ICOSSAR 2021
The 13th International Conference on Structural Safety and Reliability (ICOSSAR 2021), June 21-25, 2021, Shanghai, P.R. China
J. Li, Pol D. Spanos, J.B. Chen & Y.B. Peng (Eds)

Performance Assessment of Buildings Subjected to Earthquake-Induced Landslides through Gaussian Process Regression

Giacomo Miluccio[a], Roberto Gentile[b], Carmine Galasso[b], Fulvio Parisi[a]

[a]*Department of Structures for Engineering and Architecture, University of Naples Federico II, E-mails: giacomo.miluccio@unina.it, fulvio.parisi@unina.it.*
[b]*Department of Civil, Environmental and Geomatic Engineering, University College London, E-mail: r.gentile@ucl.ac.uk, c.galasso@ucl.ac.uk*

ABSTRACT: Earthquake-induced landslides can significantly increase structural damage to buildings located on slopes with respect to the damage due to ground shaking only. Accordingly, multi-hazard vulnerability modelling of structures should account for the cumulative damage due to earthquake-induced ground motions and landslide-related actions. This paper presents a numerical procedure based on surrogate modelling for the computationally efficient prediction of structural response to earthquake shaking and resulting landslides if any. Specifically, discriminant classifiers and Gaussian process regression are used to develop metamodels mapping the main structural characteristics (e.g., geometric and material properties) to engineering demand parameters of interest. The proposed procedure is applied to an archetype reinforced concrete frame building designed only for gravity loads and located at the toe of a slope. Analysis results show that surrogate models allow a quick yet accurate prediction of structural performance.

KEYWORDS: buildings; earthquakes; landslides; Gaussian process regression; multi-hazard performance assessment.

1 INTRODUCTION

In most instances, damage induced by earthquakes' secondary hazards (such as explosions/fire, tsunamis, and landslides) is heavier than that caused by ground shaking. Recent events highlighted that earthquake-induced landslides could lead to severe economic and human losses (e.g., Yin, 2008; Cui et al., 2009). Indeed, landslides can be classified as one of the most dangerous secondary hazardous events triggered by earthquake-induced ground shaking. Many past studies focused on slow-type landslides (e.g., Fotopoulou & Pitilakis, 2017a,b) characterised by velocity almost equal to 0.6 mm per day. In this case, damage to buildings and infrastructure is mainly induced by differential soil settlements under those structures on the crest of slopes (Fotopoulou & Pitilakis, 2017a,b). In contrast, flow-type landslides – characterised by a velocity of up to 30 m/s – can induce heavy damage to buildings downstream of slopes due to earth-flow impacts. Moreover, different objects such as rocks, debris and cars can crash on the building façade, further exacerbating damage (e.g., Mavrouli et al., 2014).

Empirical vulnerability models for structures subjected to flow-type landslides were developed by Fuchs et al. (2007). Moreover, Zeng et al. (2015) collected historical damage data of past landslides by performing field investigations to investigate the failure modes of reinforced concrete (RC) columns impacted by earth flows. More recently, Parisi & Sabella (2017) investigated RC frame buildings subjected to landslide impact via 2D nonlinear structural models and simplified models of masonry infill walls, deriving analytical (or numerical) fragility relationships (i.e., likelihood of damage given a hazard intensity measure). Miluccio et al. (2020a,b) developed fragility relationships for 3D nonlinear structural models subjected to earthquake-induced landslides. The high computational effort associated with the derivation of fragility models for a building archetype representative of a building class (or type) is one of the main challenges of this type of analysis, particularly when both within-building and building-to-building variabilities are

considered.

In this study, two discriminant classifiers and a Gaussian Process (GP) regression are trained to investigate the structural performance of a selected structural type (aka building class) subjected to earthquake-induced ground motion and landslide. The computational procedure is based on the sequential pushover analysis procedure presented by Miluccio et al. (2020a,b), the results of which are used to define training datasets for surrogate modelling. The proposed surrogate models allow the computationally-efficient prediction of 1) the cumulative damage caused by a sequence of earthquake shaking and landslide impact on the considered building class, and 2) fragility estimates with high accuracy.

A pre-code European building class (i.e., designed before 1980) is considered as a case study. Specifically, the proposed surrogate models are trained for an archetype structure representative of such a case-study building class. Their accuracy is shown through confusion matrices for the classifiers, root-mean-square error (RMSE), normalised RMSE, and 95% coverage probability for GP regression.

2 METHODOLOGY

Two classifiers are calibrated to predict collapse due to 1) earthquake-induced ground motion; 2) and subsequent landslide impact. Moreover, a GP regression model is trained to estimate seismic demand in terms of chord rotation of the corner column that is also impacted by the subsequent landslide (the column 101 hereafter; see also Figure 3) for the non-collapse cases. The datasets to train and validate the surrogate models are both generated as described in the flow chart in Figure 1; Figure 2 shows the uses of these datasets to train the corresponding surrogate model.

The first step in the proposed methodology is uncertainty characterisation, including uncertainties related to geometric and material properties, loads, and capacity models. Next, a sample dataset is developed using a Latin Hypercube Sampling Method (LHS) (i.e., for the training dataset) or a full factorial Design of Experiments (DoE) approach (i.e., for the validation dataset). A first nonlinear static (i.e., pushover) analysis is carried out on the intact structure to estimate the structural performance under earthquake ground shaking applying the Capacity Spectrum Method (CSM; Freeman, 1998). An indicator function is set to 1 or 0 if collapse is reached or not, respectively. Two different datasets are thus generated with all samples corresponding to collapse, namely, one dataset associated with earthquake-induced collapse and the other related to landslide-induced collapse (i.e., this matrix considers the previous collapse due to seismic action as well) (Figure 1). If the indicator function is equal to 0, a dataset associated with the considered seismic demand (in terms of chord rotation of column 101) is generated, and a pre-damaged inelastic structural model is developed, considering the damage caused by the seismic load. Subsequently, another nonlinear static analysis with force control is performed to predict the structural performance under landslide impact. Then, the earthquake-induced landslide dataset is generated by including all collapse cases induced by earth-flow impact.

