2019 SECED Conference
Earthquake risk and engineering towards a resilient world
9-10 September 2019, Greenwich, London

GAUSSIAN PROCESS REGRESSION FOR SEISMIC FRAGILITY ASSESSMENT OF BUILDING PORTFOLIOS

Roberto GENTILE[1], Carmine GALASSO[2]

Abstract: *Seismic fragility assessment of building portfolios usually involves empirical approaches, or numerical, mechanics-based approaches applied to properly-sampled index buildings representative of defined structural typologies. These approaches often neglect the effect of building-to-building variability on portfolio seismic risk estimates. Alternatively, metamodeling techniques can be adopted to surrogate complex mechanical analyses and to properly include class variability. However, commonly-used metamodels rely on various simplifying assumptions. In this study, Gaussian process regression is adopted to address these limitations. The proposed method is demonstrated for seismically-deficient RC school buildings with construction details typical of some developing countries (e.g., in Southeast Asia), for which field data is available. Gaussian processes estimating the fragility statistics of such schools are fitted based on thousands non-linear time-history analyses for over 100 building realisations within the considered structural class. 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" through the Simple Lateral Mechanism Analysis (SLaMA) method. Four validation structures (outside the training set) are defined and analysed through the same approaches. Preliminary results from this study show predicted-to-"observed" errors below 10%, highlighting the accuracy of the fitted metamodels. Moreover, non-linear static approaches (SLaMA or numerical pushover), coupled with the capacity spectrum method, produce sound results, drastically reducing the computational burden in the model calibration.*

1. Introduction

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 ground-motion *Intensity Measure* (IM). Fragility relationships describe such conditional probability for increasing values of the ground-motion IM, taking the form of cumulative distribution functions (CDFs). These relationships are a key ingredient in any seismic risk assessment exercise, particularly in the case of building portfolios. Fragility relationships can be obtained empirically, using observed (post-event) damage data, through expert opinion, or simulating the structural response through computational models and analysis types of different complexity and computational demand.

If the number of structures of interest is particularly large (e.g., in the case of building portfolios), it may be unfeasible to run refined analyses (using refined models) for each individual structure within the portfolio. Therefore, various building classes are often defined in terms of few parameters (e.g., material/lateral-load resisting systems, height, age of construction), assigning a specified fragility relationship to each class. Such relationships, usually derived for an "average" archetype building of the class (an "index" building), are adopted for all the buildings in the class. To include class (i.e., building-to-building) variability in practical seismic risk assessment applications, surrogate metamodels may be adopted for each building class. A metamodel is a model of a model: it defines a relationship between a given set of inputs and outputs, obtained by analysing a relatively-small set of samples (e.g., structures in the same class) and subsequently fitting a function for the outputs that will replace (or surrogate) the real model.

This allows one to have a computationally-efficient tool to capture the complex and implicit relationship between model input(s) and outputs, such as seismic fragility parameters as a function of structural geometry, materials, detailing, etc. At the same time, metamodels allow to capture both structure-specific uncertainties (materials, geometry/detailing, modelling) and class-specific uncertainties.

[1] Research Fellow, Institute for Risk and Disaster Reduction, University College London, London, UK, r.gentile@ucl.ac.uk
[2] Associate Professor, Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

These are especially important when assessing individual structures, for instance through the Performance-Based Earthquake Engineering (PBEE) framework. However, commonly-adopted metamodels suffer for some drawbacks, such as: 1) the user should specify the functional form for the fitting, potentially generating substantial misfit or, conversely, physically-unsound overfit; 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).

To address the above issues, it is proposed here to adopt Gaussian Process (GP) regressions (Rasmussen & Willams, 2006) to develop a flexible, fast and accurate metamodel for the seismic fragility parameters of building classes. 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 field data is available. To this aim, a *Design of Experiment* (DoE) is defined considering (>1000) combinations of selected geometrical and mechanical properties of the case-study portfolio. For each combination, a fragility curve is defined based on non-linear time-history analyses for 150 unscaled real (i.e., recorded during past events) ground motions. Several GPs are fitted to the numerical results of the structural analyses. Four extra buildings (not considered in the training of the model) are also numerically analysed.

It is worth noting that significant computational effort may still be needed to calibrate such metamodel for a single structural class, i.e., thousands of time-history analyses are needed. To increase the tractability of the approach, two simplified but accurate analysis methods are also adopted as alternatives. In particular, the same metamodeling methodology is applied deriving force-displacement curves through the analytical approach Simple Lateral Mechanism Analysis (SLaMA, NZSEE 2017; Gentile et al., 2019a) or numerical pushover. The Capacity Spectrum Method (CSM, Freeman, 1998), adopting the same set of recorded ground motions, is applied using such curves. Therefore, two alternative sets of fragility functions are derived, together with the corresponding GP metamodels. Although a systematic analysis of the bias and dispersion in the fragility parameters due to the simplified methods is outside the scope of this paper, the preliminary results from this study indicate that such an approach might be feasible in practice.

