

Surrogate probabilistic seismic demand modelling of inelastic single-degree-of-freedom systems for efficient earthquake risk applications

Roberto Gentile^{1,2}  | Carmine Galasso^{2,3} 

¹ Institute for Risk and Disaster Reduction, University College London, London, UK

² Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

³ Scuola Universitaria Superiore IUSS, Pavia, Italy

Correspondence

Roberto Gentile, Institute for Risk and Disaster Reduction, University College London, London, UK.

Email: r.gentile@ucl.ac.uk

Funding information

European Union's Horizon 2020, Marie Skłodowska-Curie Action (MSCA) - IF-2018, Grant/Award Number: 843794

Abstract

This paper proposes surrogate models (or metamodels) mapping the parameters controlling the dynamic behaviour of inelastic single-degree-of-freedom (SDoF) systems (i.e., force-displacement capacity curve, hysteretic behaviour) and the parameters of their probabilistic seismic demand model (PSDM, i.e., conditional distribution of an engineering demand parameter [EDP] given a ground-motion intensity measure [IM]). These metamodels allow the rapid derivation of fragility curves of SDoF representation of structures. Gaussian Process (GP) regression is selected as the metamodeling approach because of their flexibility in implementation, the resulting accuracy and computational efficiency. The metamodel training dataset includes 10,000 SDoF systems analysed via cloud-based non-linear time-history analysis (NLTHA) using recorded ground motions. The proposed GP regressions are tested in predicting the PSDM of both the SDoF database (through ten-fold cross validation) and eight realistic reinforced concrete (RC) frames, benchmarking the results against NLTHA. An application is conducted to propagate such modelling uncertainty to both fragility and vulnerability/loss estimations. Error levels are deemed satisfactory for practical applications, especially considering the low required modelling effort and analysis time. Regarding single-building applications enabled by the proposed metamodel, this paper presents a first attempt at a direct loss-based design procedure, which allows setting a target loss level for the designed structure (shown for a realistic RC frame). An earthquake risk model involving dynamic exposure and vulnerability modules is illustrated as an example of building portfolio applications. Specifically, the proposed application considers a retrofit-based seismic risk-reduction policy for a synthetic building portfolio, for which it is possible estimating the loss evolution over time.

KEYWORDS

direct loss-based design, dynamic exposure risk modelling, earthquake risk modelling, gaussian process regression, statistical modelling

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Earthquake Engineering & Structural Dynamics* published by John Wiley & Sons Ltd.

1 | INTRODUCTION AND MOTIVATION

Building-level seismic fragility is defined as the probability of reaching or exceeding various damage states (DSs) as a function of a hazard intensity measure (IM). DSs are usually expressed in terms of engineering demand parameter (EDP) thresholds, such as inter-storey drift limits. In the realm of analytical (or numerical) fragility analysis methods, non-linear time-history analysis (NLTHA) of refined structural models is the best practice when it comes to building-specific applications.^{1,2} On the other hand, NLTHA of equivalent single degree of freedom (SDoF) systems is commonly used to characterise a building vulnerability class,² with the results being used for building-portfolio earthquake risk analysis.³ This reflects the trade-off between simplicity and accuracy that different stakeholders generally tolerate on the matter:⁴ private owners likely need a detailed assessment of individual buildings or small building portfolios, while governmental agencies or (re)insurance companies generally look at large portfolios accepting larger uncertainties and a lower analysis refinement level.

The above-mentioned analysis methods, respectively targeted at single-building and building-portfolio applications, allow an analyst to easily overcome computational issues, especially considering the excellent performance of ordinary modern computers. However, at least arguably, some applications may still be computationally unaffordable if using the above approaches. An example of such applications related to single buildings may be loss-based conceptual seismic design. This would require calculating a loss metric (e.g., the expected annual loss [EAL]) for (several) tentative structural configurations, considering a specific seismic hazard profile: only the configurations complying with the target loss level will be used to continue the detailed design process. Even considering SDoF-based NLTHA, the required analysis time may not be compatible with such a preliminary/conceptual design process, which requires pseudo-instantaneous analysis results. Lower-refinement results may suffice, and more refined analysis methods can be applied in the subsequent phases of the design. Considering building portfolios, the design of regional policies for seismic risk reduction (e.g., based on structural retrofit and/or seismic insurance) may be computationally unaffordable using SDoF NLTHA. In fact, this may require an earthquake risk model with many exposure scenarios (which may also be time-dependent), allowing selecting the optimal policy using one or more loss metrics.

Metamodelling approaches may be suitable for the applications mentioned above. A metamodel is “a model of a model”: it defines a (statistical) relationship between a given set of inputs and outputs (obtained through numerical simulations). In this particular case, metamodels mapping the parameters controlling the dynamic behaviour of inelastic SDoF systems (e.g., force-displacement capacity curve, hysteretic behaviour) and their probabilistic seismic demand model (PSDM; i.e., conditional probability distribution of an EDP given an IM) seems suitable. In fact, those can easily lead to building-level fragility curves for a set of structure-specific DSs, enabling the computation of various risk metrics/decision variables (e.g., economic losses, casualties, downtime). Different metamodels were proposed in the literature for the seismic response of structures, often referring to refined multi-degree-of-freedom (MDoF) structures (e.g., ref. 5,6). For the scope of this paper, the most relevant example of such metamodels is SPO2IDA,⁷ surrogating the different percentiles of incremental dynamic analysis (IDA) curves of SDoF systems from their piecewise static pushover (SPO) curve, which include elastic, hardening and degrading branches. This parametric model is based on regressions, and it is reasonably accurate within the range of its training dataset. However, the authors highlight its limitations related to the low number of ground-motion records (i.e., 30) and the single hysteresis model (moderately pinching) adopted for the training of the model. The authors provide a methodology to extend the training dataset to other periods by using their proposed functional forms. Although a similar methodology may be applied to extend the range of applicability of SPO2IDA to different parameters (e.g., considering other hysteresis models), a fixed functional form may fail to capture the complex interaction between the input parameters and the parameters of the resulting PSDMs. An example of this process is presented in ref. 8.

In this paper, a metamodelling approach is proposed to surrogate SDoF PSDMs while addressing the above issues, enabling efficient seismic risk assessment applications. First, it is recognised that most widely adopted DS descriptions (e.g., ref. 9–11) often cover DS levels up to the so-called “near collapse.” This generally involves a structural deformation level corresponding to the peak of the pushover curve or to a 20% post-peak strength drop (which can be reasonably close to the peak, depending on the softening branch of the pushover). In this context, characterising PSDMs above such deformation levels may be avoided by accepting an approximation of the results. Restricting the PSDM characterisation to pre-peak deformation levels allows drastically reduce the number of SDoF systems in the training set. Moreover, a cloud-based analysis approach¹² is herein preferred to IDA, to use a higher number of seed ground-motion records at a lower computational cost. Most importantly, a Gaussian Process (GP) regression is adopted since it does not require any prior definition of the output functional form (a GP regression is a non-parametric model). Not only this approach would

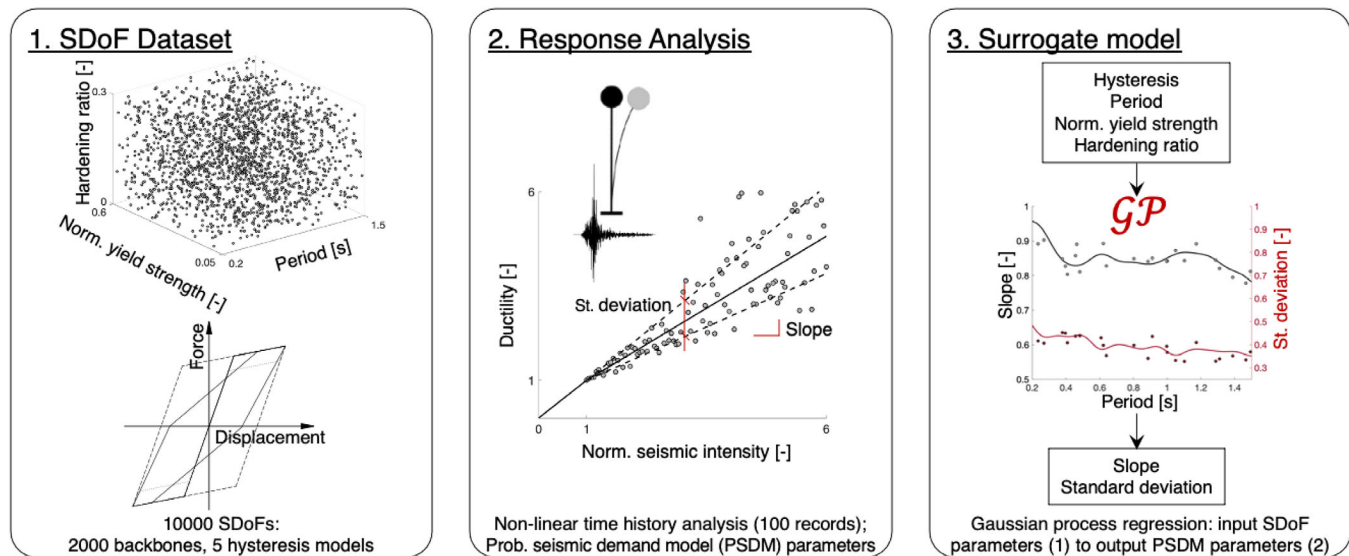


FIGURE 1 Steps for the calibration of the proposed surrogate model

ensure high accuracy and flexibility in implementation for earthquake engineering applications (e.g., ref. 13–15), but it will also result in an infinitely scalable surrogate model. In fact, it will be possible to add new case studies to the training dataset (e.g., adding more hysteresis models or damping ratios) and re-fit the GP regressions without the need to change the metamodel structure.

After describing the metamodeling strategy (Section 2), its accuracy is shown in Section 3, considering both SDoF systems and eight reinforced concrete (RC) frame case studies. Finally, Section 4 shows two use cases for the provided metamodel: direct loss-based seismic design (DLBD) and scenario-based dynamic earthquake loss modelling for risk-informed decision making.

2 | METAMODELLING STRATEGY

The mapping between the parameters controlling the dynamic behaviour of the considered inelastic SDoF systems and their PSDM consists of three phases (Figure 1). First, a number of input parameters directly affecting the PSDM are selected, together with their desired range (described in detail in Section 2.1). Plain Monte Carlo sampling is performed to generate a database of case-study SDoF systems. Second, each SDoF system is analysed via cloud analysis adopting a suite of 100 ground-motion records. A PSDM is constructed for each SDoF sample in the training dataset (described in detail in Section 2.2). A GP regression is finally fit to surrogate the complex relationship between the input SDoF parameters and the output PSDM ones (Section 2.3). It is worth mentioning that the GP regressions are tested with a ten-fold cross validation (Section 3).

The code implementing the proposed metamodel is freely available (<https://github.com/robgen/surrogatedPSDM>). In addition, the same repository includes the adopted training datasets and the code to train the GP regressions so that users can filter, extend or modify the training dataset and update the fitting.

2.1 | Generation of the SDoF database

The selected input parameters defining each SDoF system (Figure 1A) are the considered hysteresis model “*hyst*,” governing stiffness degradation under unloading-reloading conditions; the fundamental period, T , controlling the elastic stiffness; the yield shear strength (normalised to the total weight), f_y ; the hardening ratio, h . It is worth mentioning that no within-cycle strength degradation is considered since the surrogated PSDMs will not be used for predictions exceeding the ductility at the peak pushover strength. For the same reason, no explicit limit for the ductility (or displacement) at peak capacity is considered.

