time-history analysis-based logarithmic standard deviations, for which the RMSE is higher for DS3 and DS4. Again, this is due to the observed cases of dynamic instability, which increases the variability of the training dataset.

To have a dimensionless metric for the error, the RMSE of each surrogated fragility parameter is also divided by the mean of the actual fragility-parameter values in the validation set. A maximum normalised RMSE equal to 9.2% is recorded, among all the considered fragility parameters, analysis methods and sampling techniques. An error smaller than 10% is generally deemed satisfactory for most of the practical seismic risk applications, especially considering the other sources of uncertainty involved in several practical applications (e.g. record-to-record variability, in this case). As a practical rule, additional samples could be added to the training dataset if such error is higher than 10%.



**Figure 9 Predictive power of the GP regressions: a) DoE-based Root Mean Square Error, b) DoE-based 95% coverage, c) MCS-based Root Mean Square Error, d) MCS-based 95% coverage.**

Overall, the DoE-based GP regressions appear to be more effective than the MCS-based ones. They require a smaller number of analyses to generate the training data while showing a predictive power of the same order of magnitude of the MCS-counterparts. Moreover, they allow a visual validation of the results (e.g. Figure 8), which is less effective for the MCS-based GP regressions. However, in the present form, the DoE-based GP regressions cover a smaller range of the input variables, if compared to the MCS. Adopting a different, wider DoE would involve a higher computational burden. As an example, a DoE based on four samples in each input dimension would require  $150 \times 3^4 = 12,150$  seismic response analyses, while the adopted MCS-based approach requires 15,000.

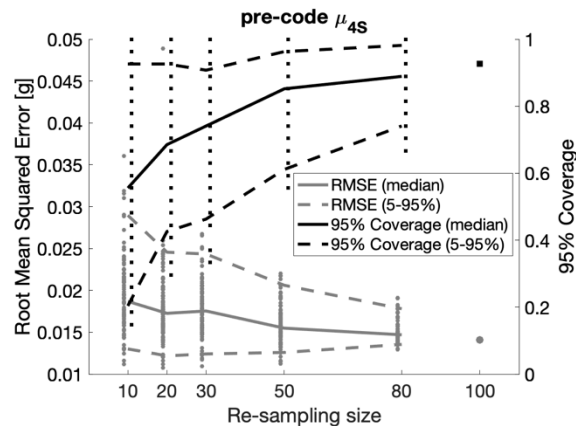
Based on the results of this specific application, it is fair to suggest that the choice between the two approaches can be based on the (evolving) availability of the input data. In situations for which such availability is limited, a DoE sampling technique

would be preferable, for example based on (assumed) uniform distributions for the input data, based on engineering judgment. On the other hand, if more detailed data is available, the MCS sampling technique could be more appropriate. In this way, the “resolution” of the GP regression can be higher in regions of the input space where more sampling will be needed when using the GP for portfolio risk assessment applications. Moreover, if the knowledge about the input data increases with time, an advantage of the MCS-sampling technique is the possibility to enlarge the training dataset by drawing more samples according to the new information.

#### 5.4. Sensitivity of the GP predictive power to the training sample size

The last part of this study investigates the sensitivity of the GP predictive power to the size of the training dataset. To study this, smaller sub-sets are generated from the MCS-based training datasets (100 samples per each building class). The chosen re-sampling sizes are equal to 10, 20, 30, 50 and 80, respectively. For each re-sampling size, 100 sub-sets of that size are uniformly generated. For each sub-set, GP regressions are estimated for each building class and fragility parameter, calculating the RMSE and 95% coverage. This allows to map the predictive power of the GP regressions as a function of the sample size (reported in Table A 2). It is worth mentioning that the sub-sets are generated without replacement, to allow a fair comparison among the sub-sets with the same re-sampling size. Indeed, if a sub-set of nominal size  $N$  has two equal samples, the training of the related GP regressions would be effectively based on  $N - 1$  samples, rather than  $N$ .

An example of the mapping is provided in Figure 10, which shows the calculated RMSE and 95% coverage for the DS4, SLaMA-based fragility median of the *pre-code* building class, including their median and 95% confidence intervals. The figure shows that increasing the re-sampling size leads to a minor improvement in terms of RMSE, while the greatest gain is related to the 95% coverage probability.