2. Gaussian Process Regression

GPs overcome all the downsides of the commonly-used surrogate modelling techniques and, for this reason, they are selected in this paper. A GP is a non-parametric statistical method that finds a multivariate Gaussian distribution over the possible functions that can fit a set of observed data. The mathematical formulation of a GP regression is briefly described in Section 2.1.

2.1. Overview of Gaussian Process regression

Given a set of training data, a GP defines a multivariate Gaussian distribution over all the functions $f(\mathbf{x})$ that fit such data. In other words, the GP outputs for any arbitrary set of inputs defines a joint normal distribution. The mathematical form of a GP is described in Eq. 1). $\mathbf{x}$ and $\mathbf{x}'$ are two different input vectors (of arbitrary dimension), $m(\mathbf{x})$ is a mean function, Eq. 2), defined as the expected value of the GP, and $k(\mathbf{x},\mathbf{x}')$ is a covariance function (or kernel), Eq. 3), capturing the correlation among different input sets reflecting it in the output.

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}),k(\mathbf{x},\mathbf{x}')) \tag{1}$$

$$m(\mathbf{x}) = \mathbb{E}\left[f(\mathbf{x})\right] \tag{2}$$

$$k(\mathbf{x},\mathbf{x}') = \mathbb{E}\left[\big(f(\mathbf{x})-m(\mathbf{x})\big)\big(f(\mathbf{x}')-m(\mathbf{x}')\big)\right] \tag{3}$$

The starting point of the fitting procedure is the definition of a prior distribution, which will be converted into a posterior based on the observed data (in a Bayesian approach). Generally, a zero mean function is assigned for the prior, since this has a negligible influence on the posterior. The topology of the output functions - with particular reference to its smoothness - is governed by the covariance function, which specifies the covariance between pairs of random variables. The nature of the covariance function is the main input from the user and should reflect the expected behaviour of the output. A common choice in engineering applications is the squared exponential covariance, Eq. 4), which reflects the “stability” of the involved physical quantities (i.e., a small perturbation of the input produces small changes in the output). In such equation, $(x_i - x'_i)$ is the distance between two input sets, σ_i^2

represent the length scale of the output in each input dimension. σ_f^2 is the signal variance, which can be seen as the variance of single observations (i.e. the covariance function is equal to σ_f^2 when the distance between $\mathbf{x}$ and $\mathbf{x}'$ is zero). The parameters of the covariance are called hyperparameters since those are not specified by the user but learnt from the data (GPs are non-parametric models).

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

The posterior incorporates in the GP the knowledge embedded in the training data. In particular, this is obtained by conditioning the joint prior distribution, Eq. 5, to the observed data. Equations 6,7 and 8 describe this by assuming a vector of (known) training outputs **y**, for which each component is related to a vector of inputs. The column vector inputs are collected in the **X** matrix. The pairs ($\mathbf{f}_*, \mathbf{X}_*$) are related to the (unknown) test data. σ_n^2 is the noise variance, which is a hyperparameter that reflects the uncertain nature of the training pairs.

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_n^2\mathbf{I} & \mathbf{K}(\mathbf{X},\mathbf{X}_*) \\ \mathbf{K}(\mathbf{X}_*,\mathbf{X}) & \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix}\right) \tag{5}$$

$$\mathbf{f}_* \mid \mathbf{X},\mathbf{y},\mathbf{X}_* \sim \mathcal{N}(\overline{\mathbf{f}}_*, \mathrm{cov}(\mathbf{f}_*)) \tag{6}$$

$$\mathbf{f}_* = \mathbf{K}(\mathbf{X}_*,\mathbf{X})[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_n^2\mathbf{I}]^{-1}\mathbf{y} \tag{7}$$

$$\mathrm{cov}(\mathbf{f}_*) = \mathbf{K}(\mathbf{X}_*,\mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*,\mathbf{X})[\mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_n^2\mathbf{I}]\mathbf{K}(\mathbf{X},\mathbf{X}_*) \tag{8}$$

The last step of the procedure is the determination of the hyperparameters. This is done by maximising the likelihood, $p(\mathbf{y} \mid \mathbf{X}, \theta)$, which is the probability of predicting the training data **y**, given the training input set **X** and a set of hyperparameters **θ**. This is carried out adopting numerical optimisation algorithms often available by default in many programming languages.