The above steps must be applied for each sample generated and each value of the considered ground-motion intensity measure (or IM; peak ground acceleration, PGA, in this study) and landslide IM (i.e., the impact velocity, v, in this study) to obtain the training and validation datasets.

The uses of the training datasets generated through this procedure are shown in Figure 2, where the two classifiers are introduced. The first classifier allows predicting whether earthquake-induced collapse is reached or not (Figure 2a). A GP regression model is trained using the seismic demand training dataset (Figure 2b). Finally, the second classifier is calibrated via the earthquake-induced landslide training dataset (Figure 2c), which consider the landslide-induced collapse cases.

The entire analysis procedure was implemented in MATLAB (The MathWorks Inc, 2020).

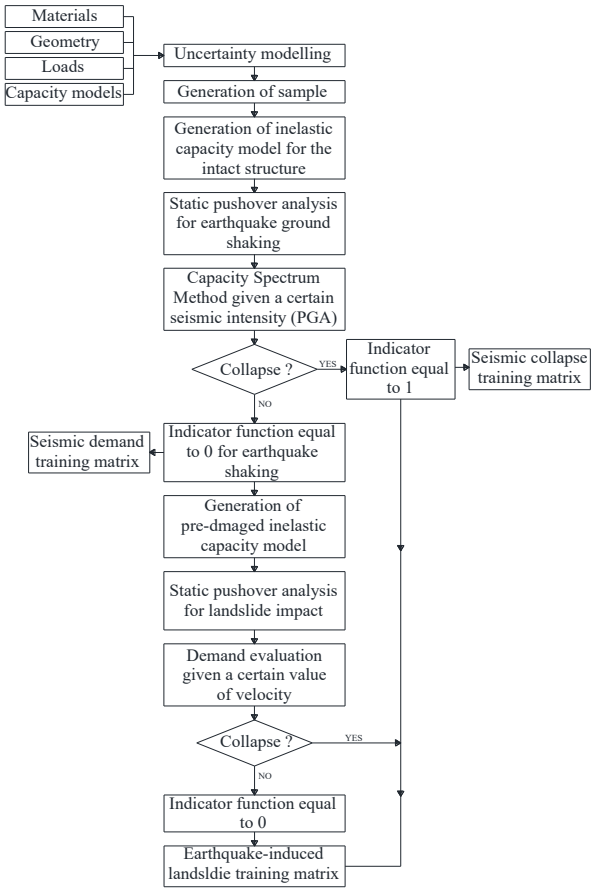

Figure 1. Analysis flowchart for one training point.

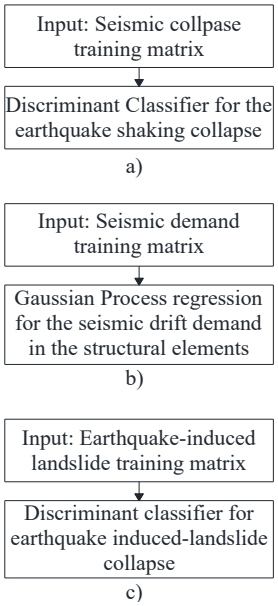

Figure 2. Training of the proposed surrogated models: a) Discriminant classifier for earthquake-induced collapse; b) Gaussian Process regression to evaluate the drift demand on structural elements due to seismic ground shaking ; c) Discriminant classifier for landslide-induced collapse.

Section 2.1 describes the two different sampling techniques adopted and the numerical structural modelling approach to define the training and validation sample sets. The analysis procedure is presented in Section 2.2, which deals with the discriminant classifiers and GP regression model.

2.1 *Sampling techniques and numerical structural modelling*

In this study, geometry, material, and capacity model parameters are considered random variables (RVs). Those RVs are assumed statistically independent. Each RV is defined through a probability distribution model characterised by 1) mean and standard deviation (or the coefficient of variation, CoV, i.e., the ratio of the standard deviation to the mean) when a normal or lognormal variable is defined; 2) a range (minimum and maximum values) when a uniform distribution is considered. Two sampling techniques are used, i.e. the LHS and the full factorial DoE to generate the training and validation dataset samples, respectively.

The full factorial DoE is implemented by considering n realisations of each RV to build the whole dataset by combining each value obtained. Based on the number of random parameters, N, N^n samples are generated.

LHS considers N hyperplanes, where N is the number of random parameters, and each of them is divided into several equally probable intervals N_{sim} (i.e., considered number of samples). In this way, N_{sim} hypercubes are created, and a value of each parameter is randomly sampled within each hypercube – i.e., a type of stratified Monte Carlo is considered.

The OpenSees software (McKenna et al., 2004) is used to build structural models according to a lumped plasticity approach. Eurocode 8 (EC8) – Part 3 (CEN, 2005) is used to define the plastic hinge behaviour through nonlinear springs with trilinear moment-chord rotation (M–θ) relationships. Each consists of a first elastic branch up to the yielding point, followed by a hardening branch up to the maximum strength ($M_{\max}$) and a degrading branch up to zero strength. The yielding point (θ_y, M_y) is defined by evaluating the yielding chord rotation, θ_y, through the EC8 formulation and the yielding moment, M_y, via sectional analysis. The

ultimate flexural chord rotation, $\theta_{u,f}$, is evaluated as the sum of the plastic rotation, θ_{pl}, and the yielding rotation, θ_y, according to EC8. The ultimate moment, M_u, is obtained as $0.8M_{max}$, estimating M_{max} as the moment leading to the ultimate strain of concrete in the most compressed fibre of the RC cross-section. The corresponding chord rotation, θ_{max}, is set equal to $0.75\theta_{u,f}$. The compressive behaviour of the concrete is modelled through a parabola-rectangle stress-strain diagram, whereas the behaviour of the reinforcing steel is assumed to be elastic-perfectly plastic.