Five different hysteresis models are adopted for the model training (Figure 1A):

- Kinematic hardening, KIN, showing no stiffness degradation, mimicking the response of steel structures. Such an interpretation implicitly neglects the Bauschinger effect, which is likely not relevant for deformation levels smaller than the peak pushover response;
- Modified Takeda “fat,”¹⁶ MTf, appropriate for structures with a non-linear response dominated by members with low axial load (e.g., newly-designed RC frames dominated by beam behaviour). The unloading (α) and reloading (β) stiffness coefficients are respectively equal to 0.3 and 0.6;¹⁷
- Modified Takeda “thin,” MTt, appropriate for structures with a non-linear response dominated by members with large axial load (e.g., soft storey-like RC frames; RC bridge piers). The α and β stiffness coefficients are respectively equal to 0.5 and 0;¹⁷
- Modified Sina, MS, appropriate for existing structures exhibiting pinching response (e.g., RC frames dominated by joint failure; timber structures dominated by the failure of their connections). The unloading and reloading stiffness coefficients are respectively equal to 0.3 and 0.6, while pinching is defined by a closing force equal to 25% of the yield force;¹⁷
- Flag shape, FS, appropriate to simulate the behaviour of hybrid pre-stressed structures with re-centring behaviour. The energy dissipation coefficient (γ), controlling the force level where the re-centring action starts, is equal to 0.5.

In fitting the GP regression, the hysteresis type is considered a categorical (i.e., non-numerical) variable, as opposed to explicitly mapping the response of the SDoF systems to the numerical parameters of the hysteresis models (e.g., α and β for the MT). This is done for three reasons: (1) because the hysteresis-parameter types are different for each hysteresis model; (2) to avoid excessive interpolation error (e.g., between the two sets of MT hysteresis parameters); (3) to avoid a high increase in the number of required SDoF case studies (to consider more hysteresis-parameters sets).

For each hysteresis model, 2,000 realisation of the parameters (T, f_y, h) are sampled via a plain Monte Carlo approach, adopting uniform distributions. As discussed in Rasmussen and Williams,¹⁸ this is an appropriate assumption when adopting GP regressions as the focus is the input-output mapping rather than on modelling the specific distribution of the input parameters. The specific parameters of the assumed distributions, clearly defining the scope of the final metamodels, are $T \sim U(0.2s, 1.5s)$; $f_y \sim U(0.05, 0.6)$; $h \sim U(0, 0.3)$.

2.2 | Seismic response analysis and PSDM

Each SDoF system in the dataset is analysed to derive a cloud of points in the EDP vs IM space. The selected EDP in this study is the ductility demand μ . The selected IM is the pseudo-spectral acceleration at the SDoF period, normalised to the yield base shear coefficient, $R = SA/f_y$. Although more advanced IMs are available (e.g., ref. 19), such a simple choice may simplify the hazard analysis in practical applications or even exploit existing hazard models.

A set of 100 natural (i.e., recorded) ground motions is selected from the SIMBAD database, “Selected Input Motions for displacement-Based Assessment and Design.”²⁰ As per ref. 13, the three-component 467 records in the database are ranked according to their peak ground acceleration, PGA, values (by using the geometric mean of the two horizontal components) and then keeping the horizontal component with the largest PGA value. The first 100 records are arbitrarily selected, characterised by moment magnitudes in the range 5–7.3, station-to-source distance smaller than 35 km; and PGA ranging between 0.29 and 1.77 g. Such an ground-motion selection approach is consistent with the adopted response analysis method (i.e., cloud analysis). Hazard-consistent, site-specific record selection is outside the scope of this work since the aim is to calibrate a flexible metamodel of the PSDM for a large number of hazard/site conditions.

NLTHAs are conducted scaling the records to ensure non-linear response ($\mu > 1$) for every case study (which show a wide range of period of vibration and yield strength), considering 100 guesses target values of ductility demand, equally spaced between 1 and 6. The equal displacement rule ($\mu \approx R$) is used to derive reasonable guesses of the scale factors (SF): (1) a record and a guess ductility demand are randomly selected (without repetition) from the relevant sets; (2) the scale factor is calculated assuming that the scaled spectral acceleration is equal to $f_y \mu_{guess}$, and using the spectral acceleration for the unscaled record at the SDoF period. The adopted scale factors range between 0.20 and 5.6, keeping the bias introduced in the response analyses to acceptable levels.²¹ It is worth mentioning that a recent study²² suggests that this level of scaling might create bias in estimating collapse probability. Therefore, for some specific applications, this aspect should be further checked and possibly corrected by using procedures available in the literature (e.g., ref. 23).

The ductility demand of each analysis is checked against the guessed value. Analysis results leading to a ductility demand outside of the target range are excluded from the PSDM fit described below, ensuring that at least 70 records are adopted for each cloud.

The adopted bi-linear PSDM (Figure 1, step 2) is defined according to Equation 1. Bi-linear PSDMs have been demonstrated effective in previous work (e.g., ref. 24). Clearly, the PSDM in Equation 1 does not imply the equal displacement rule. Characterising the behaviour in the elastic range is trivial and follows from the definition of an elastic SDoF ($\mu = R$), not requiring any analysis for its calibration. The inelastic range is obtained performing a linear regression in the logarithmic space, where $\sigma_{\ln(\mu-1)|R-1}$ (henceforth simply called σ) is the logarithmic standard deviation of the pairs $\mu - 1$ versus $R - 1$, and ε is a standard normal variable. Therefore, the median relationship is the line $\mu = a(R - 1) + 1$, where a is its slope. Such a model choice, compatible with other literature studies (e.g., ref. 7), implies a lognormal distribution of the residuals, which is desirable in calculating lognormal fragility curves (see Section 2.4). Such model also implies homoscedasticity for $\mu > 1$. As demonstrated previously,¹² this assumption is deemed satisfactory for a variety of engineering systems (including SDoF systems).

$$\begin{cases} \mu = R & \mu \leq 1 \\ \ln(\mu - 1) = \ln(a) + \ln(R - 1) + \varepsilon\sigma_{\ln(\mu-1)|R-1} & \mu > 1 \end{cases} \quad (1)$$

Depending on the considered practical applications, the SDoF parameters may be affected by a degree of variability. Although this is out of the scope of this paper, both the provided GP regressions and/or the training data can be used to appropriately account for the effect of such variability on the predicted seismic response (e.g., ref. 25).

2.3 | Gaussian process regressions for the $(hyst, T, f_y, h) - (a, \sigma)$ map

For each SDoF realisation depending on a vector of input parameters $\mathbf{x} = \{hyst, T, f_y, h\}^T$, the methodology in Section 2.2 allows defining two PSDM output parameters (a, σ) . Two independent training datasets, (X, \mathbf{y}) , are composed by the matrix X collecting the input vectors for all the SDoF realisations, and the vector \mathbf{y} which includes the related a or σ outputs. Based on such training datasets, a GP regression is fitted for each PSDM parameter. This takes a vector of unique inputs (\mathbf{x}) and produces an output/target $y = f(\mathbf{x})$ using a statistical model, while being computationally cheaper than the methodology shown in Section 2.2. Based on a training dataset made of inputs (covariates) and known outputs, a GP regression is fitted so that it is possible to make predictions for any input vector outside the training dataset.

Although a detailed mathematical description of GP regressions (and their fitting) is outside the scope of this paper, it is worth providing a general, high-level perspective on the matter. Rasmussen and Williams¹⁸ present an exhaustive mathematical description/derivation of GP regression. A step-by-step mathematical description for an earthquake engineering application is provided, for example, ref. 13

In a GP regression, $f(\mathbf{x})$ is regarded as a realised value of a GP. A GP is a generalisation of the Gaussian probability distribution model, describing the distribution of functions $f(\mathbf{x})$ rather than scalars or vectors. Its mean and covariance functions fully specify a GP. According to a Bayesian framework, the first step in a GP regression is to set a prior distribution for all the possible functions $f(\mathbf{x})$, reflecting the starting knowledge about the output before having any data. This is done by assigning some properties of the mean and the covariance functions (e.g., smoothness). Then, the prior distribution is converted into a posterior distribution (over functions) based on the observed data, and such a posterior distribution is used for predictions. The properties of the output function $f(\mathbf{x})$ —with particular reference to its smoothness—are governed by the covariance function, which captures the correlation among different input vectors and reflects it in the output. The training of a generic, scalar-input function is shown in Figure 2.

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 SDoF parameters produces a slight change in the output PSDM parameters). The covariance function parameters are called hyperparameters since they are not assigned (a GP regression is a non-parametric model) but inferred from the training dataset (usually with a maximum likelihood estimator). Given the hyperparameters, the posterior predictive distribution of the GP is obtained by conditioning the prior distribution to the training dataset. The expected value of the posterior distribution is adopted to make predictions.

With regard to the GP regression fitting for this particular study, a constant basis function is assumed for the posterior distribution of the mean function. In addition, a squared exponential covariance function with separate length scales is

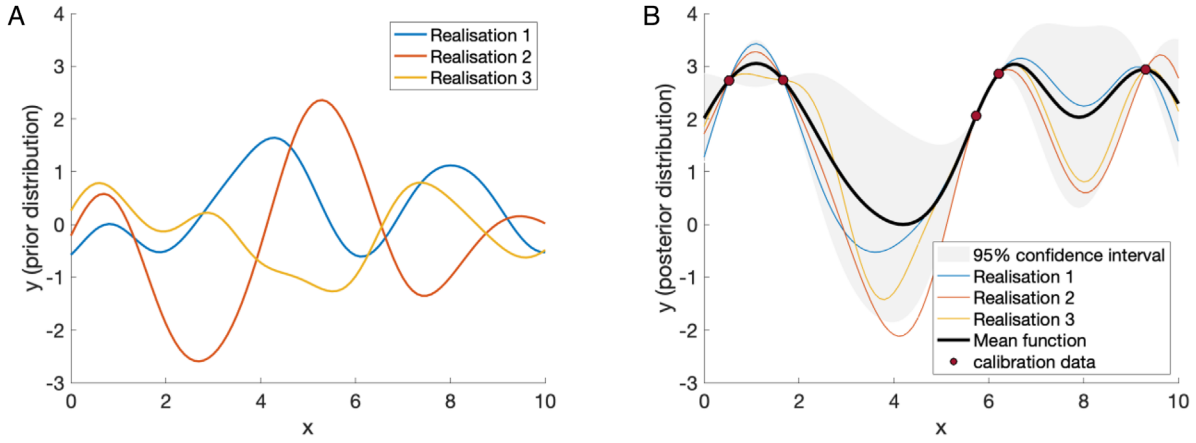


FIGURE 2 Example of Gaussian process regression for a generic $y = f(x)$ relationship. (A) Prior distribution; (B) posterior distribution

adopted. This is done to have a reasonable amount of hyperparameters while reflecting the mechanics of the phenomenon under investigation. A quasi-Newton method²⁶ is adopted to optimise the hyperparameters, including the noise variance.

2.4 | Seismic fragility and loss analysis

Once the PSDM parameters (a, σ) are evaluated using the trained GP regressions, it is possible to derive fragility functions to perform seismic risk and/or loss analysis. Building-level fragility curves are calculated for a set of structure-specific DSs, identified by the thresholds EDP_{DS} . One possibility involves choosing four DSs, slight, moderate, extensive and complete damage, as defined according to HAZUS.⁹ Other definitions of the DSs are possible,^{11,27} and the proposed framework is independent of their particular choice.