3. Methodology

This study involves the derivation of seismic fragility curves for a class of seismically-deficient RC school buildings representative of the construction practice in some developing countries (Section 4). The main objective of the work is to develop a set of GP surrogate metamodels to predict the median and logarithmic standard deviation of the fragility functions defined for four different damage states. The desired surrogate metamodel is depicted in Figure 1 while the necessary steps to obtain it are described in the following:

1. Design of Experiment. An experimental design matrix is defined considering combinations of selected geometrical and mechanical properties of the school buildings. A three-factorial DoE has been selected for this study. The adopted values are shown in Section 4, Table 2. As a result, a group of 196 building realisations are defined;

2. Analysis. For each entry of the DoE, seismic fragility functions are estimated using different analysis techniques with increasing refinement. The first step of this process is to obtain a cloud of points in the Engineering Demand Parameter (EDP) vs IM space. For this study, the maximum inter-storey drift has been selected as the EDP while the IM is defined as the geometric mean (*Avg SA*) of the pseudo-spectral acceleration in the interval [$0.2T_{1,min}$, $2T_{1max}$], where $T_{1,min}$, and T_{1max} are respectively the minimum and maximum elastic period for the buildings in the DoE. 150 unscaled natural ground motions are selected from the SIMBAD database (Selected Input Motions for displacement-Based Assessment and Design, Smerzini et al., 2014). As in Rossetto et al., 2016, these records are 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 (arbitrarily) keeping the component with the largest PGA value.

The EDP values to be used in the cloud analysis are computed in three different ways. Firstly, SLaMA is adopted to derive analytically a force-displacement capacity curve and the expected plastic mechanism for the considered structure. Then, the CSM, is adopted to calculate the inter-storey drift demand for each of the 150 ground motions. A refined model has been defined using the FEM software Ruaumoko (Carr, 2016) to derive numerically the force-displacement curve. Such model is capable of predicting the flexural, bar slip and shear failure of RC beams and columns, together with shear failure in the beam-column joints and strength degradation. The CSM is applied to calculate the maximum inter-storey drift for the selected 150 ground motions. Finally, the most refined considered method is the full non-linear time-history analysis, conducted adopting the above-mentioned numerical model (including stiffness and strength degradation) and ground motions. For each method, the building realisations are analysed by means of two different two-dimensional analyses in the transverse and longitudinal directions.

The linear least square method is applied on the derived (EDP, IM) pairs, estimating the conditional mean and standard deviation of EDP given IM and deriving the commonly-used power-law model $EDP = aIM^b$, where *a* and *b* are the parameters of the regression. The derived probabilistic seismic demand model is used to define the median (μ) of four lognormal fragility curves, one for each DS, and the corresponding logarithmic standard deviation β (which is the same the four curves, due to the homoskedasticity assumptions in the cloud approach).

3. GP fitting. For each of the calculated output parameters (medians and standard deviation for each DS for the three considered analysis methods), a surrogate metamodel is fitted. This is done adopting a constant mean function, a squared exponential covariance function with different length scale in each input dimension and zero noise (different assumptions will be tested in future studies by the authors). This last setting reflects the deterministic nature of any numerical analysis (i.e., if the analysis is repeated, the output value does not change). Moreover, additional (less complex) Gaussian processes are fitted to surrogate the non-linear static force-displacement curve. In particular, the displacement and base shear at yielding and ultimate (both SLaMA- and numerically-based) are treated as output parameters.

Finally, four additional buildings (not considered in the DoE, and outside the training set of the GPs) have been analysed with the same above-mentioned procedure. The results are compared with the predictions of the surrogate metamodels.

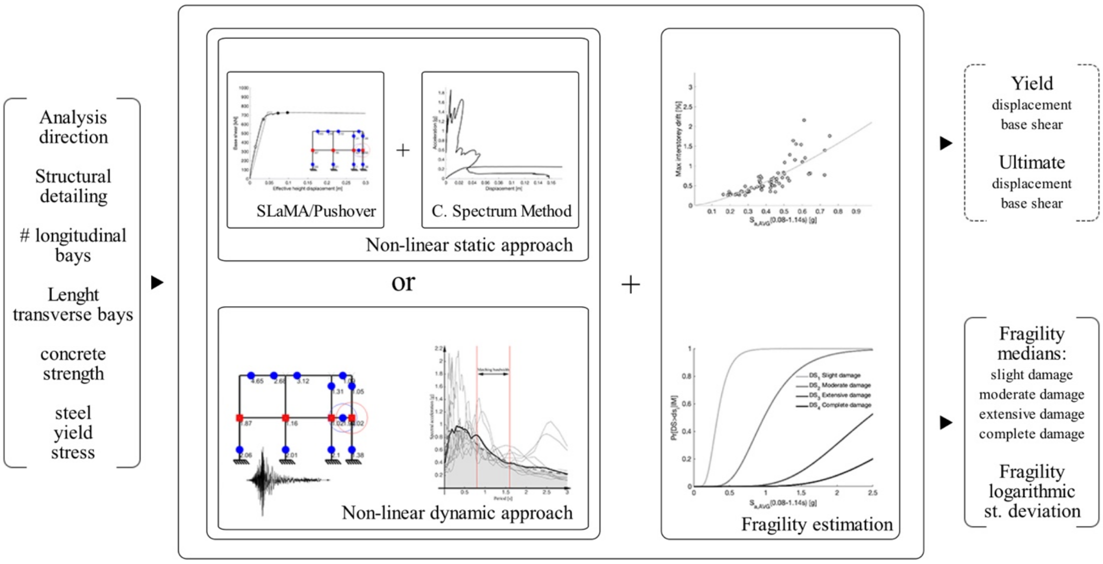

Figure 1: Description of the surrogate model approach.