The shear capacity model proposed by Biskinis et al. (2004) is adopted to consider the degrading shear behaviour of plastic hinges. Moreover, the linear capacity loss proposed by Zhu et al. (2007) is deemed to simulate shear capacity degradation accurately.

As discussed in Zeng et al. (2014), earth-flow impacts can induce damage to the midspan of the impacted columns. This motivated the implementation of a third nonlinear spring at the midspan column section together with the springs at the element ends. All of them are defined through the M–θ relationship described before.

A bare framed structural system is modelled, accounting for infill masonry walls as gravity loads. Floor slabs are considered rigid in their plane, whereas a fully fixed boundary condition is assigned to the base of the building (i.e., the soil-structure interaction is neglected). Uniformly distributed loads on beams are considered to simulate gravity loads according to their tributary floor areas and the presence of infill masonry walls.

2.2 *Analysis procedure*

Eigenvalue analysis is used to evaluate the modal shapes and vibration periods of each random realisation of the elastic structural model. Then, nonlinear static (pushover) analysis with displacement control is performed by considering the intact structure, modelled according to the lumped plasticity approach described above. A uniform load profile is used to represent the mass distribution over the building height. Pushover analysis is run by assuming an incidence angle ϑ_s of the lateral loads, hence considering their effects along the principal directions of the building plan. Finally, the CSM approach is applied to evaluate displacement demands and, consequently, structural performance for each pushover analysis. It is worth mentioning that the CSM is carried out adopting the equivalent viscous damping formulation provided in Priestley et al. (2007), using the elastic response spectrum proposed by EC8 – Part 1 (Type 1 spectrum) (CEN, 2004). Based on the biaxial interaction model proposed by Di Ludovico et al. (2013) for RC-column cross-sections, a single Demand-to-Capacity Ratio (DCR) for each structural member is evaluated to quantify damage due to earthquake-induced ground motion.

If the maximum DCR reaches or exceeds unity, then the building is assumed to collapse due to the seismic ground motion, the procedure is stopped, and the sample is included in the seismic collapse training dataset. Otherwise, DCRs related to all frame members are considered to define a second inelastic model representative of a pre-damaged structure. Pushover analysis with force control is thus carried out to assess the effects of the landslide impact on the pre-damaged structure. In this case, all structure samples with $DCR < 1$ (i.e., non-collapse cases) are considered.

The landslide impact results in lateral pressures over the height of the impacted columns that can be modelled as the sum of a linearly distributed pressure (hydrostatic component) and a uniformly distributed pressure (kinetic component), according to Zanchetta et al. (2004), as follows:

$$p_{dx}(z_s) = \rho g(D - z_s) + \rho v^2 \cos^2 \delta_f \quad (1)$$

$$p_{dz}(z_s) = \rho g(D - z_s) + \rho v^2 \sin^2 \delta_f \quad (2)$$

where: g = acceleration of gravity; D = flow depth; ρ = flow density; z_s = height of generic soil layer from the sliding surface of the flow; v = impact flow velocity; δ_f = angle between flow direction and perpendicular axis of impacted building façade ($\delta_f = 0$ in case of perpendicular impact).

In this study, according to Parisi and Sabella (2017), a uniform equivalent pressure (p_{eq}) is computed to define a distributed load per unit length over the height of the columns by assuming the earth-flow height equal to the inter-story height, as follows:

$$q_x = p_{eq} h_c \cos \delta_f \quad (3)$$

$$q_z = p_{eq} b_c \sin \delta_f \quad (4)$$

where b_c and h_c indicate the dimensions of

column cross-section along the X- and Z-axis of the building plan, respectively. Moreover, the flow impact angle (δ_f) and the extension of landslide on the plan (L_f) are considered as variables.

Thus, landslide-related loading is modelled as uniformly distributed pressures both in plan and over the height of columns, as shown in Figure 3.

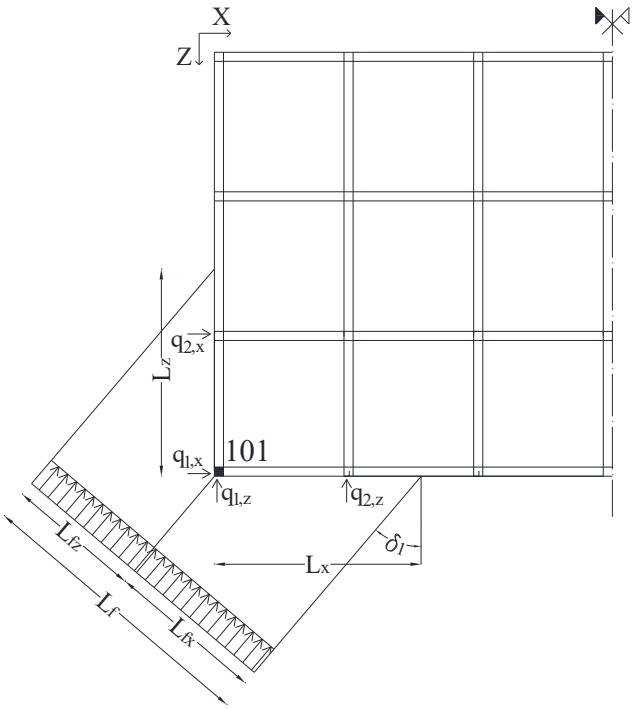

Figure 3. Landslide load distribution in plan (q_x and q_z are the load components along the building axes, L_x and L_z measure the portions of building façade along the building axes impacted by the landslide; L_f is the landslide size in plan).