According to the properties of the adopted PSDM (Equation 1), lognormal fragility curves for each DS, representing the DSs' exceeding probability, $F_{DS_i} = P(\mu \geq \mu_{DS_i} R)$, are completely specified by their median η_{DS} and logarithmic standard deviation β (simply called dispersion), which are specified in Equation 2 both for the elastic and inelastic ranges. It is worth mentioning that collapse cases (corresponding to ground motions leading to dynamic instability of the analysis or exceedance of a conventional 10% drift threshold) are not expected since the SDoF systems are only subjected to pre-peak ductility demand levels.

$$\mu_{DS_i} \leq 1; \begin{cases} \eta_{DS_i} = \mu_{DS_i} \\ \beta = 0 \end{cases} \quad \mu_{DS_i} > 1; \begin{cases} \eta_{DS_i} = \frac{\mu_{DS_i}^{-1}}{a} + 1 \\ \beta = \sigma \end{cases} \quad (2)$$

Loss analysis is based on vulnerability curves, which can be derived using a building-level consequence model relating the repair-to-replacement cost to structural and non-structural DSs. This involves defining the expected building-level damage-to-loss ratios ($DLRs$) for each DS. The (mean) loss ratio (LR) for a given IM value is defined according to Equation 3, allowing deriving a mean vulnerability curve (involving the difference between the exceeding probability of the $(i+1)^{th}$ and i^{th} DSs). More advanced, component-based fragility/loss methodologies are available (e.g., ref. 28), which are appropriate for more refined applications, such as the detailed seismic loss estimation of individual buildings, for instance.

$$LR(IM) = \sum_{i=1}^4 (F_{DS_{i-1}}(IM) - F_{DS_i}(IM)) DLR_{DS_i} \quad (3)$$

One of the relevant loss metrics commonly adopted in practice is the EAL. Assuming a hazard curve, this is calculated according to Equation 3, where λ_{IM} is the mean annual frequency of exceeding a given value of the IM.

$$EAL = \int_0^{\infty} LR(IM) \left| \frac{d\lambda_{IM}}{dIM} \right| dIM \quad (4)$$

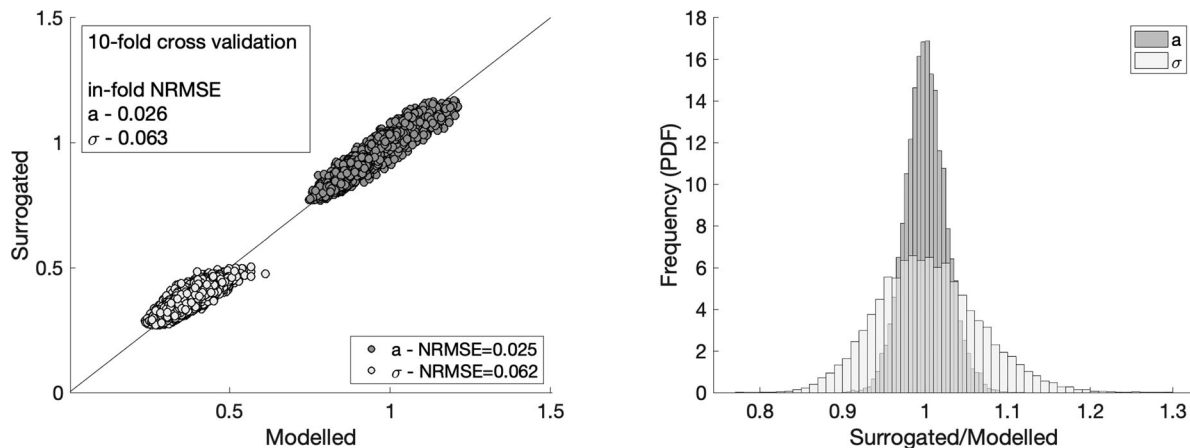


FIGURE 3 PSDM parameters: (A) Surrogated (GP regression) versus modelled (SDoF cloud analysis) plot; (B) Surrogated versus modelled ratio histograms. NRMSE: normalised root mean squared error

3 | RESULTS AND DISCUSSION

3.1 | Surrogated-vs-modelled error and cross-validation of the GP regressions

The trained GP regressions are subjected to various tests to investigate their predictive power. First, the PSDM parameters are predicted for the entire SDoF database, measuring the surrogate-vs-modelled normalised root mean squared error, $\text{NRMSE} \left(\frac{\text{mean}(\sqrt{(s_i - m_i)^2})}{\text{mean}(m_i)} \right)$. Figure 3A shows that the surrogated-vs-modelled points are particularly close to the $y = x$ line, with the NRMSE being equal to 2.5% for the slope and 6.2% for the logarithmic standard deviation of the PSDMs. Considering the various sources of uncertainty commonly involved in seismic performance assessment or risk/loss models, such error levels are deemed satisfactory for practical applications.

To test the predictive power of the GP regressions outside the training set, ten-fold cross validation is carried out. This first involves randomly dividing the training dataset into ten subsets made of 1,000 SDoF samples each. Then, ten new GP regressions are fitted alternatively, excluding one subset from the new training dataset. For each subset, the fitted GP regressions are used to make (in-fold) predictions for the subset kept out of the training dataset. The in-fold prediction errors for each of the ten sub-sets are used to calculate the in-fold NRMSE, which is 0.1% higher than the previously calculated one for both PSDM parameters, thus confirming the validity of the error estimates. It is worth mentioning that such error estimates are consistent with the results in ref. 7, which, for the case studies most similar to those in this paper, report a 3%–6% error on the PSDM estimation (although those are based on 30 ground-motion records).

As expected for GP regressions, Figure 3B shows that the prediction errors (herein quantified as surrogated-vs-modelled ratios) follow a Normal distribution with a mean equal to 1.0001 and 1.0002 (for a and σ) and standard deviation essentially equivalent to the NRMSE. This result is particularly important to infer the propagation of such error in fragility or risk/loss estimations, as discussed below in this Section.

The surrogated PSDMs for the entire SDoF dataset are used to estimate seismic fragility curves. Those are derived for two DSs related to ductility thresholds, respectively, equal to 3 and 4, representing DS3 (extensive damage according to HAZUS) and DS4 (complete damage according to HAZUS). Elastic DSs are not considered since, according to Equation 2, no error is expected for those. Figure 4A shows the surrogated vs modelled plot for the median of the fragility curves, also indicating a satisfactory NRMSE, respectively equal to 2.0% and 2.2% for DS3 and DS4. Consistently with the definition of the PSDM (Figure 1 and Equation 1), a specified error in the slope parameter a propagates to higher error levels for the fragility median as the ductility threshold increases. It is worth mentioning that no further discussion is needed for the fragility dispersion since this is exactly equal to the PSDM logarithmic standard deviation. Also in this case, as expected, the surrogated-vs-modelled ratios for the median of the fragility curves (Figure 4B) follow a Normal distribution (with mean equal to 0.9993 and standard deviations equal to 0.018 and 0.02 for DS3 and DS4, respectively). Quantifying how the uncertainty on the SDoF parameters affects the prediction of the PSDM and fragility ones is considered out of scope herein. However, after characterising the probability distribution of the SDoF parameters (and their correlation), a user can adopt the proposed GP regressions to carry out the above uncertainty quantification.

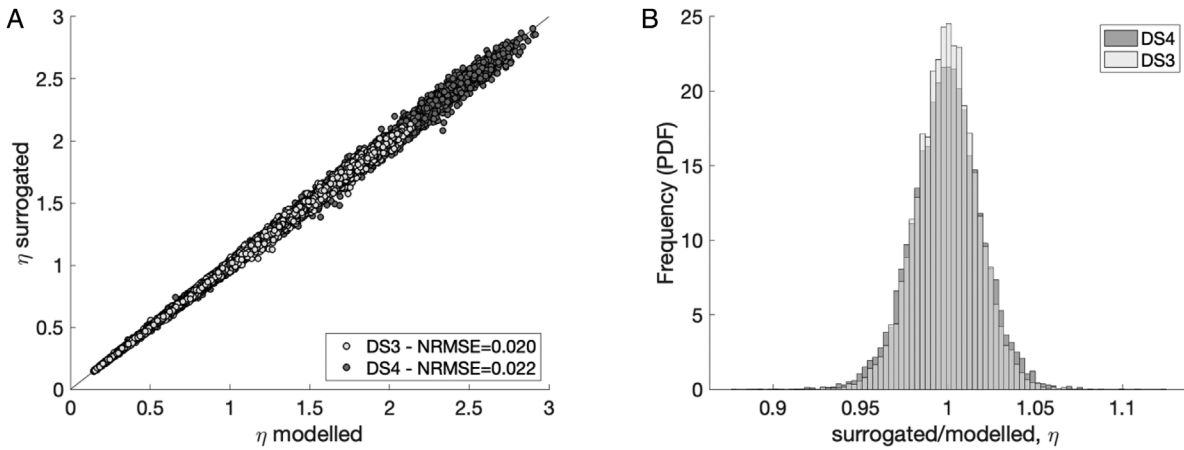


FIGURE 4 DS3-DS4 fragility medians: (A) surrogated (GP regression) versus modelled (SDoF cloud analysis) plot; (B) Surrogated versus modelled ratio histograms. NRMSE: normalised root mean squared error

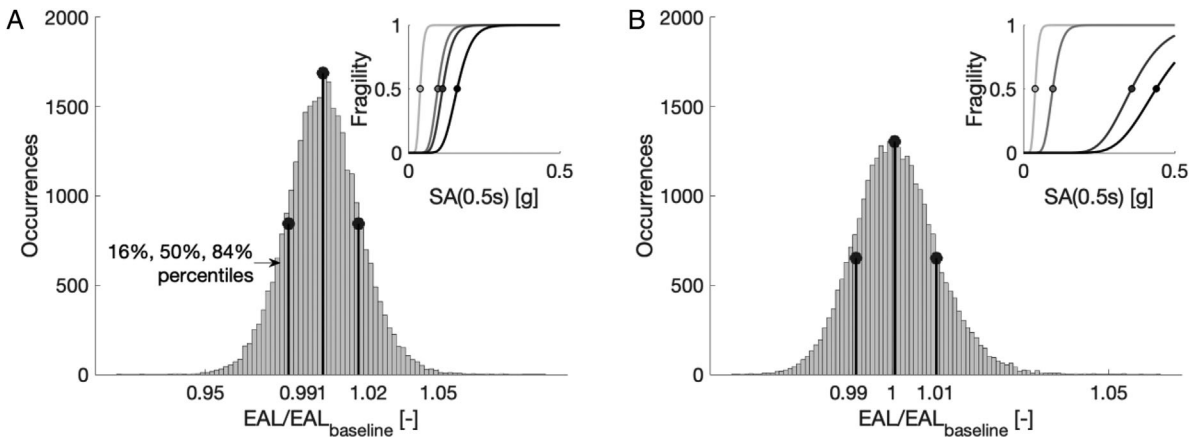


FIGURE 5 Propagation of fragility error on the calculated expected annual loss

The last step of this error-propagation exercise involves calculating both vulnerability curve and EAL for two illustrative (arbitrary) case-study SDoF systems with an elastic period equal to 0.5s. Once again, both systems are characterised according to the four structure-specific DSs qualitatively defined according to HAZUS and quantitatively measured on the SDoFs backbones. The median of the fragilities for the elastic-range DSs is the same for both case studies and is equal to 0.04 g and 0.1 g for DS1 and DS2, respectively. The first case study represents a structure with poor seismic performance; thus, the DS3 and DS4 medians are equal to 0.12 g and 0.16 g, respectively. For the second case study, showing considerably higher performance, the DS3 and DS4 medians are equal to 0.36 g and 0.44 g, respectively. A fragility dispersion equal to 0.24 is adopted for all DSs (both elastic and inelastic) for both case studies. A detailed motivation to assume the same dispersion for both the elastic and inelastic ranges is given in Section 3.2.