4. Illustrative application

4.1. Description of the case study and Design of Experiment

The case-study building class selected for this study represents seismically-deficient RC school buildings, defined based on large data collection exercises (e.g., Nassirpour et al., 2018; Gentile et al., 2019b) involving Rapid Visual Surveys for over 200 school buildings carried out to collect administrative, geometric and mechanical data related to the investigated buildings.

The analysis of the collected data allows to firstly define building archetypes based on the most common lateral load resisting system, average geometrical and mechanical features and structural details. Then, the distributions of the collected data are used to define the DoE, selecting the most relevant varying parameters and their samples. The archetype building (Figure 2) is a two-storey rectangular-plan frame building and it represents approximately 80% of the surveyed schools. According to the collected geometrical data, the parameters with a significant variation are the number of longitudinal bays and the length of the transverse bays, while other parameters such as the number of storeys, the length of the corridor bay, the dimension of the beams/columns have negligible variability within the survey sample.

The knowledge on the material properties is comparatively lower than for geometry, given the visual nature of the data collection. Indeed, average values related to Indonesian statistics (Saputra, 2017) for the concrete cylindrical strength and steel yield stress have been used (24MPa and 400MPa, respectively). Moreover, Coefficients of Variation (CoV) respectively equal to 18% and 5% are selected based on literature studies (Nowak & Szerszen, 2003; Galasso et al., 2014). It is worth mentioning that both the analytical and 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.

Structural detailing is the last major variable in the DoE. Although measured data is not available, two different simulated design approaches are conducted to reflect two different nominal seismic performances. To this aim, the buildings are simulated designed according to the Uniform Building Code (1997) and the American Society of Civil Engineers (ASCE) 7-10. In fact, building codes in the developing countries typically refer (and are fully consistent with) to the UBC and/or the United States codes. However, some of the provisions in such codes have not been fully applied (e.g. stirrups in the joints) in the simulated design to somehow take into account the potential lack of code enforcement (observed during the field survey). The resulting detailing for the two categories (Table 1) leads to “Pre-Code” and “Low-Code” configurations, as defined in HAZUS MH4 (HAZard United States; Kircher et al., 2006).

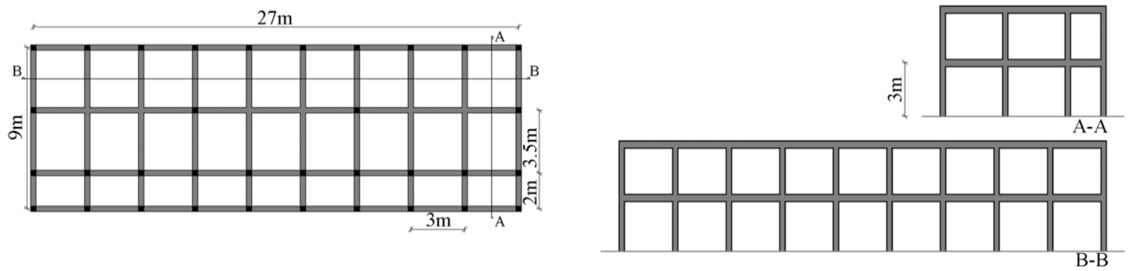

Figure 2: Geometry of the archetype building.

According to the above considerations, the selected parameters for the DoE are structural detailing, the number of longitudinal bays, the length of the transverse bays, concrete cylindrical strength and steel yield stress. The sampling values for the three-factorial design are shown in Table 2. Clearly, the number of longitudinal bays have no influence on the analysis in the transverse direction (the analogous is valid for the length of the transverse bays). Therefore, 54 models are sampled for each building direction, leading to a total of 108 models x 150 ground motions = 16200 analyses for each of the three considered analysis techniques.

	Typical beam	**Typical column**	**Typical joint**
Pre-Code	3ϕ16 top 3ϕ16 bottom ϕ10@150mm stirrups	3ϕ16 top 3ϕ16 bottom ϕ10@200mm stirrups	No stirrups
Low-Code	3ϕ16 top 3ϕ16 bottom ϕ10@150mm stirrups	3ϕ16 top 3ϕ16 bottom ϕ10@100mm stirrups	No stirrups

Table 1: Structural details for the as-built configuration.