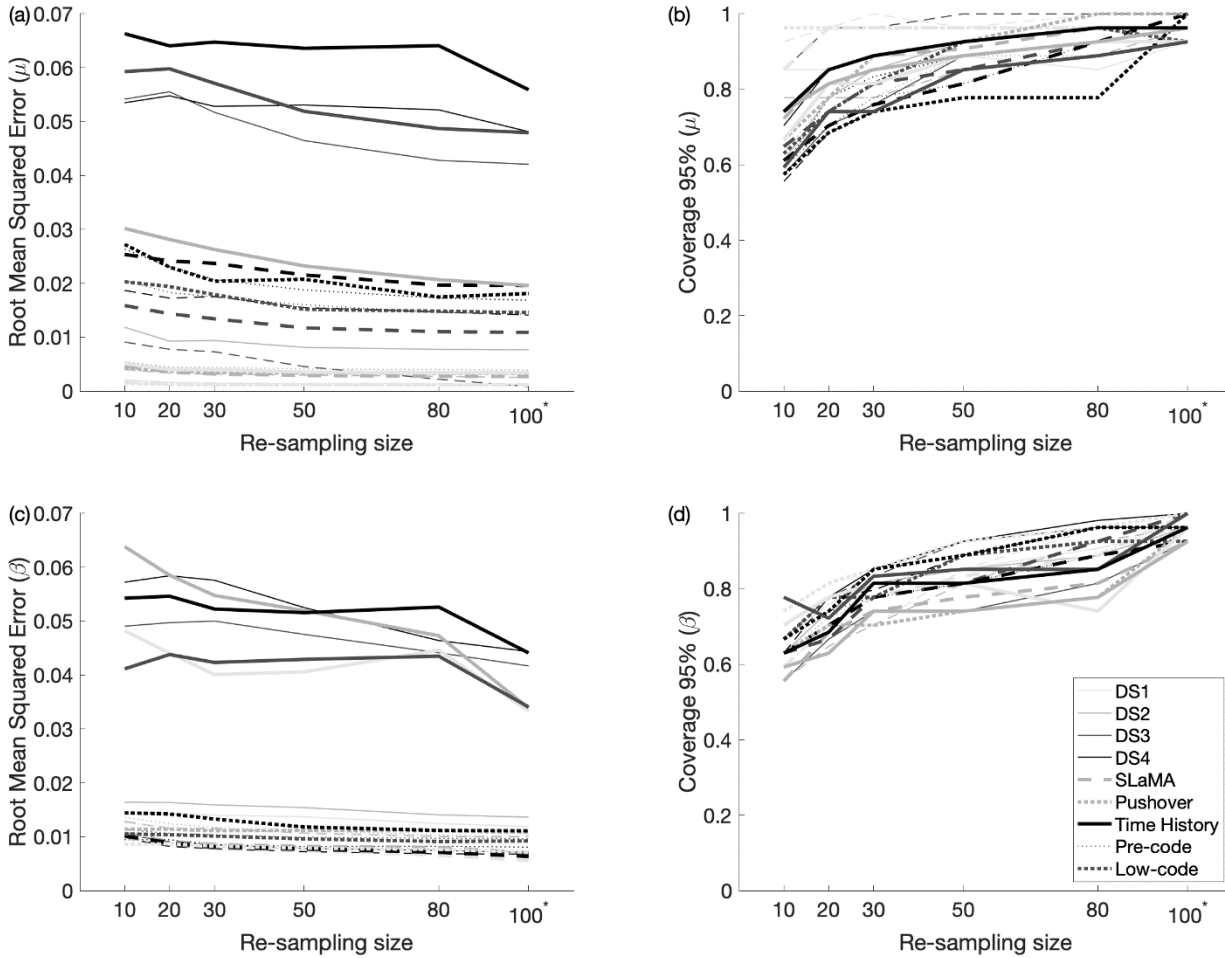


20

Figure 10 GP statistics for the DS4 fragility median based on SLaMA (Monte Carlo sampling). \*based on the full training dataset; no resampling.

The full picture of the sensitivity analysis is provided in Figure 11, which shows the median curves for the RMSE and 95% coverage for all the fragility parameters. Firstly, the higher RMSE values reported in Section 5.3 for the time-history analysis-based GP regressions are confirmed here, regardless of the considered building class and DS.

A relatively-horizontal trend is shown for the RMSE, both for the medians and the logarithmic standard deviations. This applies regardless of the considered building class, seismic response analysis method, and DS. A decrease of the RMSE is registered when moving from a re-sampling size of 80 to 100. However, the data related to the re-sampling size of 100 refers to a single GP fitting, rather than the median over 100 sub-sets, and it is plotted for completeness only. No inference can be made with regard to the 80-100 branch of the curves.



10 **Figure 11** Statistics of the GP regressions based on Monte Carlo sampling: median for 100 sub-sets. Root mean square error for the median fragility (a) and the dispersion (c); 95% coverage for the median fragility (b) and the dispersion (d) . \* based on the full training dataset; no resampling.