DLRs equal to 2%, 10%, 43.5%, and 100% of the total reconstruction cost (DS1 to DS4) are adopted for this example, consistently with the assumptions in ref. 13. Vulnerability curves are calculated according to Equation 3. Finally, the EAL is calculated according to Equation 4, assuming the hazard curve appropriate for L'Aquila, a high-seismicity town in central Italy. Such a curve is consistent with the official Italian hazard model (Stucchi et al.²⁹), implemented in the current Italian building code.³⁰ Although the absolute values of the hazard curve are deemed less relevant for this error-propagation exercise, the adopted hazard model is consistent with those used in Section 4.1.

Starting from the “baseline” fragility curves described above, a plain Monte Carlo approach is used to simulate 30,000 new realisations of the fragility parameters such that $\eta_{sim} = Err_{\eta} * \eta_{baseline}$, $\beta_{sim} = Err_{\beta} * \beta_{baseline}$. The variables Err_{η} , Err_{β} are consistent with the empirical distributions of the surrogated-vs-modelled ratios shown in Figure 4B (for the medians) and Figure 3B (for the dispersion). By re-calculating both the vulnerability for each realisation, the fragility error is propagated to the EAL. For both case studies, Figure 5 shows the empirical distribution of the EAL as a ratio of the baseline EAL (the insertions in each panel show the baseline fragility curves for the case studies). The overall distribution

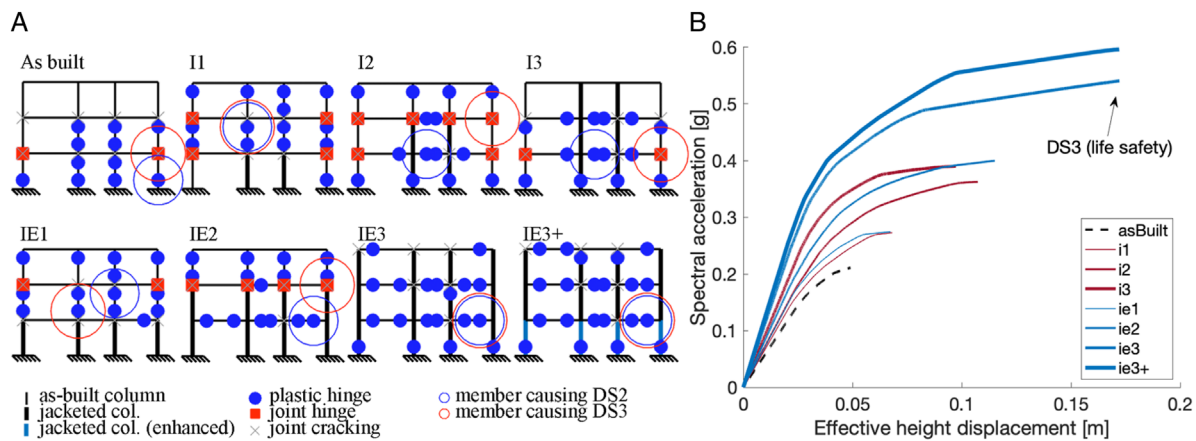


FIGURE 6 Selected incremental retrofit solutions: (A) plastic mechanism at the life-safety damage state, DS3; (B) pushover curves (modified after Gentile et al.³²)

of the error is a particularly narrow Normal distribution with a mean respectively equal to 1.0007 and 1.0006 and standard deviation equal to 0.015 and 0.01 for the lower- and higher-performance case studies, respectively. In the Authors' opinion, such low error levels confirm the suitability of the proposed surrogate model in loss assessments involving SDoF-based fragility curves and building-level damage-to-loss ratios (e.g., ref. 2).

3.2 | Application to realistic RC frame case studies

Most practical applications involve more complex MDoF structures rather than SDoF systems. Therefore, it is crucial to test the predictive power of the trained GP regression against realistic structures. The selected case study is the central longitudinal frame of a three-storey RC building with rectangular plan geometry and structural details of beams, columns and joints consistent with pre-1976 design according to an older Italian building code.³¹ Apart from this as-built configuration, consistent with gravity-only design and neglecting any capacity-design provision, seven retrofitted configurations providing incremental seismic performance are designed implementing the RC jacketing technique (Figure 6). Detailed descriptions and illustrations of geometry, load analysis and structural details of the as-built and retrofitted members are provided in ref. 32. They are not repeated here for brevity.

A 2D lumped-plasticity model (bare frame) is developed using the finite element software Ruaumoko³³ for each configuration. The adopted numerical modelling strategy was extensively validated against experimental results.³⁴ 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 since they are deemed negligible for three-storey frames. A 5% tangent stiffness-proportional elastic damping is assigned to all frequencies. 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 lateral member response. The capacity model of the members included in the analysis is modified (if appropriate) to include those failure mechanisms accordingly, considering slab-related flange effect for the negative beam moment capacity, lap splice failure, shear capacity, bar buckling.¹³ The modified Takeda hysteresis model¹⁶ 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,¹⁶ which can capture their pinching behaviour. Each frame configuration is first analysed via displacement-control pushover with a linear force profile (Figure 6). This allows quantifying four displacement thresholds (Δ_{DSi}) compatible with the DS definitions in HAZUS (described in Section 2.4), to be adopted for fragility derivation. Those refer to the first member in the frame reaching first cracking, yielding, $\frac{3}{4}$ of the near collapse drift and the near collapse drift respectively for DS1-DS4. It is worth mentioning that the member "causing" a given DS may change for different DSs.

The as-built configuration shows a storey-level failure mode developing at the first storey (Figure 6A), characterised by plastic hinging of the columns and shear failure of the exterior beam-column joints. Although the shear failure of the exterior joints "avoids" the soft-storey behaviour, this failure mode results in a low strength (Figure 6A) approximately equal to 0.2 g spectral acceleration capacity for the equivalent SDoF system, whose elastic period is 0.77s. The acceleration-displacement capacity (Figure 6B) of the equivalent system is derived from the pushover force-displacement curve assuming a (first-mode) effective mass equal to 90% of the total building mass (obtained averaging the participating mass of the

TABLE 1 Ductility-based DS thresholds

	μ_{DS_1}	μ_{DS_2}	μ_{DS_3}	μ_{DS_4}
As built	0.54	1.00	1.05	1.40
I1	0.46	1.00	1.25	1.67
I2	0.55	1.00	2.36	3.14
I3	0.71	1.00	2.76	3.68
IE1	0.52	1.00	1.40	1.86
IE2	0.58	1.00	2.65	3.54
IE3	0.65	1.00	4.45	5.93
IE3+	0.66	1.00	4.54	6.05

case studies, ranging between 88% and 93%). Moreover, the effective height displacement is adopted as an EDP, using the effective height formulation by Priestley et al.¹⁷

The adopted retrofit strategy has the objective of inverting the local hierarchy of strength at the sub-assembly level and transforming such a localised failure mode into a more reliable Beam-Sway global mechanism (with flexural plastic hinges forming in all beam ends and at the ground section of the columns). Such a goal is obtained “incrementally,” adopting concrete column jacketing as the selected retrofit technique. This results in seven retrofit solutions, for which Figure 6A shows the plastic mechanism at DS3 (extensive damage according to HAZUS; generally associated with life safety). For the solutions I1, I2 and I3, only the interior columns are jacketed (respectively up to the first, second or third storey). Similarly, the solutions IE1, IE2, and IE3 include column jacketing for both interior and exterior columns. Finally, the IE3+ retrofit solution improves IE3 by involving enhanced jacketing for the first-storey columns to provide higher strength for the frame. It is worth mentioning that the beam-column joints located between two jacketed columns are reinforced with horizontal stirrups having the same layout as the jacketed columns, thus significantly enhancing their shear capacity (and avoiding shear hinging).

The response of each frame configuration is independently analysed using two analysis methods with increased refinement. The cloud capacity spectrum method (Cloud CSM,³⁵) is first adopted: this involves applying the CSM,³⁶ assuming a suite of as-recorded ground-motion spectra. In this case, the set of 100 records described in Section 2.2 is adopted without using any scale factors. The response (effective height displacement) for each ground motion is calculated by using the pushover curves in Figure 6B. Moreover, by adopting the same record suite, cloud-based NLTHA¹² is carried out for each frame configuration, registering the peak effective height displacement. The spectral acceleration at the fundamental period, $SA(T_1)$, is finally calculated for each record and adopted as an IM. T_1 ranges between 0.77s for the as built configuration and 0.55s for the IE3+ one.

The two alternative (EDP vs IM) sets are adopted to derive the conditional mean and standard deviation of EDP given IM and derive PSDMs for each frame configuration. Consistently with the common practice, the power-law model $EDP = p_1 IM^{p_2}$ is obtained via the least square method, where p_1 and p_2 are the parameters of the regression. This

allows deriving fragility curves compatible with those derived using the GPs, where $\eta_{DSi} = \exp\left(\frac{\ln(\frac{\Delta_{DSi}}{p_1})}{p_2}\right)$ and $\beta = \frac{\sigma}{p_2}$. It is worth mentioning that some ground motions led to collapse, which is herein defined as 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 the nominal threshold of 10% maximum inter-storey drift. The information carried out in these analyses with non-numerical EDP values are included in the definition of the fragility curves according to the procedure in ref. 37, which is described in detail in ref. 13.