The entire procedure is consistent with the assumption that landslide impact follows seismic shaking. Sequential pushover analysis is performed to assess the cumulative damage to structural elements due to earthquake-induced ground motion and landslide impact. In this respect, it is assumed that structural collapse is reached when column 101 achieves either the ultimate chord rotation or maximum shear strength. Figure 4 shows the sequence of structural analyses performed under lateral actions representing the effects of earthquake-induced ground motion and lateral actions representing the landslide impact.

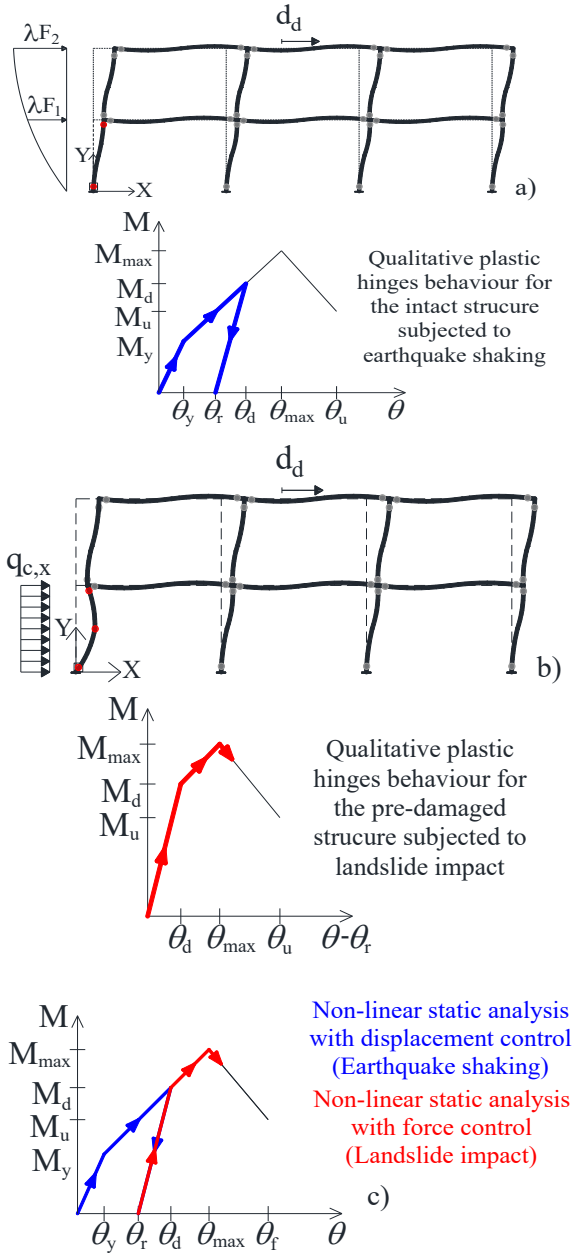

Figure 4. Sequential nonlinear static analysis: a) Pushover analysis on intact structure, b) landslide impact on the pre-damaged structure, c) plastic hinge behaviour during the whole procedure (plastic hinge behaviour characterised by yielding chord rotation θ_y, yielding moment M_y, maximum chord rotation θ_{max}, maximum moment M_{max}, ultimate chord rotation θ_u, ultimate moment M_u, rotation demand θ_d and moment demand M_d induced by earthquake shaking, and residual chord rotation θ_r).

2.3 *Discriminant classifier and GP regression*

The sampling techniques described in Section 2.1 are adopted to define a vector $\boldsymbol{x}$ that includes any geometric characteristics and material properties considered in the analysis. Thus, the

matrix $\boldsymbol{X}$ is set to comprise all input vectors $\boldsymbol{x}$ (for all the training points). In addition, a vector $\boldsymbol{y}$ is defined to include the selected engineering demand parameter (EDP) values, i.e., the computed chord rotation seismic demand, for all the considered training points.

The discriminant analysis is used to train a discriminant classifier and to predict the collapse or non-collapse cases for the buildings under study. In this regard, two data groups are considered as input and output, namely, collapse and non-collapse.

The probability that K = k given a certain observation $\boldsymbol{x}$ of $\boldsymbol{X}$ (i.e., Y is the class predicted, and k is the observed class) can be obtained by applying the Bayes theorem. Thus, Pr(Y = k | $\boldsymbol{x}$), can be defined as follows:

$$\Pr(K = k \mid \boldsymbol{x}) = \frac{f_k(\boldsymbol{x})\pi_k}{\sum_{i=1}^{n_K} f_i(\boldsymbol{x})\pi_i} \quad (5)$$

where $f_k(\boldsymbol{x})$ is the class density function; π_k is the prior probability of class (i.e., the proportion of training data that belongs to the k^{th} class).

In linear discriminant analysis (LDA), the class density function is defined through a multivariate Normal model assuming that all classes have the same covariance matrix. Since LDA via least squares does not use a Gaussian assumption for the predictors, its applicability extends beyond the realm of Gaussian data (Friedman et al., 2008). Therefore, LDA is performed, and linear decision boundaries among classes are defined by inserting the given $f_k(\boldsymbol{x})$ into Eq. 5, comparing two classes, and considering their log-ratio. Therefore, the linear discriminant function is given by Eq. 6:

$$\delta_k(\boldsymbol{x}) = \boldsymbol{x}^T \sum^{-1} \mu_k - \frac{1}{2}\mu_k^T \sum^{-1} \mu_k + \log(\pi_k) \quad (6)$$

where δ_k is a linear function of $\boldsymbol{x}$; hence, linear decision boundaries exist between the classes.