The real gain in increasing the re-sampling size is seen in terms of 95% coverage, which shows a rather linearly-increasing overall trend for all the considered fragility parameters. The 95% coverage changes from the range (66.7%-96.3%) to (92.6%-100%) for a re-sampling size of 50 and 100, respectively.

5 These results are limited to the adopted case study. However, the procedure employed for this sensitivity analysis can be suggested for practical applications, to achieve the desired level of predictive power of the GP regressions. In particular, one can first base the full process on a small number of building samples (e.g. 20-30), and then gradually increase such number until the desired level of 95% coverage is achieved. Clearly, statistics for the RMSE and 95% coverage will be available only for the first  $N - 1$  values of the re-sampling size.

## 6. Conclusions

10 This paper proposed using gaussian process (GP) regressions to develop flexible and accurate metamodels that can be used to account for building-to-building variability in simulation-based seismic risk assessment of building portfolios. This involves analysing a small number of samples (i.e. building realisations of the same class) and subsequently fitting GP regressions to the seismic fragility parameters obtained through cloud-based non-linear time-history analysis of each building realisation. Significant computational effort may still be needed to calibrate such metamodels for a single structural type, i.e. thousands of  
15 time-history analyses are needed. To increase the tractability of the approach, two simplified-yet-accurate seismic response analysis methods have been proposed as an alternative to non-linear time-history analysis. In particular, non-linear static force-displacement curves are derived through the analytical method SLaMA or through a standard numerical pushover (with invariant force profile). The Capacity Spectrum Method is adopted to derive the seismic demands.

An illustrative application of the proposed framework has been shown for seismically-deficient reinforced concrete school  
20 buildings with construction details typical of developing countries in Southeast Asia, for which the empirical distributions of geometrical/mechanical data are available. Two different sampling techniques have been independently adopted to sample realisations within the analysed building class: a Design of Experiment (DoE) and a plain Monte Carlo sampling (MCS). This is done to simulate alternative scenarios in which the probability distributions of the building input parameters are available (MCS) or unavailable (DoE), totally or partially. For each building realisation and each considered structure-specific damage  
25 state, a fragility curve has been derived using 150 unscaled real ground motions and independently adopting SLaMA, pushover and time-history analysis. A GP regression has been fitted for each resulting parameter defining lognormal fragility relationships for the four damage states.

Findings from this study can be summarised as follows:

- GP regressions offer a flexible, non-parametric approach that does not rely on simplified assumptions for the  
30 quantification of the uncertainties, i.e. the modelling of the fitting error is explicitly dependent on the value of the input variables. Moreover, GP regressions do not require any a priori definition for the functional form of the fitting.

This is particularly desirable if class-specific parameters have to be considered or a high number of input variables is involved. In fact, in such cases it can be non-trivial to a priori define a functional form accurately capturing the mechanics of the phenomenon under investigation;

- For both the DoE and MCS approaches, the fitted GP regressions have a high predictive power in surrogating the computed results. Indeed, the RMSE is fairly low for all the investigated parameters, i.e. median and logarithmic standard deviation of the fragilities. Due to dynamic instability, which increases the variability of the training dataset (especially for the fragilities of more severe damage states), the RMSE of the time-history analysis-based GP regressions is approximately two to three times greater than for the two non-linear static methods. Moreover, the calculated 95% coverage probability falls in the range (92.6%-100%) for all the considered GP regressions, which is reasonably close to the theoretical value of 95%;
- For a given range of the input variables, both adopted sampling techniques (DoE and MCS) seemed leading to a similar computational burden. It is fair to suggest that choosing between DoE and MCS can be based on the (evolving) availability of the input data. In situations for which such availability is limited, a DoE sampling technique would be preferable, for example based on (assumed) uniform distributions for the input data, based on engineering judgment. If limited information is available for the building parameters (e.g. mean and standard deviation only), this can be used for a more-refined DoE. If the distributions of the building parameters are available, the MCS sampling technique based on such distributions could be more appropriate. In this way, the “resolution” of the GP regression can be higher in regions of the input space where more sampling will be needed when using the GP for portfolio risk assessment applications. Moreover, if the knowledge about the input data increases with time, the MCS-sampling technique allows enlarging the training dataset by drawing more samples according to the new information;
- At least for the analysed building classes, it is demonstrated that non-linear static methods are promising alternatives to non-linear time-history analysis to generate training datasets for the GP regressions. Coupling the GP with simplified response analysis methods is deemed necessary for practical portfolio applications, where a high number of building classes are generally involved (thus rendering hard the use of time-history analysis). The fragility-related errors of the simplified methods (with respect to time-history analysis) are within acceptable limits for portfolio-scale analysis, especially considering the mechanics-based approach and the explicit consideration of building-to-building variability. The SLaMA-based fragilities are generally slightly conservative: the mean error ratio (SLaMA-to-time history) for the median and logarithmic standard deviation of the fragilities ranges between 0.85 and 0.98, with a standard deviation within 0.03 and 0.27. On the other hand, the mean error ratio for the pushover is in the range (0.86, 1.08), with standard deviation in the range (0.03, 0.22);
- A much wider validation is required to generalise such results, and therefore demonstrate that training the GP regressions based on non-linear static based methods is feasible in the practise. In particular, a systematic validation of the CSM method using natural ground motions is needed. Moreover, the bias and dispersion of the SLaMA- or