The parameters of the fragility curves are finally computed using the trained GP regressions, after bi-linearising the pushover curves,³⁸ adopting the MTF hysteresis type (deemed appropriate for RC frames not developing a soft-storey mechanism), and using DS ductility thresholds consistent with the above-mentioned Δ_{DSi} (Table 1). Based on the GP results, literature formulations can be used to calculate other EDPs such as displacement profiles¹⁷ or peak floor accelerations.²⁸

Before comparing fragility estimates of a surrogated SDoF model and an explicit MDoF model, it is worth highlighting the various sources of error affecting it. First, the two models involve a different functional form of the PSDM, respectively suited for SDoF and MDoF models. Moreover, the comparison adopts the spectral acceleration at the fundamental period as IM, although more sufficient and efficient IMs are available (e.g., ref. 19). Moreover, although the same set of records are used, those are scaled in amplitude in the metamodel training, possibly leading to further bias.²¹

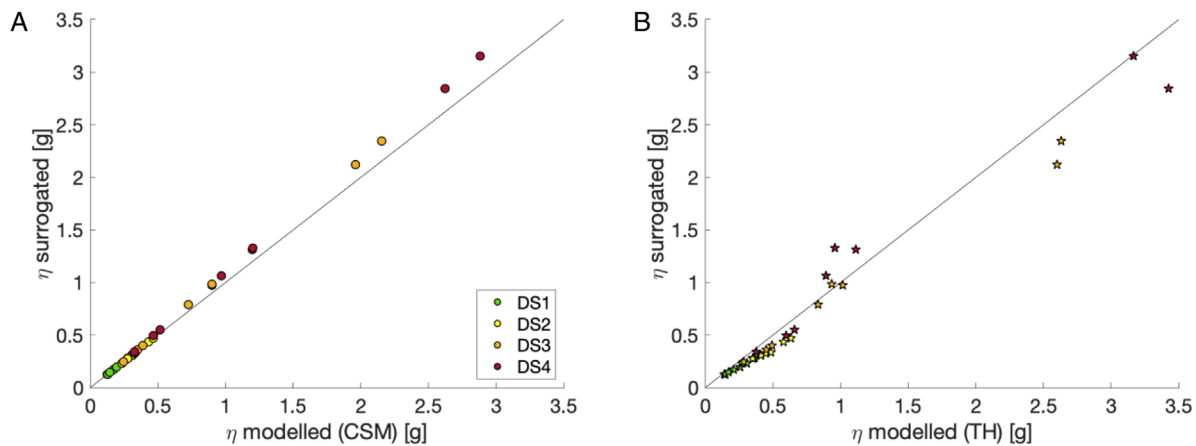


FIGURE 7 Surrogated versus modelled fragility median for the MDoF case study. Modelling is alternatively based on (A) capacity spectrum method and (B) non-linear time history analysis

As shown in Figure 7A, comparing the surrogated medians with those calculated via the Cloud CSM highlights a very low prediction error: the NRMSE is equal to 1.0%, 0.6%, 10.4%, and 11.2% for DS1, DS2, DS3, and DS4, respectively. Such error level increase to 24.0%, 26.0%, 17.7%, and 19.0% if the benchmark analysis method is the NLTHA of the MDoF numerical models (Figure 7B). Although such an increase is substantial, it is still deemed consistent with the simplicity of the GP metamodel as opposed to the required modelling effort and time involved in the MDoF NLTHA.

From a qualitative point of view, the GP regressions provide a conservative underestimation of the medians (with respect to the NLTHA), except for the DS4 estimation for three case studies. Consequently, the GP regressions would provide conservative overestimations in risk/loss analyses. Confirming the discussion in Section 3.1, the error levels are generally higher for severe DSs and higher-performing structures (higher fragility medians). However, in risk/loss estimations, such error levels will likely have a low impact given the lower hazard frequency related to high intensity (e.g., in Equation 4 they will be weighted down by the lower λ_{IM} values as IM increases). This effect, valid for any combination of building models and sites, is further discussed in Section 4.1, also showing an example.

On the other hand, the GP regressions show a lower predictive power in estimating the fragility dispersion. In fact, the fragility dispersion predicted with the GP regressions for the non-linear range lies within 0.35 and 0.4 for all the considered frame configurations. Despite an NRMSE equal to 23.6% and 29.0%, respectively, assuming the Cloud CSM and the NLTHA as benchmark models, in both cases, the error for the single datapoints approximately ranges between 9% and 59%, and it is generally an underestimation.

It is worth highlighting that the fragility dispersion, in turn depending on the EDPIIM standard deviation, is strongly affected by the refinement of the model (e.g., a higher dispersion is expectable for MDoFs than for SDoF systems, e.g., due to higher-mode effects and lower sufficiency of the IMs). Accordingly, the adopted PSDM model for MDoFs allows accounting for the EDPIIM uncertainty also in the elastic range, while this is not present in the SDoF response. To partially compensate for this deficiency, it is suggested to use the GP-based fragility dispersion also for the elastic DSs (which would have $\beta = 0$ according to Equation 2). On the other hand, considering collapse cases in the MDoF model usually reduces the fragility dispersion, especially if the structure develops a plastic mechanism for low seismic intensity levels. The combination of such effects, not adequately captured in the GP regression approach, is case-dependent, and it is arguably challenging to predict a general trend.

For this reason, although the fragility dispersion errors are numerically higher than those of the medians, those are still comparable to those obtained explicitly modelling an SDoF system (arguably the standard for portfolio analyses) subjected to cloud-based NLTHA.

4 | EXPLOITING THE PROPOSED SURROGATE MODEL

4.1 | Direct loss-based seismic design: a first attempt

The first application relying on the proposed surrogate model is the DLBD. A *first attempt* DLBD procedure, among other possible alternatives (e.g., refs. 39–41), is proposed in this Section, along with an application to a realistic RC frame

structure analysed along one horizontal direction. This procedure is not intended as definitive, and future research work is suggested, and underway, to refine and validate it, possibly modifying some of its steps. By analogy with the words of Priestley,¹⁷ the fundamental philosophy of the proposed procedure is to design a structure that would achieve, rather than be bounded by, a given loss-related metric (e.g., EAL) under the relevant site-specific seismic hazard. Moreover, the adjective “direct” refers to the ability to specify the desired level of loss as a fundamental input parameter before performing all the relevant steps to achieve such loss level with a reasonable tolerance.

The loss estimation module in this procedure is based on building-level vulnerability curves defined in terms of IM (Section 2.4). Although more refined approaches with vulnerability defined in terms of EDP are available, possibly at a component-based level, the selected approach is arguably deemed more appropriate for a preliminary design phase. For example, an EDP-based loss estimation would allow explicitly considering acceleration-sensitive non-structural components. However, since their exact typology and number is most likely not known in the preliminary design phase, it is preferred to make a more generic (and flexible) assumption to embed them in the definition of the DLRs. Moreover, assumptions to include indirect losses can be embedded in the definition of DLRs, rather than needing to add specific vulnerability functions.

The proposed DLBD can be outlined in the following steps, selecting the EAL as the fundamental loss metric for design:

1. Select EAL_{target} (e.g., 0.5% of the total reconstruction cost). This is the level of EAL that the designed structure will achieve;
2. Retrieve an appropriate set of site-specific hazard curves in terms of SA in a wide range of periods. Also, select a set of DSs (e.g., DS1-DS4 as defined in Section 2.4) relevant to the considered structural typology (e.g., RC frames). Relatively to the ductility capacity at peak strength μ_{cap} , which is an intermediate design parameter calculated in step 5, select reasonable guesses for the DS thresholds consistent with the qualitative definition of the DSs (e.g., $\mu_{DSi} = [0.5 \ 1 \ 0.75\mu_{cap} \ \mu_{cap}]$ may be consistent with the DS definition above). The main idea is to provide μ_{DSi} values as close as possible to those obtainable from the numerical pushover analysis of the final design case. Select a relevant damage-to-loss model consistent with the involved structural typology and the adopted DSs (e.g., $DL \ R_{DSi} = [7 \ 15 \ 50 \ 100]$ % of the total reconstruction cost);
3. Select the basic geometric and material properties of the considered structure, also according to gravity-load preliminary design. For a geometrically-regular RC frame building, the parameters to be (tentatively) set are the yield stress and elastic modulus of steel (f_{sy}, E_s), the number of storeys and bays (N_{storey}, N_{bay}), the centre-to-centre inter-storey height H_{int} , the centre-to-centre length of the beams l_b , the depth of beams and columns (h_b, h_c), the seismic storey mass m_i directly affecting the considered frame;
4. Select a number of combinations of yield strength and ductility capacity to define *seed SDoF systems*. For example, those can be N_1 equally-spaced points within $f_{y,min}$ and $f_{y,max}$ and N_2 equally-spaced points between $\mu_{cap,min}$ and $\mu_{cap,max}$, leading to N_1N_2 seed SDoF systems. The range for such parameters should be carefully selected based on engineering judgement, to ensure including possible design cases. According to the relevant structural type, complete the definition of the seed SDoF systems consistent with the GP regression inputs (i.e., bilinear backbone and expected hysteresis type). For RC frames, for example, the yield displacement (Δ_y) only depends on geometry,¹⁷ and it can be calculated according to Equation 5, where H_i is the height of storey i from the ground and δ_i is the displacement shape ($\delta_{roof} = 1$) assumed according to Priestley et al.¹⁷ Therefore, yield strength ($f_{y,seed}$) is linearly related to stiffness, and the elastic period (T_{seed}) of the seed SDoF systems is calculated via Eq. 6. The hardening ratio of the seed SDoF systems can be reasonably guessed (e.g., $h = 0.05$ for RC frames) or alternatively included in the initial grid-based definition. Therefore, the peak normalised based shear of the SDoF system is equal to $f_{peak} = f_y (1 + \mu_{cap}h)$. Finally select the most appropriate hysteresis model for the analysed structure (e.g., MTF for newly-designed RC frames);

$$\Delta_y = \theta_y H_{eff} \quad \text{where} \quad \theta_y = \frac{0.5 \frac{f_{sy}}{E_s} (l_b - h_c)}{h_b} \quad \text{and} \quad H_{eff} = \frac{\sum_i m_i \delta_i H_i}{\sum_i m_i \delta_i} \quad (5)$$

$$T_{seed} = 2\pi \sqrt{\frac{\Delta_y}{9.81 f_{y,seed}}} \quad (6)$$

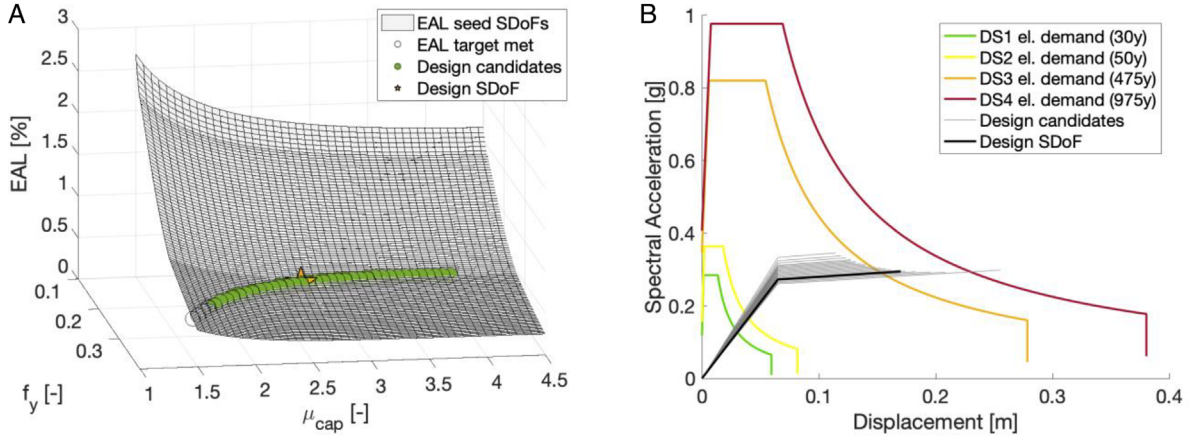


FIGURE 8 (A) Map of the expected annual loss for the trial SDoF configurations; (B) capacity spectra of the candidate design SDoF systems