4.2. Results and Discussion

As explained in Section 3, for each realisation in the DoE, three increasing-refinement analysis methods are adopted. This is done to quantify the bias of the simplified methods (with respect to the full non-linear time-history analyses) in the estimation of the fragilities (median and dispersion). Although a surrogate metamodel based on time-history analyses can substantially reduce the computational effort needed for regional scale seismic fragility applications, its calibration for a single class of buildings still requires thousands of time-history analyses. Therefore, simplified methods may be needed to render feasible such a metamodeling approach for a high number of building classes.

Variable	Samples		
Structural detailing	Pre-Code	Low-Code	
Number of longitudinal bays, $N_{bays,x}$ [-]	6	9	12
Length of transverse Bays, L_{by} [m]	2.92	3.71	4.50
Concrete cylindrical strength, f_c [MPa]	19.68	24.00	28.32
Steel yield stress, f_y [MPa]	380	400	420

Table 2: Design of experiment.

Figure 3a shows the longitudinal fragility analysis results for a Low-Code building in the DoE. 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. 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 3b, where the EDP vs IM cloud is shown, as predicted according to the CSM adopting the selected150 ground motions as an input. Indeed, the largest SLaMA vs Pushover discrepancies for maximum inter-storey drift are recorded for high values of the IM (geometric mean of the pseudo spectral acceleration), which forces high drift demand to the structure. Interestingly, very good match is shown, overall, among the non-linear static methods and the non-linear dynamic approach, considering the simplified nature of the formers. Figure 3c shows the estimated fragility curves for four different damage states (DS_i, Slight Damage, Moderate Damage, Extensive Damage, Complete Damage), which are based on drift limits equal to [0.25 0.6 1.5 2] %, which represent average values valid for the entire DoE. The highest error on the estimation of the fragility is registered for DS_4: respectively +7.3% and -8.7% on the median for SLaMA and pushover with respect to the time-history (+30.2% and 38.0% for the dispersion).

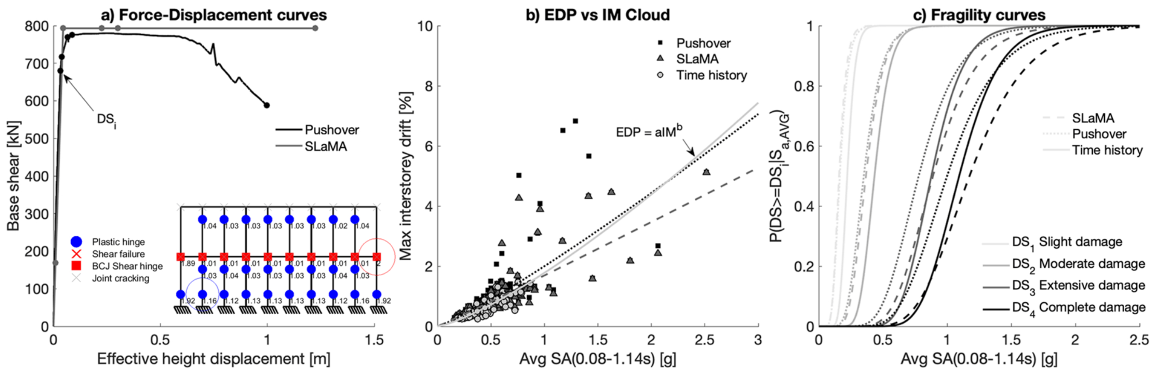

Figure 3: Longitudinal analysis results for the Low-Code, 9-bays, f_y-400MPa, f_c-19.7MPa building realisation.

Although the results presented in Figure 3 are promising and are qualitatively representative of the majority of the analysed buildings in the DoE, there are some other case studies in which the error trend increases among the non-linear static approaches and the refined time-history. Figure 4 shows one of such cases, which suggest that more investigations are needed before suggesting applying simplified non-linear static approaches for fragility analysis. In particular, such greater error in the fragility estimation (Figure 4c) could be ascribed to the application of the CSM using the spectra of recorded ground motions. Indeed, using a linear factor to downscale a natural elastic spectrum according to the expected ductility demand might not be appropriate since the resulting inelastic spectrum is not sufficiently smooth.