More details about LDA can be found in (Lachenbruch and Goldstein 1979; Friedman et al., 2008; McLachlan 2004).

Next, a GP regression is derived. The relationship $\boldsymbol{y} = f(\boldsymbol{x})$ between $\boldsymbol{x}$ and $\boldsymbol{y}$ is surrogate through a statistical model. Thus, the GP can predict the EDP of interest given a given input vector faster than the numerical analysis.

The GP describing the relationship $f(\boldsymbol{x})$ is defined by its mean function $m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})]$ and its covariance $k(\boldsymbol{x},\boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))]$ where $\boldsymbol{x}$ and $\boldsymbol{x}'$ are two different input vectors:

$$f(\boldsymbol{x}) \sim GP(m(\boldsymbol{x}),(\boldsymbol{x},\boldsymbol{x}')) \quad (5)$$

The properties of the output function - 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. Therefore, the structure of the covariance function is selected to 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 material properties produces small changes in the considered output EDP). In this case, two hyperparameters are used to describe the kernel, σ_i and σ_f. The former is representative of the length scale of the output given an input dimension i. Therefore, in the case of small values of σ_i, the GP regression outputs can change quickly with respect to the building's attributes. In contrast, σ_f is the signal variance and represents the variability of the output given an input $\boldsymbol{x}$. The prediction of the training data $\boldsymbol{y}$ is defined by maximising the likelihood $p(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \{\sigma_i, \sigma_f\}$ is a vector of hyperparameters and $\boldsymbol{X}$ is the training input dataset. This procedure is carried out through MATLAB using a quasi-Newton numerical optimisation algorithm.

3 CASE STUDY

An archetype structure representative of the European pre-code building class is chosen as a case study, considering low-rise buildings designed for gravity loads only. Figure 5 shows the structural system, consisting of two storeys with the same inter-storey height and a rectangular plan with two by six bays in the two principal global directions. Both material and geometric properties are described in Table 1 as RVs (Crowley et al., 2004; Parisi & Sabella, 2017).

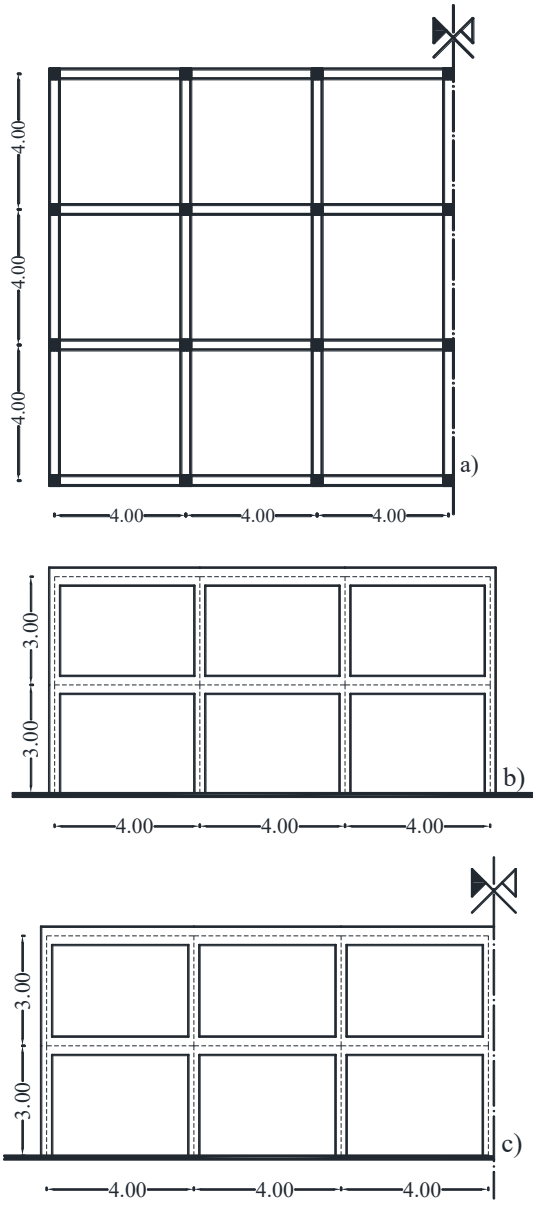

Figure 5. a) Building plan; b,c) building elevation.

Beams are assumed to have a rectangular cross-section with depth and width equal to 0.5 m and 0.3 m, respectively. LHS and DoE sampling techniques are applied to define the validation and training sample datasets, respectively, as discussed above. Specifically, LHS is implemented by assuming uniformly distributed RVs in the range $[\mu - \sigma, \mu + \sigma]$ (denoting mean and standard deviation of each RV as μ and σ, respectively) and by generating 3000 samples. Conversely, DoE is implemented by considering three values for θ_s, δ_f, L_f, f_c, f_y (i.e., $\mu - \sigma$; μ; $\mu + \sigma$). Instead, four values of b_c and $V_{u,exp}/V_{u,theor}$ (i.e., $\mu - \sigma$; $(\mu - \sigma)/2$ μ; $(\mu + \sigma)/2$) are considered due to their high influence on the structural response of the case-study structure for the specific problem under investigation. The total samples generated through the DoE are 3888.

Table 1. Considered random variable (RVs) and their probability density function (PDFs).

Item	RV	μ/*Range*	CoV	PDF
Earthquake	θ_s	0 – 90°	-	Uniform
Landslide	δ_f	0 – 90°	-	Uniform
	L_f	$h_{col} - d_b$	–	Uniform
Concrete	f_c	2.73 MPa	31%	Lognor.
Steel	f_y	245 MPa	10%	Normal.
Column	$b_c = h_c$	-1.03 m	37%	Lognor.
Capacity Model	$V_{u,exp}/V_{u,th.}$	-0.03	25%	Lognor.