5. Using the herein-proposed GP regressions, calculate fragility curves for each seed SDoF system consistent with the selected DS thresholds μ_{DSi} . According to the selected DLR_{DSi} and the hazard curve (interpolated based on T_{seed}), use Equations 3 and 4 to calculate the building-level vulnerability curve and the EAL for each seed (see Figure 8A);
6. Select all the SDoF seeds that meet the target EAL level within a set tolerance (e.g., $EA L_{tol} = 0.01EA L_{target}$). Clearly, the finer the (f_y, μ_{cap}) grid is, the smaller such tolerance can be. For each selected seed, run the CSM using spectra for each DS demand (code-based, as per Figure 8B, or site-specific ones) and disregard the cases not meeting the seismic demand for any DS (i.e., $\mu_{dem,DSi} > \mu_{DSi}$). In addition, calculate the frequency of exceeding the complete damage DS, by integrating the complete damage fragility with λ_{IM} , analogously to Equation 4, and disregard the cases above a conventionally-established threshold (e.g., between 10^{-5} and 10^{-4})⁴². The seed SDoF systems meeting the target EAL, the complete damage exceeding frequency bound and complying with the code-based displacement check are equally valid *candidate design SDoF systems*. One of those can be arbitrarily selected as the final *design SDoF* system, possibly according to design requirements not related to seismic actions;
7. Among many other alternatives, the principles of displacement-based design are herein suggested to detail each member of the structure to comply with the design SDoF's backbone and achieve a favourable plastic mechanism (e.g., plastic hinges for the base column sections and all beams' end sections, for an RC frame). For a geometrically-regular RC frame, the beams' plastic drift demand ($\theta_{bi,p}$) compatible with the displacement capacity at peak force (Δ_{cap}) of the design SDoF is calculated according to Equations 7 and 8, where θ_i is the inter-storey drift profile compatible with the displacement profile Δ_i . On the other hand, the drift demand on the first-storey columns (θ_c) is simply equal to the inter-storey drift at that level. Such an approach involves a degree of error with respect to more refined, NLTHA-based approaches. However, pros and cons of displacement-based design, including the calibration of all the involved equations and assumptions are provided in ref. 17;

$$\theta_{bi,p} = \theta_{i,p} / (1 - h_c/l_b). \quad \text{Where} \quad \theta_{i,p} = \theta_i - \theta_y \quad (7)$$

$$\Delta_i = \frac{\Delta_{cap}}{\delta_{eff}} \delta_i. \quad \text{Where} \quad \delta_{eff} = \frac{\sum_i m_i \delta_i^2}{\sum_i m_i \delta_i} \quad (8)$$

8. The strength demand for the members developing inelastic behaviour is computed via an equilibrium approach. For RC frames, this involves the overturning moment (OTM) equilibrium,¹⁷ according to Equation 9, where $V_B(\Delta_{cap}) = f_{peak} m_{eff}$ is the SDoF base shear, $m_{eff} = \frac{\sum_i m_i \Delta_i}{\Delta_{cap}}$ is the effective mass, $M_{c,k}$ is the base moment of column k , L_{build} is the frame length, $V_{b,i,j}$ is the shear of the beam at storey i and bay j . Any allocation of strength to the members developing inelastic behaviour that complies with Equation 11 would satisfy the design objective. One possible design choice may involve assuming (Equation 10) that the contra-flexure point of the first-storey columns locates at 60% of their height so that capacity design of the first-storey beam-column joints is ensured. Moreover, if the beams have the same detailing in both end sections and within the frame, Equation 11 holds, where $M_{b,i,j}(\theta_{bi,p})$ is the beam moment

demand corresponding to its plastic drift demand. With such assumptions, it is possible to invert Equation 9 and calculate the beams' moment demand $M_{b,ij}$ corresponding to the target base shear $V_B(\Delta_{cap})$. The moment demand of the base columns, at $V_B(\Delta_{cap})$, can be reasonably proportioned assuming that the interior columns will be approximately twice as strong as the exterior ones, thus $M_{c,ext} = 0.6H_1 \frac{V_B(\Delta_{cap})}{2(N_{bay}-1)+2}$ and $M_{c,int} = 2M_{c,ext}$.

$$OTM(\Delta_{cap}) = V_B(\Delta_{cap}) H_{eff} = \sum_k M_{c,k}(\theta_{c,k}) + L_{build} \sum_{i,j} V_{b,ij}(\theta_{bi,p}) \quad (9)$$

$$\sum_k M_{c,k}(\theta_{c,k}) = \sum_j 0.6H_1 V_{c,k} = 0.6H_1 V_B(\Delta_{cap}) \quad (10)$$

$$V_{b,ij}(\theta_{bi,p}) = \frac{2M_{b,ij}(\theta_{bi,p})}{l_b} \quad (11)$$

9. Once the theoretical deformation and strength demand of the members developing inelastic behaviour are obtained, the structural details of such members are designed. For an RC frame, this involves the demand for beams ($\theta_{bi,p}; M_{b,ij}$) and first-storey columns ($\theta_{c,k}; M_{c,k}$). The structural details of such members can be designed via a moment-curvature approach. Generally, the provided detailing will lead to a degree of deviation from the theoretical values above. Thus, Equation 9 can be used to re-calculate the peak base shear $V_B(\Delta_{cap})$ provided by the design. Moreover, with a reasonable assumption for the hardening (e.g., $h = 0.05$ for RC frames), and by knowing Δ_y , the entire backbone can be calculated and compared to that of the design SDoF system (Figure 9A). Such match is particularly important since it ensures the match between the EAL of the designed frame and the target EAL, at least under the assumptions adopted for the fragility and vulnerability models. Possibly, a pushover analysis can be conducted to further verify the simplified equilibrium/compatibility-based calculations;
10. Any undesired mismatch between the SDoF and the frame backbone curves may be corrected via: (a) iterations in the member design, thus repeating step 8; (b) revising Δ_y or h , thus restarting from step 3; (c) based on the results of a pushover analysis, revising the relative definition of μ_{DSi} with respect to the ductility capacity at peak strength μ_{cap} , thus restarting from step 1;
11. As per any seismic design procedure involving non-linear behaviour, members intended to remain elastic must be capacity protected. Moreover, some member characteristics may be modified to comply with code-based requirements.

This procedure is demonstrated for a four-storey RC case-study building in the city of L'Aquila and designed to achieve a 0.5% EAL. According to the Italian seismic risk classification, this would correspond to a building in the A+ risk class.⁴³ Both μ_{DSi} and DLR_{DSi} exemplified in step 1 are adopted for the design (consistently with those proposed in ref. 43 for Italian concrete buildings). The code-based Italian seismic hazard model is adopted²⁹; thus, no ad hoc site-specific probabilistic seismic hazard analysis is explicitly performed. For each point of a 5 km-spaced grid, the above model provides both PGA and spectral acceleration values for nine probabilities of exceedance in 50 years. The model provides the 16th, 50th, and 84th percentiles of the spectral accelerations for 10 period values between 0.1s and 2s, assuming rock conditions. The Italian code³⁰ also provides analytical approximations of the uniform hazard spectra consistent with the above model, including correction factors to obtain, among others, different soil and topography conditions. For the present illustration, a C-type soil category according to the Italian code is assumed (shear wave velocity in the first 30 m within the range 180–360 m/s).

The rectangular building is composed of four 9 m-spaced parallel frames with $N_{bay} = 3$, $l_b = 6m$, $H_{int} = 3.6m$, $h_b = 0.6m$, $h_c = 0.6m$, $f_{sy} = 300MPa$, $E_s = 200GPa$. A storey mass approximately equal to 400Ton is calculated by adopting, apart from the self-weight of the structure, a superimposed dead load equal to 2KPa and a factored live load equal to 0.9KPa. Thus, the storey mass affecting one frame is $m_i = 100Ton$.

Figure 8A shows the EAL of the seed SDoF systems, defined using a vector of 100 equally-spaced points between 0.1 and 0.4 for f_y and 100 equally-spaced points within 1.5 and 6 for μ_{cap} , thus leading to 10000 seed SDoF systems. A linear displacement shape is assumed for this four-storey frame,¹⁷ and the calculated yield displacement is $\Delta_y = 0.065m$, with $H_{eff} = 10.5m$. The MTf hysteresis model is used for this case study. Among the SDoF seeds meeting the target EAL (with a tolerance equal to $0.01EAL_{target}$), those complying with the code-based demand as per step 5 (design candidates) are represented in Figure 8B, together with the seismic demand corresponding to respectively, 30, 50, 475, and 975 year mean return periods. Those are related to the operability, damage limitation, life safety and near collapse DSs, defined

consistently with HAZUS. The final design SDoF system is arbitrarily selected among those, and it is characterised by $T = 0.97\text{s}$ (secant-to-yielding), $f_y = 0.27$, $\mu_{cap} = 2.65$, $h = 0.05$. The frequency of exceeding the complete damage DS for the design SDoF system is equal to 1.4×10^{-6} , which is deemed appropriate according to the above-mentioned limits.

According to the displacement capacity (at DS4) of the design SDoF system ($\Delta_{cap} = 0.17\text{m}$), steps 6 and 7 of the above procedure returned a demand equal to 325kNm(@0.8% plastic drift) for the beams, 347kNm(@0.8% plastic drift) for the exterior columns and 694kNm(@0.8% plastic drift) for the interior ones. Based on such theoretical values, cross-section design is performed characterising the flexural capacity of beams and columns via moment-curvature analysis (including gravity axial load). The model by Mander⁴⁴ is used for concrete, including calculating its ultimate strain (accounting for confinement). King et al.⁴⁵ is used for steel. Plastic hinge length is calculated according to ref. 46 Such a cross-section design required four iterations to reasonably match the SDoF design curve by appropriately balancing the (combined) strength of the members to achieve the desired OTM.

The RC frame is equipped with $0.6 \times 0.3\text{ m}$ beams reinforced with $4\phi 22\text{ mm}$ bars both in the tension and the compression sides of the cross section (340kNm yield moment). The interior columns have a $0.6 \times 0.4\text{ m}$ cross section reinforced with $4\phi 26\text{ mm}$ bars on both sides of the section and $2\phi 26\text{ mm}$ bars in the section centreline (729kNm yield moment). Finally, the exterior columns have a $0.5 \times 0.3\text{ m}$ cross section equipped with $8\phi 22\text{ mm}$ bars equally spaced along the perimeter (336kNm yield moment). The transverse reinforcement of the listed members is composed of $\phi 12\text{ mm}$ stirrups with 100 mm spacing, and the clear cover is equal to 30 mm. The unconfined compressive strength of the adopted concrete is equal to 30 MPa.

A pushover analysis (as per Section 3.2) is first carried out (Figure 9A). The comparison of the numerical pushover curve (both in the refined or bi-linearised forms) to the design SDoF curve is particularly satisfactory since the curves are essentially superimposed. The eigenvalue analysis of the frame indicates a fundamental period equal to 0.93s (secant-to-yielding), which is satisfactorily close to the SDoF one (0.98s). As discussed in Section 3.2, the member-level pushover results are used to identify refined DS thresholds. Figure 9A shows that the comparison of such thresholds (circular marks) with those assumed in step 1 of the procedure (triangular marks) is satisfactory since the maximum registered (displacement-wise) error is 11% (for DS4).

To thoroughly test the DLBD of this case study, NLTHA is conducted adopting the set of 100 records described in Section 2.2 (without using any scale factors). As per the frames in Section 3.2, fragility and vulnerability curves are herein derived based on the NLTHA results. In Figure 9A, such fragility curves (solid lines) are compared with those obtained via the GP regressions (dashed lines), respectively adopting the refined and simplified DS thresholds. Similarly, to the discussion in Section 3.2, the GP regressions generally under-estimate the median of the fragilities, with higher errors for more severe DSs (15.3%, 11.2%, 21.2%, and 19.3% for DS1 to DS4). The fragility dispersion obtained with the GP regressions underestimates the NLTHA-based one by 29.2% for this particular case. The fragility estimation error propagates to the vulnerability curves, with higher errors as IM increased. However, as anticipated in Section 3.2, such error levels are less relevant in the loss estimation since they correspond to hazard levels with low MAF of exceedance. This is reflected in the final calculation of EAL based on NLTHA, which is equal to 0.46% (conservatively predicted within 8% from the target value).