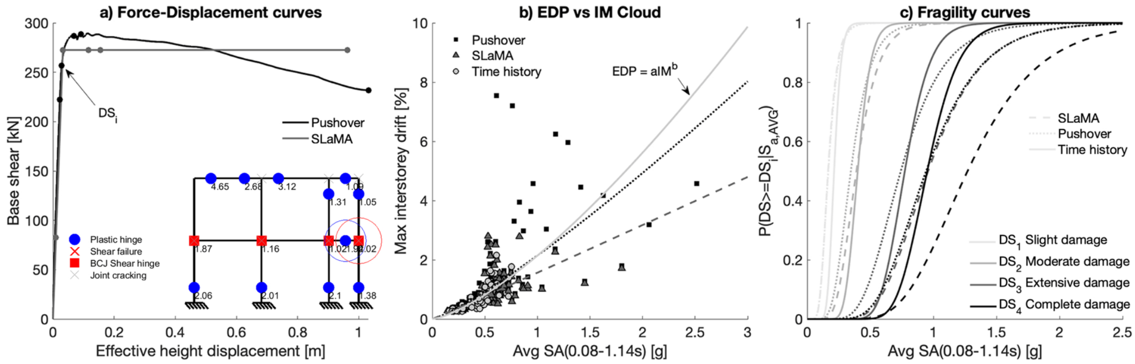

Figure 4: Transverse analysis results for the Pre-Code, L_{by}-4.5m, f_y-380MPa, f_c-24MPa building realisation.

The output from the analytical and numerical analyses for the DoE is adopted to fit a number of GPs. In particular, the median and the dispersion of the fragility curves are used ($\mu_1, \mu_2, \mu_3, \mu_4, \beta$). Moreover, each force-displacement curve is represented in bi-linear form and a GP is fitted for the displacement and base shear of the yielding and ultimate point (both for SLaMA and pushover, $\delta_y, \delta_u, V_{By}, V_{Bu}$). An additional subscript is added to such parameters to indicate the adopted analysis method (i.e. SL, PO and TH for SLaMA, pushover and time-history respectively). A single GP is fitted for each of these parameters, for each structural vulnerability class (Pre-/Low- code) and for each analysis direction, leading to 92 different GPs.

To demonstrate the achieved goodness of fit, Figure 5 shows the fitted GP related to the median of the DS2 fragility based on non-linear time-history ($\mu_{2,TH}$), related to the transverse analysis of the Pre-Code vulnerability class. The higher-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 length of the transverse bay (L_{by}), represented by each line in the panels. It is evident that, after the maximisation of the likelihood function, the mean function of the GP is capable of interpolating the observed data (negligible noise is considered for numerical purposes). Moreover, the shaded areas in the figure represent the 95% confidence bounds of the predictions. Those clearly indicate that the highest uncertainty is expected for predictions for intermediate values between two observed points, although the highest registered 95% confidence range is approximately equal to $\pm$5% with respect to the mean. The central panel of this figure also shows the kernel adopted hyperparameters (length scale in each input dimension, $\sigma_{L_{by}^2}, \sigma_{f_c}^2, \sigma_{f_y}^2$, and signal variance σ_f^2).

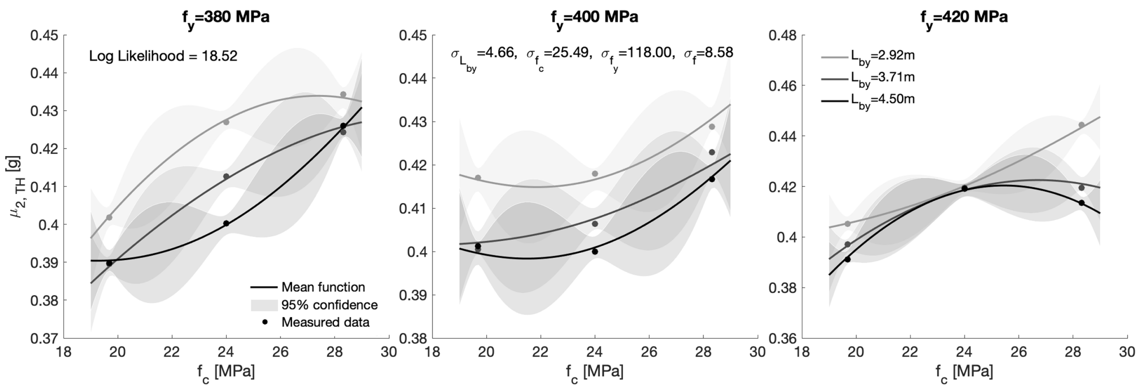

Figure 5: Pre-Code, transverse direction GP for the DS2 fragility median based on full time-history.

For a thorough validation of the entire set of calibrated GPs, four "test" buildings are defined according to geometrical and mechanical parameters within the ranges of the DoE (Table 3). Such buildings are analysed (SLaMA, pushover and time-history) to obtain the above-mentioned parameters that define both the force-displacement and the fragility curves ("observed" data). The same parameters are therefore predicted adopting the calibrated GPs, and a relative error is calculated in the form of "(predicted – observed) / observed".