The training and validation datasets for the case-study building are defined via the procedure summarised in Figure 1, assuming 15 values of PGA (i.e., from 0.05g to 0.75g with a step of 0.05g) and 11 values of landslide velocity (i.e., form 0 m/s to 10 m/s). Therefore, the discriminant classifiers are trained by considering collapse cases associated with seismic shaking and landslide impact. In this regard, two training matrices are defined: (i) the matrix associated with seismic collapses includes 45000 realisations (i.e., 3000 samples multiplied for 15 values of PGA) and four predictors (f_c, f_y, b_c, PGA); (ii) the matrix related to the landslide collapses comprises 495000 realisations (i.e., 3000 samples multiplied for 15 values of PGA and 11 velocities) and five predictors (f_c, f_y, b_c, PGA, v). Moreover, the GP regression is calibrated based on the seismic demand training matrix (Figure 1) that includes the same predictors of (i). In this study, four predictors are chosen to train the surrogate models in order to implicitly include the loads and capacity models variability into them. In this way, the models can be used considering the material and geometry variables that can be easily derived from literature or surveys.

On the other hand, the validation datasets include: (i) 6480 realisations for the seismic collapses matrix (i.e., 432 samples timed 15 values of PGA) due to the fact that the landslide

load parameters (i.e., L_f and δ_f) do not influence the seismic structural response (all samples related to their variability are not considered in this matrix); (ii) 641520 realisations for the earthquake-induced landslide collapses matrix (i.e., 3888 samples times 15 values of PGA and 11 values of velocity).

The surrogate models prediction power isevaluated by considering the validation dataset. Finally, confusion matrices for the classifiers, RMSE, normalised RMSE and 95% coverage probability for GP regression are assessed to define the surrogate models' prediction quality.

3.1 *Results*

A confusion matrix is one of the most useful tools to describe classifier performance. Such a matrix collects the actual and predicted outputs to evaluate the quality of predictions. Three parameters can be inferred from the confusion matrix: accuracy, precision, and recall. Accuracy is the number of correct predictions over the number of samples adopted. Precision is defined as the percentage of correct predictions for each class. Recall, also called sensitivity in binary classification, is the percentage of correct predictions over the actual samples in the class.

The discriminant classifier produces a squared confusion matrix, where the main diagonal represents the correct predictions of the classifiers. Figure 6 shows the confusion matrices obtained for both classifiers by applying the validation sample sets (i.e., the sample sets generated through the DoE method).

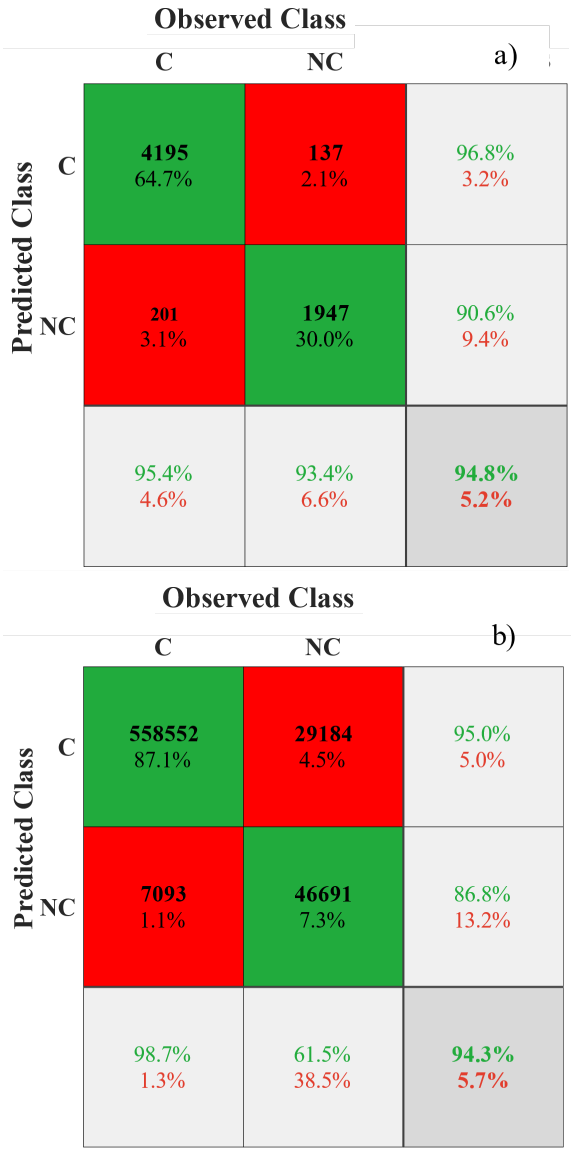

Figure 6. Confusion matrix of discriminant classifiers for (a) earthquake shaking; and (b) landslide impact.

The accuracy is displayed in the bottom right corners in terms of percentage (i.e., the green text). In contrast, the precision and recall are shown in the last column (on the right-hand side) and the last row (at the bottom) of the confusion matrices, respectively.

The analysis results show that the discriminant classifiers have an accuracy of 94% in both cases. Moreover, the precision and recall values are both higher than 85%. These results show good quality for both classifiers, as reported in Mangalathu and Jeon (2019). Thus, the proposed classifiers have a high power to predict the collapse/non-collapse of the considered structure. It is worth noting that the proposed classifiers

consider the uncertainties in material and geometric properties (i.e., f_c, f_y, b_c) and loads parameters (θ_s ,L_f, δ_f).