The result of this practical illustration, although encouraging, cannot be regarded as general. The tentative DLBD procedure proposed here is, at least arguably, worth a more comprehensive investigation/refinement/modification involving

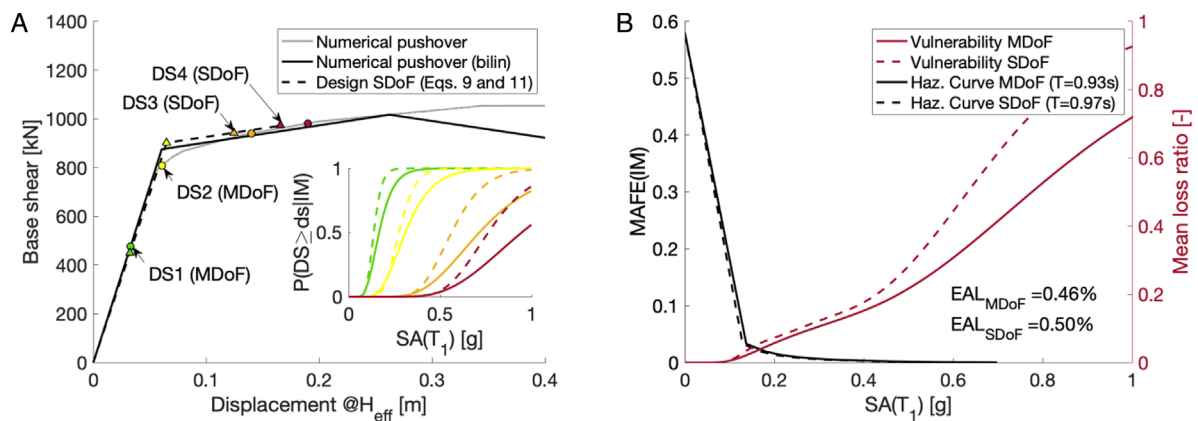


FIGURE 9 Design SDoF case versus design MDoF case: (A) pushover and fragility curves; (B) vulnerability and hazard curves

different case studies (in terms of material, lateral resisting system, geometry, etc.), different hazard levels, different loss metrics (e.g., the tail value at risk) and different target values of such metrics. Detailed comparison of design cases against refined NLTHA is required (also including 3D effects). If the procedure is proved successful, a final verification against more refined, component-by-component methodologies for loss analysis (e.g., FEMA P58, as opposed to the adopted building-level approach) is needed to estimate the confidence with which DLBD allows to set a target loss level.

4.2 | Dynamic earthquake risk modelling

A flexible and reliable surrogate model for the PSDM of structures (and hence for the derivation of fragility and vulnerability curves) can enable the fast development of a high number of earthquake risk models with different scenario exposure. Among all the possible examples falling in this category, this Section demonstrates a dynamic earthquake risk model with exposure changing over time, representing the implementation of a retrofit-based seismic risk reduction policy for a region.^{47,48}

This illustrative example involves a synthetic portfolio of 100 RC buildings. In the as-built condition (time $t = 0$), the buildings SDoF parameters are simulated based on the uniform distributions of $T \sim (0.2s, 1.5s)$, $f_y \sim (0.1, 0.25)$, $h \sim (0.01, 0.1)$, and $\mu_{peak} \sim (1.2, 3)$. All the buildings are characterised by the MTF hysteresis model.

An earthquake risk model can be used to simulate the effects, in terms of portfolio economic losses, of a risk-mitigation policy of this kind: each householder in the area must retrofit at a time $t \sim U(0, 15\text{years})$, and they must increase the yield strength such that $\Delta f_y \sim U(0, 0.15)$ and ductility capacity such that $\Delta \mu_{peak} \sim U(0.2, 1.5)$. Given the inherent freedom of the householders, the implementation of this policy may be regarded as a random retrofit process. Using the provided GP regressions, simulating such a process to obtain the distribution of the evolving portfolio losses becomes computationally feasible. This can be done according to Algorithm 1.

Algorithm 1. *Dynamic earthquake risk model to simulate a portfolio retrofit process.* \mathbf{Ret}_{sim} : number of simulations of the retrofit process; \mathbf{N}_{build} : number of buildings in the portfolio; \mathbf{t} : time;

1. procedure

2. calculate median fragility and dispersion for each building and each damage state
 3. generate a stochastic event set using relevant probability distributions:
occurrence time, rupture characteristics/event location, magnitude
 4. simulate spectral acceleration at a number of closely-spaced periods, $SA(T)$, for each site in the portfolio
 5. calculate portfolio losses at $t = 0$ (as-built condition)
 6. **for** $s = 1$ to \mathbf{Ret}_{sim} **do**
 7. simulate retrofit times t_i for each building; sort them in ascending order
 8. **for** $i = 1$ to \mathbf{N}_{build} **do**
 9. update retrofitted building parameters (T, f_y, h, μ_{cap})
 10. update fragility parameters of the retrofitted building (i)
 11. update building losses according to the relevant $SA(T)$ and the updated fragility curves
 12. aggregate portfolio losses at $t = t_i$
 13. calculate median portfolio loss curve for simulation s and time t_i and the related $EAL(s, t_i)$
 14. **end for**
 15. **end for**
 16. Calculate empirical distributions of the median portfolio loss curve and EAL
 17. **end procedure**
-

Figure 10 shows an application of Algorithm 1, considering one single simulation of the retrofit process described above ($\mathbf{Ret}_{sim} = 1$). The details of the earthquake risk modelling methodology and related assumptions are deemed less relevant for this illustrative application. They are only briefly introduced here, while they are described in detail in ref. 49. The buildings in the portfolio are equally spaced in a rectangular grid (Figure 10A) in the vicinity of a case-study strike-slip line fault. A 10,000-year stochastic catalogue is considered; for each event, 500 realisations of the ground motion fields are

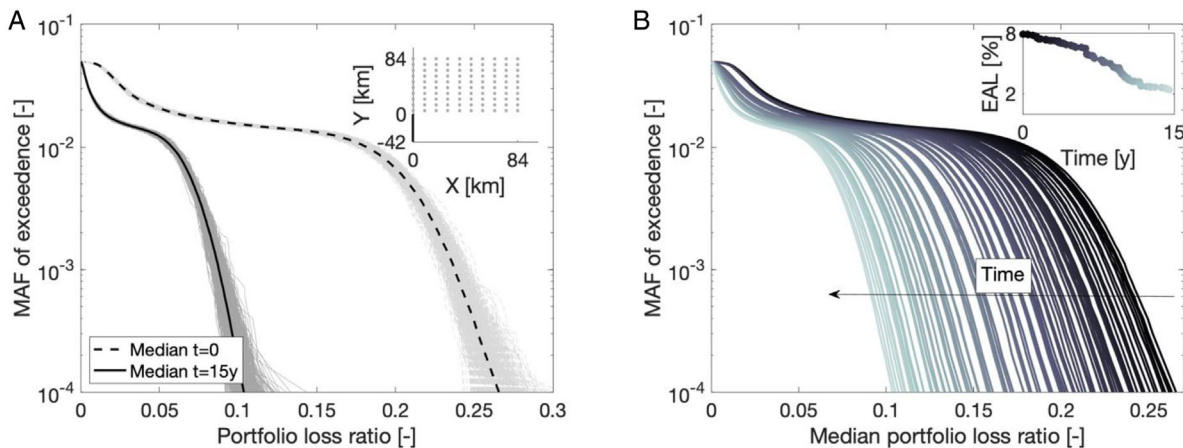


FIGURE 10 (A) Simulated portfolio loss curves at $t = 0$ (as-built portfolio) and $t = 15$ years (completed retrofit process); (B) Evolution of the median portfolio loss curves during the retrofit process

considered. Those are expressed in terms of SA in the period range (0.2s, 1.5s) with 0.1s spacing, appropriately considering the correlation coefficients at different vibration periods. The considered fault follows a log-linear recurrence relationship (i.e. frequency of occurrence vs moment magnitude M_w) for moderate (non-characteristic) events ($5 \leq M_w < 6.5$), and a constant probability branch of occurrence for events with $6.5 \leq M_w \leq 7$. Any pulse-like feature of the ground motions is neglected.

The fragility curves for each building in the portfolio are derived by adopting the fitted GP regressions. Those are related to DSs qualitatively defined according to HAZUS and quantified by the ductility thresholds $\mu_{DS} = [0.5, 1, 0.75\mu_{peak}, \mu_{peak}]$. Conservatively, the dispersion related to the non-linear range (Eq. 2) has been assigned to every DS. DLRs equal to 2%, 10%, 43.5%, and 100% of the total reconstruction cost (DS1 to DS4) are adopted. Figure 10A shows 500 simulations of the portfolio loss curves (mean annual frequency, MAF, of loss exceedance vs loss ratio) for the as-built portfolio ($t = 0$) and after the full simulation of the retrofit process ($t = 15$ years), showing the significant reduction of losses. Figure 10B shows the time evolution of the median portfolio loss curve, together with the evolution of the portfolio EAL.

5 | CONCLUSIONS

This paper proposed a metamodeling approach mapping the parameters controlling the dynamic behaviour of SDoF systems (i.e., force-displacement capacity curve, hysteretic behaviour) and their probabilistic seismic demand model (i.e., EDP vs IM distribution). This metamodel allows rapidly calculating fragility curves of SDoF representation of structures to be used in seismic risk and/or loss analyses.

The selected approach is the GP regression since it does not require any a priori definition of the output functional form (a GP is a non-parametric model), and therefore, it results in an infinitely-scalable surrogate model. The dataset used to train the GP regressions is based on a Monte Carlo sampling of 10,000 SDoF systems, each analysed via a cloud-based NLTHA using 100 ground-motion records. The code to make predictions using the proposed set of GP regressions is freely available (<https://github.com/robgen/surrogatedPSDM>), together with the adopted training datasets. With the same code, the users are also allowed to filter, extend or modify the training dataset and customise the fitting.

The results of this study can be highlighted as follows:

- Within the training dataset, the GP predictions for the slope of the PSDMs show an NRMSE equal to 2.5%, while this is equal to 6.2% for the logarithmic standard deviation. Such error estimates increase by 0.1% if a ten-fold cross validation test is conducted. Estimates of the median of the fragility curves consistent with DS3 and DS4 show an NRMSE respectively equal to 2.0% and 2.2%, while the error on the fragility dispersion remains equal to 6.2%.
- By selecting two baseline case-study SDoF systems, the distribution of the above fragility errors is propagated to the EAL. Simulating 30,000 realisations of the fragility parameters affected by such errors, the resulting

simulated-over-baseline EAL error follows a Normal distribution with mean equal to 1 and standard deviation equal to 0.015 and 0.01, respectively for the two case studies.