	Details	$N_{bays,x}$	L_{by} [m]	f_c [MPa]	f_y [MPa]
Test 1	Pre-Code	9	3.5	21.9	416
Test 2	Pre-Code	6	4.0	27.6	390
Test 3	Low-Code	12	4.3	26	385
Test 4	Low-Code	6	3.7	23	419

Table 3: Building parameters for the test cases.

Figure 6 shows the relative error for all the parameters and all the test cases. Although the results are still preliminary, it is interesting to highlight that the error is considerably below 10% for most of the parameters, which highlights the excellent performance of the GPs in surrogating such a complex mechanical analysis. This is particularly valid for the force-displacement parameters (mean absolute error 1.4%, maximum 5.9%) and the fragility parameters based on the non-linear static methods (mean absolute error 1.7%, maximum 13.1%). On the other hand, the fragility parameters based on time-history analyses present a higher error trend (mean absolute error 4.2%, maximum 21.7%). Although more investigation is needed, this is likely to be due to the relatively "unstable" nature of the observed fragility (see also Figure 5). In turn, the preliminary results indicate that this might be due to the "random" P-delta related collapses in the time-history analyses. Therefore, it is deemed that a more refined fitting of the clouds to include such cases in the probabilistic seismic demand model can significantly improve the overall goodness of fit, along with the accuracy of the mechanics-based results.

It is worth mentioning that such results refer to the surrogated vs observed error, and therefore refer only on the ability of the GP to effectively surrogate complex mechanical analyses. They do not give any insight on the relative effectiveness of one analysis method with respect to the others. As mentioned above, this aspect requires more investigation before such a research question can be fully answered.

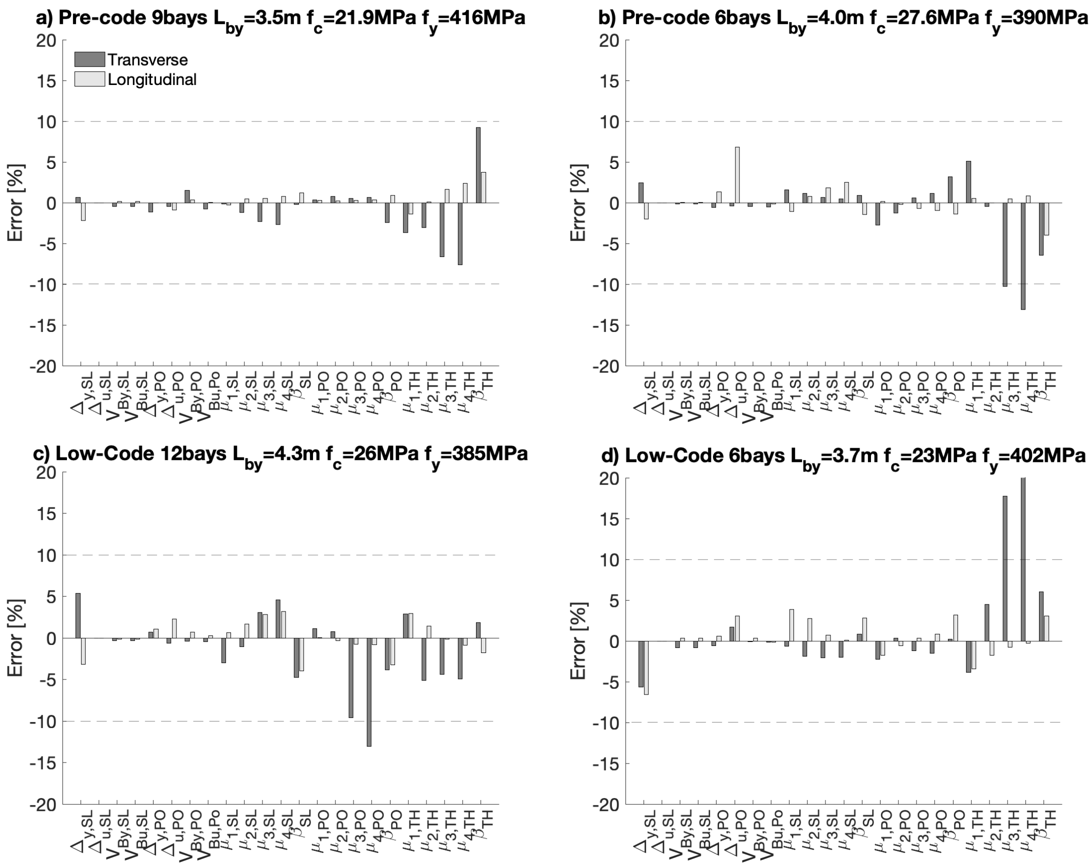

Figure 6: Predicted vs observed error for the four test buildings.