pushover-based fragility parameters (with respect to non-linear time-history analysis) should be estimated for different building classes, possibly defining correction factors. In fact, SLaMA (or pushover analysis) can be used to calibrate GP regressions, while correcting the results accounting for such bias/dispersion. This could effectively reduce the required computational burden while providing accurate predictions.

## 5 Acknowledgements

This study was performed in the framework of the “*INSPIRE: Indonesia School Programme to Increase Resilience*” and “*i-RESIST: Increasing Resilience of Schools in Indonesia to earthquake Shaking and Tsunami*” projects, funded by the British Council through the Newton Institutional Links scheme and Research England through the UCL Global Challenges Research Fund (GCRF) Small Research Grants scheme. Additional funding was received from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 843794. (Marie Skłodowska-Curie Research Grants Scheme MSCA-IF-2018: MULTIRES, MULTI-level framework to enhance seismic RESilience of RC buildings).

## References

- [1] Cornell CA, Jalayer F, Hamburger RO, Foutch DA. Probabilistic basis for 2000 SAC federal emergency management agency steel moment frame guidelines. *J Struct Eng* 2002. doi:10.1061/(ASCE)0733-9445(2002)128:4(526).
- 15 [2] Federal Emergency Management Agency. *Seismic Performance Assessment of Buildings. Volume 1 - Methodology*. Washington, DC: 2012.
- [3] Gokkaya BU, Baker JW, Deierlein GG. Quantifying the impacts of modeling uncertainties on the seismic drift demands and collapse risk of buildings with implications on seismic design checks. *Earthq Eng Struct Dyn* 2016;45:1661–1683. doi:10.1002/eqe.2740.
- 20 [4] Liel AB, Haselton CB, Deierlein GG, Baker JW. Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. *Struct Saf* 2009;31:197–211. doi:10.1016/j.strusafe.2008.06.002.
- [5] Iervolino I, Manfredi G, Polese M, Verderame GM, Fabbrocino G. Seismic risk of R.C. building classes. *Eng Struct* 2007;29:813–20. doi:10.1016/j.engstruct.2006.06.019.
- 25 [6] Silva V. Uncertainty and correlation in seismic vulnerability functions of building classes. *Earthq Spectra* 2019;35:1515–39. doi:10.1193/013018EQS031M.
- [7] De Risi R, Jalayer • F, De Paola • F, Iervolino • I, Giugni • M, Topa • M E, et al. Flood risk assessment for informal settlements. *Nat Hazards* 2013;69:1003–32. doi:10.1007/s11069-013-0749-0.
- [8] Miano A, Jalayer F, De Risi R, Prota A, Manfredi G. Model updating and seismic loss assessment for a portfolio of bridges. *Bull Earthq Eng* 2016. doi:10.1007/s10518-015-9850-y.
- 30 [9] Seeger M. Gaussian processes for machine learning. *Int J Neural Syst* 2004;14:69–106.

doi:10.1142/S0129065704001899.