The proposed GP regression is trained to predict the seismic chord rotation demand on column 101. The high accuracy in the GP regression's prediction is first shown in Figure 7, where the multi-dimensional function is reduced to a 3D plot by considering the variability of two parameters and assuming a constant value of the others. Therefore, the GP regression is tested by applying the DoE dataset. In particular, Figure 7 shows the GP regression by assuming f_c = 16 MPa and f_y = 220.5 MPa. The shaded area in Figure 7 represents the 95% confidence interval of the predictions, whereas the grey surface is the mean value of regression that fits the testing points in red.

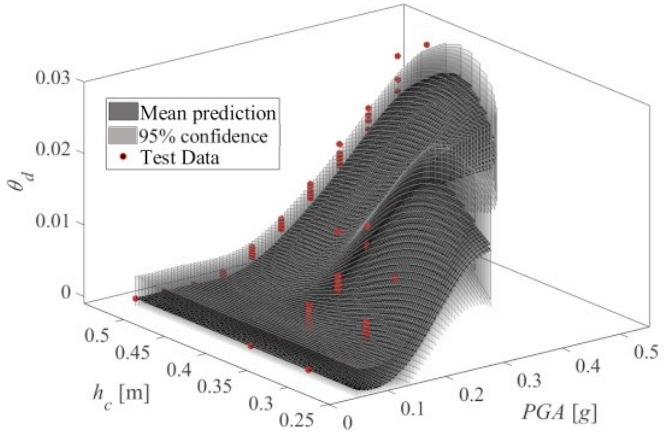

Figure 7. Example of Gaussian Process prediction for column 101 chord rotation due to earthquake ground shaking (f_c = 16.0 MPa, f_y = 220.5 MPa).

It can be noted that the GP regression is not fitted on the entire variables' intervals (e.g., for PGA values of 0.*05g* to 0.75*g*) because in some regions of the defined hyperspace, there are no training data. Indeed, there is a lack of training points for high values of PGA and lower values of b_c. This aspect is easily explained by following the procedure illustrated in Figure 2. Indeed, the classifier identifies the collapses due to earthquake shaking, and these input data are not included in the subsequent GP regression. Thus, the seismic collapse classifier acts as a filter.

The RMSE has a value of $3.1 \cdot 10^{-6}$ in terms of chord rotation demand. Additionally, the dimensionless RMSE is obtained by dividing the computed (absolute) RMSE by the mean value of chord rotation demand of the validation dataset. The normalised RMSE is equal to 0.08%. Moreover, the coverage probability, defined as the proportion of the confidence intervals of the surrogated model that contains the modelled data, is equal to 92.8%, which is very close to the theoretical value of 95%.

The surrogate models presented in this paper can describe the range of structural responses under earthquake shaking and resulting earthquake-induced landslides. It is important to highlight that the procedure to define the training matrices is based on sequential nonlinear static analyses that consider the cumulative damage induced by the seismic load and consequent landslide impact. Thus, the surrogate models can predict the damage caused by earthquake ground motion and predict the collapse of the pre-damaged structure subjected to the earth-flow impact. Based on the first classifier, the building class fragility curves can be easily derived in the case of structural collapse due to earthquake shaking. Moreover, fragility curves for different Damage States (DS) can be defined by applying the GP regression. Then, the last classifier should be adopted to obtain two more sets of fragility curves related to the earthquake-induced landslide collapse given a certain seismic DS or a defined seismic IM (i.e., a certain value of PGA).

4 CONCLUSIONS

This paper described a procedure to develop surrogate models for a considered archetype building representative of a building class. Uncertainties in material and geometric properties (i.e., f_c, f_y, b_c) as well as in loads parameters (θ_s ,L_f, δ_f) were modelled and propagated in the analysis. Training and validation datasets were generated by applying two different sampling techniques, namely, Latin Hypercube sampling and full factorial Design of Experiments, respectively. Two discriminant classifiers and one GP regression were trained to predict the earthquake-induced and landslide-induced collapse and structural response under horizontal seismic actions, respectively. A sequential nonlinear static analysis was performed for each sample to predict structural response and cumulative damage induced by earthquake shaking and landslide impact. In this regard, a pre-damaged structural model was developed to assess response to earth-flow impact, accounting

for the previous seismic damage.

A case study was considered to illustrate the methodology described above and train/validate the surrogate models. The building considered is representative of European low-rise buildings designed for gravity loads only (pre-code building class).

The surrogate models were trained using 45000 and 495000 realisations for the seismic and landslide matrices, respectively. On the other hand, the validation datasets comprise 6480 and 641520 realisations for the seismic and landslide matrices, respectively. The validation dataset was applied to define the prediction power of the proposed surrogate models. In this regard, the confusion matrixes and RMSE estimates were obtained for the classifiers and GP regression, respectively. The two discriminant classifiers show an accuracy higher than 94%, with precision for each class equal to about 90%. The seismic structural response (i.e., the chord rotation demand on the corner column impacted by the subsequent landslide) was finally predicted through the GP regression. The high predictive power of GP regression was highlighted by a value of RMSE equal to $3.1 \cdot 10^{-6}$ and a corresponding dimensionless RMSE equal to 0.08%. Moreover, the calculated coverage probability is equal to 92.8% which is very close to the theoretical value of 95%.

The proposed surrogate models can be used to derive building-class fragility models by including both the within-building variability and cumulative damage with high accuracy and sustainable computational effort compared to traditional fragility analyses based on nonlinear analysis of a large number of structural models. In addition, the CSM method using real (i.e., recorded) records can be applied to obtain more refined seismic performance estimates. In this regard, the recently proposed cloud-CSM (Nettis et al., 2021) should be considered in future work or even nonlinear dynamic analyses instead of nonlinear static ones. Finally, based on past landslide events, the impact of objects on the building's columns during a landslide could be a realistic scenario to consider in some cases. Therefore, future studies should also consider the dynamic action due to such collisions to define surrogate models that can predict a premature failure of impacted columns.