5. Conclusions

This study aims at developing a flexible, fast and accurate mechanics-based metamodel for seismic fragility curves of classes of structures. Such metamodel, based on GP regression, allows to account for class-variability in regional scale seismic fragility assessment applications, avoiding the need to rely on empirical methods nor using a single fragility curve for the entire structural class. Moreover, negligible computational power is needed to adopt the fitted metamodel. As opposed to commonly-used metamodeling techniques, the Gaussian process approach allows to have full probabilistic predictions, which renders it more appropriate for regional-scale risk applications.

Such method is demonstrated for seismic-deficient RC schools with construction details typical of developing countries (such as Philippines and Indonesia), for which real data is available. Gaussian processes to estimate the fragility curves of such schools are fitted on the basis of thousands of time-history analyses for over one hundred building realisations within the structural class. To further increase the tractability of the methodology, alternative metamodels are defined based on numerical pushover analyses or analytical "by hand" non-linear static analyses by means of the SLaMA method. Four "test" buildings (outside the calibration set) are defined and analysed by the same means. Predicted vs "observed" error define the accuracy of the fitted metamodels.

For the test sets, the predicted vs "observed" error is below 10% for the majority of the analysed parameters. However, a more refined fitting of the probabilistic seismic demand model is needed before confirming the observed error trends (which will be presumably reduced). Although a systematic analysis of the bias of the simplified methods is still needed, non-linear static approaches (SLaMA or numerical pushover), coupled with the CSM, seem to be capable of producing sound results, drastically reducing the computational demand. Also, the generic ground-motion set used in this study could be replaced by *ad-hoc*, site-specific (e.g., hazard-consistent) ground-motion sets.

More investigations are needed to refine these preliminary results. However, these seem to indicate that metamodels based on Gaussian process regression could be an appealing solution for the seismic fragility analysis of building portfolios. Moreover, the analytical method SLaMA could be combined to the metamodel, to effectively "control" the mechanics of the problem and avoid a black box effect. In other words, SLaMA could be easily applied to all the cases in the regional assessment, to predict the force-displacement curve and the expected plastic mechanism. Therefore, those could be used as additional inputs to the metamodel, adopted to predict the fragility curves.

References

American Society of Civil Engineers (ASCE), Minimum Design Loads for Buildings and Other Structures. ASCE/SEI Standard 7-10. Reston, Virginia, USA, 2010.

Carr AJ (2016), *RUAUMOKO2D* - The Maori God of Volcanoes and Earthquakes. Inelastic Analysis Finite Element program. Carr research ltd, Christchurch, New Zealand, 2016.

Freeman SA (1998), Development and use of capacity spectrum method. In: Proc. 6th U.S. National Conference on Earthquake Engineering, Seattle.

Galasso C, Maddaloni G, Cosenza E (2014), Uncertainly Analysis of Flexural Overstrength for Capacity Design of RC Beams. *Journal of Structural Engineering*, 140(7).

Gentile R, Del Vecchio C, Pampanin S, Raffaele D, Uva G (2019a), Refinement and validation of the Simple Lateral Mechanism Analysis (SLaMA) procedure for RC bare frames. *Journal of Earthquake Engineering*, DOI: 10.1080/13632469.2018.1560377, 2019a.

Gentile R, Galasso C, Idris Y, Rusydy I, Meilianda E (2019b), From rapid visual survey to multi-hazard risk prioritisation and numerical fragility of school buildings. Nat. Hazards Earth Syst. Sci. In press.

International Conference of Buildings Officials (ICBO), Uniform Building Code. Whittier, California, USA, 1997.

Kircher CA, Whitman RV , Holmes WT (2006), HAZUS Earthquake Loss Estimation Methods". *Natural Hazard Review*, 7: 45-59.

Nassirpour A, Galasso C, D'Ayala D (2018), Multi-Hazard Physical Vulnerability Prioritization of School Infrastructure in the Philippines. 11th U.S. National Conference on Earthquake Engineering (11NCEE). Los Angeles, CA, USA, June 25-29.

Nowak AS, Szerszen MM (2003), Calibration of design code for buildings (ACI 318). Part 1: Statistical models for resistance. ACI *Structural Journal*, 100(3): 377–382.

NZSEE, New Zealand Society for Earthquake Engineering (2017), The seismic assessment of existing buildings - technical guidelines for engineering assessments. Wellington, New Zealand.

Rasmussen CE and Williams CKI (2006), Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X.

Rossetto T, Gehl P, Minas S, Galasso C, Duffour C, Douglas J, Cook O (2016), FRACAS: A capacity spectrum approach for seismic fragility assessment including record-to-record variability. Engineering Structures 125: 337-348.

Saputra A (2017), Safety Performance of Concrete Structures in Indonesia. Procedia Engineering 171: 985-993.

Smerzini C, Galasso C, Iervolino I, Paolucci R (2014), Ground motion record selection based on broadband spectral compatibility Earthquake Spectra, 30 (4): 1427-1448.