- [10] Gentile R, Galasso C. A multi-level seismic fragility assessment framework for existing RC buildings. 17th World Conf. Earthq. Eng., Sendai, Japan: 2020, p. C001374, submitted.
- [11] New Zealand Society for Earthquake Engineering (NZSEE). The seismic assessment of existing buildings - technical guidelines for engineering assessments. Wellington, New Zealand: 2017.
- 5 [12] Gentile R, Del Vecchio C, Pampanin S, Raffaele D, Uva G. Refinement and Validation of the Simple Lateral Mechanism Analysis (SLaMA) Procedure for RC Frames. *J Earthq Eng* 2019.
- [13] Gentile R, Pampanin S, Raffaele D, Uva G. Non-linear analysis of RC masonry-infilled frames using the SLaMA method: part 1—mechanical interpretation of the infill/frame interaction and formulation of the procedure. *Bull Earthq Eng* 2019;17:3283–304. doi:10.1007/s10518-019-00580-w.
- 10 [14] Gentile R, Pampanin S, Raffaele D, Uva G. Non-linear analysis of RC masonry-infilled frames using the SLaMA method: part 2—parametric analysis and validation of the procedure. *Bull Earthq Eng* 2019;17:3305–26. doi:10.1007/s10518-019-00584-6.
- [15] Gentile R, Pampanin S, Raffaele D, Uva G. Analytical seismic assessment of RC dual wall/frame systems using SLaMA: Proposal and validation. *Eng Struct* 2019;188:493–505. doi:10.1016/j.engstruct.2019.03.029.
- 15 [16] Freeman SA. Development and use of capacity spectrum method. 6th U.S. Natl. Conf. Earthq. Engineering, Seattle: 1998.
- [17] Box GEP, Wilson KB. On the Experimental Attainment of Optimum Conditions. *J R Stat Soc Ser B* 1951;13:1–38. doi:10.1111/j.2517-6161.1951.tb00067.x.
- 20 [18] Gentile R, Porco F, Raffaele D, Uva G. Simplified Moment-Curvature relationship in analytical form for RC circular sections. *Bull New Zeal Soc Earthq Eng* 2018;51:145–58.
- [19] Faravelli L. Response-surface approach for reliability analysis. *J Eng Mech* 1989. doi:10.1061/(ASCE)0733-9399(1989)115:12(2763).
- [20] Bucher CG, Bourgund U. A fast and efficient response surface approach for structural reliability problems. *Struct Saf* 25 1990. doi:10.1016/0167-4730(90)90012-E.
- [21] Rabitz H, Aliş ÖF, Shorter J, Shim K. Efficient input-output model representations. *Comput Phys Commun* 1999;117:11–20. doi:10.1016/S0010-4655(98)00152-0.
- [22] Sahu D, Nishanth M, Dhir PK, Sarkar P, Davis R, Mangalathu S. Stochastic response of reinforced concrete buildings using high dimensional model representation. *Eng Struct* 2019;179:412–22. doi:10.1016/j.engstruct.2018.10.083.
- 30 [23] Sevieri G, Andreini M, De Falco A, Matthies HG. Concrete gravity dams model parameters updating using static measurements. *Eng Struct* 2019. doi:10.1016/j.engstruct.2019.05.072.
- [24] Friedman JH. Rejoinder: Multivariate Adaptive Regression Splines. *Ann Stat* 2007;19:123–41. doi:10.1214/aos/1176347973.

- [25] Hardy RL. Multiquadric equations of topography and other irregular surfaces. *J Geophys Res* 1971;76:1905–15. doi:10.1029/jb076i008p01905.
- [26] Mangalathu S, Heo G, Jeon JS. Artificial neural network based multi-dimensional fragility development of skewed concrete bridge classes. *Eng Struct* 2018;162:166–76. doi:10.1016/j.engstruct.2018.01.053.
- 5 [27] Vapnik VN. *Statistical learning theory*. Wiley; 1998.
- [28] Ghosh J, Padgett JE, Dueñas-Osorio L. Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges. *Probabilistic Eng Mech* 2013;34:189–99. doi:10.1016/j.probengmech.2013.09.003.
- [29] Minas S, Chandler RE, Rossetto T. BEA: An efficient Bayesian emulation-based approach for probabilistic seismic response. *Struct Saf* 2018;74:32–48. doi:10.1016/j.strusafe.2018.04.002.
- 10 [30] Carr AJ. RUAUMOKO2D - The Maori God of Volcanoes and Earthquakes. Inelastic Analysis Finite Element program. Christchurch, New Zealand: 2016.
- [31] Magenes G, Pampanin S. Seismic response of gravity-load design frames with masonry infills. 13th World Conf. Earthq. Eng., Vancouver, B.C., Canada: 2004.
- 15 [32] Priestley MJN, Seible F, Calvi GM. *Seismic design and retrofit of bridges*. New York, USA: John Wiley and Sons; 1996.
- [33] Kowalsky MJ, Priestley MJN. Improved analytical model for shear strength of circular reinforced concrete columns in seismic regions. *ACI Struct J* 2000;97:388–96.
- [34] Berry MP, Eberhard MO. Practical Performance Model for Bar Buckling. *J Struct Eng* 2005;131:1060–70.
- 20 [35] Saiidi M, Sozen M. *Simple and complex models for nonlinear seismic response of reinforced concrete structures*. Urbana, Illinois, USA: 1979.
- [36] Sharpe RD. *The Seismic Response of Inelastic Structures*. Department of Civil Engineering, University of Canterbury, Christchurch, New Zealand., 1976.
- [37] Montejo LA, Kowalsky MJ. Set of codes for the analysis of reinforced concrete members. Raleigh, North Carolina.: 25 2007.
- [38] Gentile R. RC Section Rect - practical manual. Free internal report and software ([www.robortogentile.org](http://www.robortogentile.org), last access: 30/07/2019). 2017.
- [39] Jalayer F, Cornell CA. Alternative non-linear demand estimation methods for probability-based seismic assessments. *Earthq Eng Struct Dyn* 2009;38:951–72. doi:10.1002/eqe.876.
- 30 [40] Smerzini C, Galasso C, Iervolino I, Paolucci R. Ground motion record selection based on broadband spectral compatibility. *Earthq Spectra* 2014;30:1427–48. doi:10.1193/052312EQS197M.
- [41] Rossetto T, Gehl P, Minas S, Galasso C, Duffour P, Douglas J, et al. FRACAS: A capacity spectrum approach for seismic fragility assessment including record-to-record variability. *Eng Struct* 2016;125:337–48.