- The GP-based fragility predictions are tested against eight realistic RC frames with different levels of seismic performance. The NRMSE of such predictions is equal to 24.0%, 26.0%, 17.7%, and 19.0% for the DS1, DS2, DS3, and DS4 medians, respectively, if compared to NLTHA-based results. On the other hand, the inherent simplicity of the metamodels leads to a higher estimation error for the fragility dispersion, despite the NRMSE is equal to 29%.
- Considering the various sources of uncertainty typically involved in the seismic performance or risk/loss models, often not captured due to the simplified nature of the models themselves, the error levels introduced by using the proposed GP regressions are deemed satisfactory for practical applications, especially considering the low modelling effort and time required for the GP-based predictions.
- The proposed metamodels enable the development of a direct loss based design, for which a tentative procedure is proposed. This allows designing structures complying with a given target EAL, which is a fundamental input selected by the designer. The procedure consists of two phases: (1) the GP regressions are used to quickly select a design SDoF system complying with both the selected EAL target and the code-based demand at different DSs; (2) the members composing the lateral resisting system of the structure are designed (via equilibrium and compatibility principles) by matching the force-displacement curve of the designed structure with a target SDoF backbone. A realistic 2D RC frame is designed to achieve 0.5% EAL. Adopting an NLTHA-based approach and a building-level damage-to-loss model, the same parameter is measured to be 0.46%, thus showing an 8% error. Although this result is encouraging, the procedure requires further validations, refinement and/or modifications before it can be regarded generality.
- The proposed GP-regression approach also enables the fast development of a high number of earthquake risk models to be used in scenario-based decision making. This paper demonstrates a dynamic-exposure earthquake risk model representing the implementation of a retrofit-based seismic risk reduction policy for a region, in which householders are required to improve the performance of their buildings within a given time frame. By using the provided GP regressions, simulating such a random retrofit process becomes computationally feasible, thus allowing to obtain the distribution of the evolving portfolio losses. A seismic risk-mitigation policy implementation for a synthetic portfolio of 100 buildings is analysed for illustration purposes.

ACKNOWLEDGEMENTS

This study was partially funded by the European Union's Horizon 2020 research and innovation program under grant agreement No. 843794 (Marie Skłodowska-Curie Research Grants Scheme MSCA-IF-2018: Multi-level Framework to Enhance Seismic Resilience of RC buildings - MULTIRES) for Roberto Gentile. Carmine Galasso was supported by the project "Dipartimenti di Eccellenza", funded by the Italian Ministry of Education, University and Research at IUSS Pavia.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at <https://github.com/robgen/surrogatedPSDM>.

ORCID

Roberto Gentile  <https://orcid.org/0000-0002-7682-4490>

Carmine Galasso  <https://orcid.org/0000-0001-5445-4911>

REFERENCES

1. D'Ayala D, Meslem A, Vamvatsikos D, et al. Guidelines for analytical vulnerability assessment - low/mid-rise. GEM Technical Report. 2013.
2. Silva V, Akkar S, Baker J, et al. Current challenges and future trends in analytical fragility and vulnerability modeling. *Earthq Spectra*. 2019;35(4):1927-1952.
3. Mitchell-Wallace K, Jones M, Hillier J, Foote M. *Natural catastrophe risk management and modelling*. New York: Wiley; 2017.
4. Gentile R, Galasso C. Simplicity versus accuracy trade-off in estimating seismic fragility of existing reinforced concrete buildings. *Soil Dyn Earthquake Eng*. 2021;144:106678.
5. Vaseghiamiri S, Mahsuli M, Ghannad MA, Zareian F. Surrogate SDOF models for probabilistic performance assessment of multistory buildings: methodology and application for steel special moment frames. *Eng Struct*. 2020;212:110276.
6. Du A, Padgett JE. Investigation of multivariate seismic surrogate demand modeling for multi-response structural systems. *Eng Struct*. 2020;207:110210.
7. Vamvatsikos D, Cornell A. Direct estimation of the seismic demand and capacity of oscillators with multi-linear static pushovers through IDA. *Earthq Eng Struct Dyn*. 2006;35(9):1097-1117.

8. Nafeh AMB, O'Reilly GJ, Monteiro R. Simplified seismic assessment of infilled RC frame structures. *Bull Earthquake Eng.* 2020;18(4):1579-1611.
9. Kircher CA, Whitman RV, Holmes WT. HAZUS earthquake loss estimation methods. *Nat Hazard Rev.* 2006;7(2):45-59.
10. European Committee for Standardisation (CEN). *EN 1998-1. Eurocode 8: design of structures for earthquake resistance. Part 1: general rules, seismic action and rules for buildings.* Brussels, Belgium: 2005.
11. New Zealand Society for Earthquake Engineering (NZSEE). *The seismic assessment of existing buildings—technical guidelines for engineering assessments.* Wellington, New Zealand: 2017.
12. Jalayer F, Cornell CA. Alternative non-linear demand estimation methods for probability-based seismic assessments. *Earthquake Eng Struct Dyn.* 2009;38(8):951-972.
13. Gentile R, Galasso C. Gaussian process regression for seismic fragility assessment of building portfolios. *Struct Saf.* 2020;87:101980.
14. Mangalathu S, Heo G, Jeon JS. Artificial neural network based multi-dimensional fragility development of skewed concrete bridge classes. *Eng Struct.* 2018;162:166-176.
15. Ghosh J, Padgett JE, Dueñas-Osorio L. Surrogate modeling and failure surface visualization for efficient seismic vulnerability assessment of highway bridges. *Probab Eng Mech.* 2013;34:189-199.
16. Saiidi M, Sozen M. *Simple and complex models for nonlinear seismic response of reinforced concrete structures.* Urbana, Illinois, USA. 1979.
17. Priestley MJN, Calvi GM, Kowalsky MJ. *Displacement-based seismic design of structures.* Pavia, Italy: IUSS Press; 2007.
18. Rasmussen CE, Williams CKI. *Gaussian processes for machine learning.* Cambridge, Massachusetts, USA: the MIT Press; 2006.
19. Minas S, Galasso C. Accounting for spectral shape in simplified fragility analysis of case-study reinforced concrete frames. *Soil Dyn Earthquake Eng.* 2019;119:91-103.
20. Smerzini C, Galasso C, Iervolino I, Paolucci R. Ground motion record selection based on broadband spectral compatibility. *Earthq Spectra.* 2014;30(4):1427-1448.
21. Luco N, Bazzurro P. Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses?. *Earthq Eng Struct Dyn.* 2007;36(13):1813-1835.
22. Dávalos H, Miranda E. Evaluation of the scaling factor bias influence on the probability of collapse using Sa(T1) as the intensity measure. *Earthq Spectra.* 2019;35(2):679-702.
23. Haselton CB, Baker JW, Liel AB, Deierlein GG. Accounting for ground-motion spectral shape characteristics in structural collapse assessment through an adjustment for Epsilon. *J Struct Eng.* 2011;137(3):332-344.
24. O'Reilly GJ, Monteiro R. Probabilistic models for structures with bilinear demand-intensity relationships. *Earthq Eng Struct Dyn.* 2019;48(2):253-268.
25. Sevieri G, Gentile R, Galasso C. A multi-fidelity Bayesian framework for robust seismic fragility assessment. *Earthq Eng Struct Dyn.* 2021. **accepted.**
26. Nocedal J, Wright SJ. *Numerical optimization.* New York, NY: Springer; 2006.
27. European Committee for Standardisation (CEN). Eurocode 8: Design of structures for earthquake resistance. Part 3: Strengthening and repair of buildings 2005.
28. Federal Emergency Management Agency. Seismic performance assessment of buildings. Volume 1 - methodology. 2012.
29. Stucchi M, Meletti C, Montaldo V, Crowley H, Calvi GM, Boschi E. Seismic hazard assessment (2003-2009) for the Italian building code. *Bull Seismol Soc Am.* 2011.
30. Consiglio dei Ministri. *DM 17 gennaio 2018 in materia di "aggiornamento delle norme tecniche per le costruzioni". Gazzetta ufficiale n.42 del 20 febbraio 2018.* 2018.
31. Consiglio dei Ministri. *Regio Decreto Legge n. 2229 del 16/11/1939. G.U. n.92 del 18/04/1940.* 1939.
32. Gentile R, Pampanin S, Galasso C. A computational framework for selecting the optimal combination of seismic retrofit and insurance coverage. *Computer-Aided Civil and Infrastructure Engineering.* 2021; <https://doi.org/10.1111/mice.12778>
33. Carr AJ. *RUAUMOKO2D - the maori god of volcanoes and earthquakes. Inelastic Analysis Finite Element program.* Christchurch, New Zealand. 2016.
34. Magenes G, Pampanin S. Seismic response of gravity-load design frames with masonry infills. *13th World Conference on Earthquake Engineering,* 2004.
35. Nettis A, Gentile R, Raffaele D, Uva G, Galasso C. Cloud capacity spectrum method: accounting for record-to-record variability in fragility analysis using nonlinear static procedures. *Soil Dyn Earthquake Eng.* 2021;150:106829.
36. Freeman SA. Development and use of capacity spectrum method. *6th U.S. National Conference of Earthquake Engineering,* Seattle: 1998.
37. Jalayer F, Ebrahimian H, Miano A, Manfredi G, Sezen H. Analytical fragility assessment using unscaled ground motion records. *Earthq Eng Struct Dyn.* 2017;46(15):2639-2663.
38. ATC 40. *Applied technology council, seismic evaluation and retrofit of concrete buildings.* Redwood City; 1996.
39. Shahnazaryan D, O'Reilly GJ. Integrating expected loss and collapse risk in performance-based seismic design of structures. *Bull Earthquake Eng.* 2021;19(2):987-1025.
40. Krawinkler H, Zareian F, Medina RA, Ibarra LF. Decision support for conceptual performance-based design. *Earthq Eng Struct Dyn.* 2006;35(1):115-133.
41. O'Reilly GJ, Calvi GM. Conceptual seismic design in performance-based earthquake engineering. *Earthq Eng Struct Dyn.* 2019;48(4):389-411.

42. Dolšek M, Lazar Sinković N, Žižmond J. IM-based and EDP-based decision models for the verification of the seismic collapse safety of buildings. *Earthq Eng Struct Dyn*. 2017;46(15):2665-2682.
43. Cosenza E, Del Vecchio C, Di Ludovico M, et al. *The Italian guidelines for seismic risk classification of constructions: technical principles and validation*. Netherlands: Springer; 2018.
44. Mander JB, Priestley MJN, Park R. Theoretical stress strain model for confined concrete. *J Struct Eng*. 1988;114(8):1804-1826.
45. King DJ, Priestley MJN, Park R. *Computer programs for concrete column design, Research Report 86/12*. Department of Civil Engineering; 1986.
46. Priestley MJN, Park R. Strength and ductility of concrete bridge columns under seismic loading. *ACI Struct J*. 1987;84:61-76.
47. Yi Z, Burton HV, Shokrabadi M, Issa O. Multi-scale cost-benefit analysis of the Los Angeles soft-story ordinance. *Eng Struct*. 2020;214:110652.
48. Liel AB, Deierlein GG. Using collapse risk assessments to inform seismic safety policy for older concrete buildings. *Earthq Spectra*. 2012;28(4):1495-1521.
49. Gentile R, Galasso C. Accounting for directivity-induced pulse-like ground motions in building portfolio loss assessment. *Bulletin of Earthquake Engineering*. 2020. <https://doi.org/10.1007/s10518-020-00950-9>

How to cite this article: Gentile R, Galasso C. Surrogate probabilistic seismic demand modelling of inelastic single-degree-of-freedom systems for efficient earthquake risk applications. *Earthquake Engng Struct Dyn*. 2021;1–20. <https://doi.org/10.1002/eqe.3576>