AKNOWLEDGEMENTS

This study was developed in the framework of PON INSIST (Sistema di monitoraggio INtelligente per la Sicurezza delle infraSTrutture urbane) research project, which was funded by the Italian Ministry for Education, University and Research (Programma Operativo Nazionale "Ricerca e Innovazione 2014–2020", Grant No. ARS01_00913).

5 REFERENCES

Biskinis D., Roupakias G.K., Fardis M.N. 2004. Degradation of shear strength of reinforced concrete members with inelastic cyclic displacements. ACI Structural Journal, 101(6), 773-783.

Crowley H., Pinho R., Bommer, J.J. 2004. A Probabilistic Displacement-based Vulnerability Assessment Procedure for Earthquake Loss Estimation. Bull Earthquake Eng 2, 173–219.

Cui, P., Zhu, Yy., Han, Ys. et al. 2009. The 12 May Wenchuan earthquake-induced landslide lakes: distribution and preliminary risk evaluation. Landslides 6, 209–223.

Di Ludovico M., Verderame G.M., Prota A., Manfredi G., Cosenza E. 2013. Experimental behaviour of nonconforming RC columns with plain bars under constant axial load and biaxial bending. Journal of Structural Engineering, 139(6), 897-914.

EN 1998-1:2004, Eurocode 8: Design of structures for earthquake resistance – Part 1: General rules, seismic actions and rules for buildings. Brussels: Comité Européen de Normalisation.

EN 1998-3:2005, Eurocode 8: Design of structures for earthquake resistance – Part 3: Assessment and retrofitting of buildings. Brussels: Comité Européen de Normalisation.

Fotopoulouand S.D. & Pitilakis K.D. 2017. Probabilistic assessment of the vulnerability of reinforced concrete buildings subjected to earthquake induced landslides. Bulletin of Earthquake Engineering, 15, 5191-5215.

Fotopoulouand S.D. & Pitilakis K.D. 2017. Vulnerability assessment of reinforced concrete buildings at precarious slopes subjected to combined ground shaking and earthquake induced landslide. Soil Dynamics and Earthquake Engineering, 93, 84-98.

Freeman SA. 1998. Development and use of capacity spectrum method. 6th U.S. Natl. Conf. Earthq. Engineering, Seattle.

Friedman J., Hastie T., and R. Tibshirani. 2008. The elements of statistical learning: Springer series in statistics. Berlin: Springer.

Fuchs S, Heiss K, Hübl J. 2007. Towards an empirical vulnerability function for use in debris flow risk assessment. Nat Hazards Earth Syst Sci, 7:495–506.

Gentile R., Galasso C. 2020. Gaussian process regression for seismic fragility assessment of building portfolios. Structural Safety, 87.

Lachenbruch, P. A., and M. Goldstein. 1979. "Discriminant analysis." Biometrics 35 (1): 69–85. https://doi.org/10.2307/2529937.

Mangalathu S. & Jeon J.S. 2019 Machine Learning–Based Failure Mode Recognition of Circular Reinforced Concrete Bridge Columns: Comparative Study. Journal of Structural Engineering, 145(10).

MATLAB R2020b. Natick, Massachusetts: The MathWorks Inc.

Mavrouli O., Fotopoulou S., Pitilakis KD., Zuccaro G., Corominas J., Santo A., Cacace F., De Gre-gorio D., Di Crescenzo G., Foerster E., Ulrich T. 2014. Vulnerability assessment for reinforced concrete buildings exposed to landslides. Bulletin of Engineering Geology and the Environment, 73(2), 265-289.

McKenna F., Fenves G.L., Scott M.H. 2004. OpenSees: Open system for earthquake engineering simulation. California: Pacific Earthquake Engineering Research Center, University of California Berkeley. Available at http://opensees.berkeley.edu.B.

McLachlan, G. 2004. Discriminant analysis and statistical pattern recognition. Hoboken, NJ: Wiley.

Miluccio G., Parisi F., Cosenza E. 2020. Impact of cumulative damage on fragility of RC framed buildings subjected to earthquake-induced landslides. Eurodyn 2020, XI International Conference on Structural Dynamics, Athens, Greece.

Miluccio G., Parisi F., Cosenza E. 2020. Fragility curves for RC framed buildings subjected to earthquake-induced landslide: comparison between 2d and 3d structural models. Eurodyn 2020, XI International Conference on Structural Dynamics, Athens, Greece.

Nettis A., Gentile R., Raffaele D., Uva G., Galasso C. 2021. Cloud Capacity Spectrum Method: Accounting for record-to-record variability in fragility analysis using nonlinear static procedures, Soil Dynamics and Earthquake Engineering, 150.

Parisi F. & Sabella G. 2017. Flow-type landslide fragility of reinforced concrete framed buildings, Engineering Structures, 131, 28-43.

Wang X., Nieand G., Wang D. 2010. Relationships between ground motion parameters and landslides induced by Wenchuan earthquake. Earthquake Science, 23(3), 233-242.

Yin, Y., Wang, F., Sun, P. 2009 Landslide hazards triggered by the 2008 Wenchuan earthquake, Sichuan, China. Landslides 6, 139–152.

Zanchetta G, Sulpizio R, Pareschi MT, Leoni FM, Santacroce R. 2004. Characteristics of May 5–6, 1998 volcaniclastic debris flows in the Sarno area (Campania, southern Italy): relationships to structural damage and hazard zonation. J Volcanol Geotherm Res, 133(1):377–93.

Zeng C, Peng C, Zhiman S, Lei Y, Chen R. 2014. Failure modes of reinforced concrete columns of buildings under debris flow impact. Landslides, 12:561–71

Zhu L, Elwood K, Haukaas T. 2007. Classification and seismic safety evaluation of existing reinforced concrete columns. J Struct Eng, 103(9):1316–30.