doi:10.1016/j.engstruct.2016.06.043.

- [42] Jalayer F, De Risi R, Manfredi G. Bayesian Cloud Analysis: efficient structural fragility assessment using linear regression. *Bull Earthq Eng* 2015;13:1183–203.
- [43] Freeman SA. Review of the Development of the Capacity Spectrum Method. *ISET J Earthq Technol* 2004;41:1–13.
- 5 [44] Priestley MJN, Calvi GM, Kowalsky MJ. Direct displacement-based seismic design of structures. Pavia, Italy: IUSS Press; 2007.
- [45] Kohrangi M, Vamvatsikos D, Bazzurro P. Site dependence and record selection schemes for building fragility and regional loss assessment. *Earthq Eng Struct Dyn* 2017. doi:10.1002/eqe.2873.
- [46] Minas S, Galasso C. Accounting for spectral shape in simplified fragility analysis of case-study reinforced concrete frames. *Soil Dyn Earthq Eng* 2019;119:91–103. doi:10.1016/j.soildyn.2018.12.025.
- 10 [47] Jalayer F, Ebrahimian H, Miano A, Manfredi G, Sezen H. Analytical fragility assessment using unscaled ground motion records. *Earthq Eng Struct Dyn* 2017;46:2639–63. doi:10.1002/eqe.2922.
- [48] Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. the MIT Press; 2006.
- [49] Nocedal J, Wright SJ. *NUMERICAL OPTIMIZATION*. New York, NY: Springer; 2006.
- 15 [50] MATLAB. version 9.5.0.944444 (R2018b). Natick, Massachusetts: The MathWorks Inc.; 2018.
- [51] Nassirpour A, Galasso C, Ayala DD. Multi-Hazard Physical Vulnerability Prioritization of School Infrastructure in the Philippines. 11th U.S. Natl. Conf. Earthq. Eng., Los Angeles, CA, USA: 2018.
- [52] Saputra A. Safety Performance of Concrete Structures in Indonesia. *Procedia Eng* 2017;171:985–93. doi:10.1016/j.proeng.2017.01.407.
- 20 [53] Nowak AS, Nowak S, Szerszen MM. Calibration of design code for buildings (ACI 318): Part 1 - Statistical models for resistance. *ACI Struct J* 2003;100:377–82. doi:10.14359/12613.
- [54] Galasso C, Maddaloni G, Cosenza E. Uncertainly Analysis of Flexural Overstrength for Capacity Design of RC Beams. *J Struct Eng* 2014;140.
- [55] International Conference of Buildings Officials (ICBO). *Uniform Building Code*. Whittier, California, USA: 1997.
- 25 [56] American Society of Civil Engineers (ASCE). *Minimum Design Loads for Buildings and Other Structures*. Reston, Virginia, USA: 2010. doi:10.1061/9780784400920.
- [57] Kircher CA, Whitman R V., Holmes WT. HAZUS Earthquake Loss Estimation Methods. *Nat Hazards Rev* 2006;7:45–59. doi:10.1061/(asce)1527-6988(2006)7:2(45).
- [58] Sevieri G, Gentile R, Galasso C. A bayesian framework for robust seismic fragility assessment based on various model classes. 17th World Conf. Earthq. Eng., Sendai, Japan: 2020, p. C001926, submitted.
- 30 [59] Alvarez M, Lawrence ND. Sparse convolved Gaussian processes for multi-output regression. *Adv. Neural Inf. Process. Syst.* 21 - Proc. 2008 Conf., 2009.



## Appendix

**Table A 1 Predictive power of the DoE-based GPs. cvg: 95% coverage.**

GP	Pre-Code		Low-Code		GP	Pre-Code		Low-Code		GP	Pre-Code		Low-Code	
	RMSE	cvg	RMSE	cvg		RMSE	cvg	RMSE	cvg		RMSE	cvg	RMSE	cvg
$\mu_1^S$	0.000	1.000	0.001	1.000	$\mu_1^P$	0.001	1.000	0.000	1.000	$\mu_1^T$	0.001	0.967	0.003	0.967
$\mu_2^S$	0.001	1.000	0.002	1.000	$\mu_2^P$	0.003	0.933	0.002	0.967	$\mu_2^T$	0.008	0.967	0.021	0.933
$\mu_3^S$	0.006	0.967	0.008	1.000	$\mu_3^P$	0.012	0.933	0.012	0.967	$\mu_3^T$	0.037	0.933	0.184	0.933
$\mu_4^S$	0.008	0.933	0.013	1.000	$\mu_4^P$	0.015	0.967	0.024	0.967	$\mu_4^T$	0.071	0.800	0.101	0.600
$\beta_1^S$	0.007	0.967	0.002	1.000	$\beta_1^P$	0.005	1.000	0.004	1.000	$\beta_1^T$	0.009	0.967	0.007	1.000
$\beta_2^S$	0.006	0.967	0.003	1.000	$\beta_2^P$	0.005	1.000	0.005	1.000	$\beta_2^T$	0.011	0.967	0.002	1.000
$\beta_3^S$	0.006	0.967	0.004	0.967	$\beta_3^P$	0.007	0.933	0.007	0.967	$\beta_3^T$	0.039	0.933	0.027	0.967
$\beta_4^S$	0.005	0.967	0.003	1.000	$\beta_4^P$	0.003	1.000	0.009	0.967	$\beta_4^T$	0.055	0.900	0.000	1.000



**Table A 2 Mapping of the predictive power of the MCS-based GPs (median for 100 sub-sets).**

Class	GP	Root Mean Squared Error						95% coverage					
		10	20	30	50	80	100*	10	20	30	50	80	100*
Pre-Code	$\mu_1^S$	0.001	0.001	0.001	0.001	0.001	0.001	0.926	1.000	0.981	0.963	0.963	1.000
	$\mu_2^S$	0.004	0.004	0.003	0.003	0.003	0.003	0.685	0.704	0.741	0.722	0.704	0.963
	$\mu_3^S$	0.010	0.009	0.008	0.008	0.008	0.001	0.759	0.889	0.889	0.926	0.926	1.000
	$\mu_4^S$	0.019	0.019	0.018	0.018	0.018	0.014	0.593	0.704	0.704	0.741	0.778	0.926
	$\beta_1^S$	0.013	0.012	0.012	0.012	0.012	0.010	0.593	0.630	0.704	0.778	0.741	0.963
	$\beta_2^S$	0.012	0.012	0.012	0.012	0.012	0.010	0.556	0.630	0.704	0.741	0.704	0.963
	$\beta_3^S$	0.010	0.010	0.010	0.010	0.010	0.007	0.667	0.704	0.741	0.796	0.778	0.963
	$\beta_4^S$	0.009	0.009	0.009	0.008	0.009	0.007	0.630	0.704	0.722	0.778	0.741	0.963
	$\mu_1^P$	0.002	0.002	0.001	0.001	0.001	0.001	0.815	0.852	0.889	0.889	0.926	0.963
	$\mu_2^P$	0.006	0.005	0.005	0.005	0.005	0.004	0.593	0.704	0.667	0.704	0.667	0.926
	$\mu_3^P$	0.021	0.020	0.020	0.020	0.021	0.014	0.593	0.704	0.685	0.704	0.704	0.963
	$\mu_4^P$	0.028	0.025	0.025	0.025	0.025	0.017	0.630	0.741	0.741	0.759	0.741	0.963
	$\beta_1^P$	0.012	0.012	0.012	0.012	0.011	0.009	0.667	0.722	0.741	0.778	0.852	0.963
	$\beta_2^P$	0.013	0.013	0.012	0.012	0.012	0.010	0.630	0.722	0.741	0.778	0.815	1.000
	$\beta_3^P$	0.012	0.011	0.011	0.011	0.011	0.010	0.630	0.778	0.778	0.778	0.852	0.963
	$\beta_4^P$	0.010	0.010	0.009	0.009	0.009	0.008	0.574	0.741	0.704	0.796	0.815	1.000
	$\mu_1^T$	0.002	0.002	0.002	0.001	0.001	0.001	0.852	0.889	0.870	0.852	0.852	0.926
	$\mu_2^T$	0.011	0.011	0.011	0.012	0.012	0.008	0.574	0.667	0.704	0.741	0.778	0.963
	$\mu_3^T$	0.053	0.057	0.055	0.052	0.049	0.042	0.519	0.704	0.741	0.815	0.852	0.963
	$\mu_4^T$	0.058	0.054	0.054	0.053	0.052	0.048	0.704	0.852	0.889	0.926	0.963	0.963
	$\beta_1^T$	0.014	0.014	0.014	0.014	0.013	0.012	0.722	0.741	0.796	0.852	0.889	0.963
	$\beta_2^T$	0.016	0.016	0.016	0.015	0.016	0.014	0.667	0.741	0.778	0.815	0.852	0.963
	$\beta_3^T$	0.052	0.050	0.050	0.048	0.048	0.042	0.481	0.630	0.741	0.815	0.852	0.926
	$\beta_4^T$	0.060	0.058	0.061	0.058	0.058	0.044	0.630	0.778	0.815	0.852	0.889	1.000
Low Code	$\mu_1^S$	0.002	0.001	0.001	0.001	0.001	0.001	0.889	0.963	0.963	0.963	0.963	0.926
	$\mu_2^S$	0.004	0.004	0.003	0.003	0.003	0.003	0.722	0.741	0.815	0.741	0.778	0.963
	$\mu_3^S$	0.016	0.015	0.015	0.015	0.015	0.011	0.593	0.741	0.704	0.741	0.778	0.926
	$\mu_4^S$	0.027	0.026	0.025	0.025	0.025	0.020	0.537	0.704	0.741	0.741	0.778	1.000
	$\beta_1^S$	0.010	0.009	0.008	0.008	0.008	0.006	0.593	0.704	0.704	0.778	0.796	1.000
	$\beta_2^S$	0.011	0.010	0.009	0.009	0.009	0.007	0.519	0.667	0.630	0.704	0.704	0.926
	$\beta_3^S$	0.010	0.009	0.009	0.008	0.008	0.007	0.574	0.741	0.704	0.741	0.741	1.000
	$\beta_4^S$	0.010	0.009	0.009	0.008	0.008	0.006	0.593	0.741	0.722	0.741	0.778	0.926
	$\mu_1^P$	0.001	0.001	0.001	0.001	0.001	0.001	0.963	0.963	0.944	0.963	0.963	0.963
	$\mu_2^P$	0.005	0.004	0.004	0.004	0.004	0.003	0.630	0.704	0.741	0.704	0.667	1.000
	$\mu_3^P$	0.022	0.021	0.020	0.021	0.020	0.015	0.630	0.704	0.741	0.778	0.852	0.926
	$\mu_4^P$	0.028	0.026	0.027	0.024	0.026	0.016	0.593	0.611	0.741	0.815	0.889	1.000
	$\beta_1^P$	0.009	0.008	0.008	0.008	0.008	0.007	0.704	0.815	0.815	0.889	0.889	1.000
	$\beta_2^P$	0.012	0.011	0.011	0.011	0.011	0.011	0.537	0.704	0.704	0.741	0.778	0.963
	$\beta_3^P$	0.011	0.010	0.010	0.010	0.011	0.009	0.630	0.778	0.815	0.833	0.889	0.926
	$\beta_4^P$	0.015	0.014	0.014	0.015	0.014	0.011	0.667	0.778	0.778	0.815	0.852	0.963
	$\mu_1^T$	0.004	0.004	0.004	0.004	0.004	0.003	0.704	0.704	0.741	0.778	0.741	0.926
	$\mu_2^T$	0.029	0.028	0.027	0.028	0.027	0.020	0.667	0.759	0.815	0.852	0.815	0.963
	$\mu_3^T$	0.061	0.064	0.061	0.057	0.056	0.048	0.556	0.722	0.704	0.815	0.852	0.926
	$\mu_4^T$	0.063	0.066	0.065	0.063	0.062	0.056	0.815	0.815	0.889	0.926	0.963	0.963
	$\beta_1^T$	0.051	0.047	0.044	0.046	0.046	0.033	0.685	0.704	0.741	0.759	0.778	0.963
	$\beta_2^T$	0.065	0.061	0.058	0.059	0.054	0.034	0.593	0.630	0.667	0.741	0.741	0.926
	$\beta_3^T$	0.045	0.042	0.044	0.042	0.042	0.034	0.667	0.778	0.796	0.889	0.926	1.000
	$\beta_4^T$	0.057	0.053	0.053	0.051	0.050	0.044	0.593	0.722	0.778	0.852	0.926	0.963

\* based on the full training dataset; no resampling.