Received: 30 October 2023 | Revised: 5 March 2024 | Accepted: 6 March 2024

DOI: 10.1002/eqe.4115

RESEARCH ARTICLE

WILEY

Stochastic emulation with enhanced partial- and no-replication strategies for seismic response distribution estimation

Sang-ri Yi[1] | **Alexandros A. Taflanidis**[2]

[1]Department of Civil & Environmental Engineering, University of California, Berkeley, California, USA

[2]Department of Civil & Environmental Engineering & Earth Sciences, University of Notre Dame, Notre Dame, Indiana, USA

Correspondence
Sang-ri Yi, Department of Civil & Environmental Engineering, University of California, Berkeley, CA 94720, USA.
Email: yisangri@berkeley.edu

Funding information
National Science Foundation, Grant/Award Number: CMMI-2131111

Abstract

Modern performance earthquake engineering practices frequently require a large number of time-consuming non-linear time-history simulations to appropriately address excitation and structural uncertainties when estimating engineering demand parameter (EDP) distributions. Surrogate modeling techniques have emerged as an attractive tool for alleviating such high computational burden in similar engineering problems. A key challenge for the application of surrogate models in earthquake engineering context relates to the aleatoric variability associated with the seismic hazard. This variability is typically expressed as high-dimensional or non-parametric uncertainty, and so cannot be easily incorporated within standard surrogate modeling frameworks. Rather, a surrogate modeling approach that can directly approximate the full distribution of the response output is warranted for this application. This approach needs to additionally address the fact that the response variability may change as input parameter changes, yielding a heteroscedastic behavior. Stochastic emulation techniques have emerged as a viable solution to accurately capture aleatoric uncertainties in similar contexts, and recent work by the second author has established a framework to accommodate this for earthquake engineering applications, using Gaussian Process (GP) regression to predict the EDP response distribution. The established formulation requires for a portion of the training samples the replication of simulations for different descriptions of the aleatoric uncertainty. In particular, the replicated samples are used to build a secondary GP model to predict the heteroscedastic characteristics, and these predictions are then used to formulate the primary GP that produces the full EDP distribution. This practice, however, has two downsides: it always requires minimum replications when training the secondary GP, and the information from the non-replicated samples is utilized only for the primary GP. This research adopts an alternative stochastic GP formulation that can address both limitations. To this

end, the secondary GP is trained by measuring the square of sample deviations from the mean instead of the crude sample variances. To establish the primitive mean estimates, another auxiliary GP is introduced. This way, information from all replicated and non-replicated samples is fully leveraged for estimating both the EDP distribution and the underlying heteroscedastic behavior, while formulation accommodates an implementation using no replications. The case study examples using three different stochastic ground motion models demonstrate that the proposed approach can address both aforementioned challenges.

KEYWORDS
aleatoric uncertainty, seismic risk estimation, stochastic emulation, stochastic Gaussian process regression, surrogate modeling

1 | INTRODUCTION

Assessment of risk in earthquake engineering applications requires proper consideration of various types of uncertainties for the seismic hazard and the infrastructure models.[1] Recent advancement of high-fidelity computational simulation models[2,3] have dramatically improved our ability to provide accurate predictions for infrastructure response through non-linear response history analysis (NLRHA), but the use of such models in assessing seismic risk creates frequently a prohibitive computational burden due to the large number of simulations needed to address the aforementioned uncertainty sources. In similar engineering problems the use of surrogate modeling techniques has emerged as an attractive approach to alleviate the computational burden associated with the use of expensive numerical models.[4–10] Surrogate models, also referred to as metamodels or emulators, replace the original high-fidelity model with a computationally efficient (fast) data-driven approximation developed using a small number of judicially chosen simulations, frequently referenced as computer experiments or training points. In earthquake engineering applications such surrogate modeling techniques can be used to establish an approximate mapping between inputs (i.e., the parameters of earthquake and structural models) and outputs of interest (structural responses), to accelerate risk assessment.[11–18]

Although attractive, surrogating seismic structural behavior is challenging because of the aleatoric uncertainty associated with the hazard description.[11] This type of uncertainty originates from the complex physical mechanism in propagation/generation of earthquakes, which precludes the complete description of the seismic excitation through parametrized explanatory variables (EVs).[1,19,20] Depending on the approach adopted for modeling the seismic excitation (ground motion time-history in context of NLRHA)[21] this variability may be expressed as high-dimensional or non-parametric uncertainty. From the perspective of the numerical simulations to predict the engineering demand parameters (EDPs) of interest, this yields simulation models that are stochastic rather than deterministic: the response output for the same explanatory seismic input parameters (e.g., intensity measure vector) will vary across simulations depending on the exact selection of the ground motion according to the underlying aleatoric uncertainty in the hazard description.[20] This variability is large and crucial to the downstream analysis of structural performance estimation[1] necessitating for a high-precision estimation of the underlying EDP distribution. Meanwhile, the level of variability depends significantly on structural and seismic explanatory inputs[22] (representing a heteroscedastic behavior) increasing the complexity of the risk assessment.

A branch of specialized surrogate modeling techniques called *stochastic emulation* has emerged as a popular approach for accommodating similar challenges associated with non-deterministic simulation models.[23–31] Many of the stochastic emulation approaches leverage specially designed configuration of simulation experiments called replications, established by repeating the simulation for identical input values. Such replications of the stochastic simulation code can be used to provide sample-based estimates of the response statistics of the output, such as the non-stationary mean and heteroscedastic variance across replications, and are used directly by the stochastic emulator to approximate the output distribution.[23,26] Albeit the demonstrated effectiveness of such techniques, the computational burden associated with the need to perform replications for the simulation experiments limits their applicability for many practical applications with expensive simulation models. This challenge has been further addressed by recent developments that reduce[26] or even entirely avoid[29,31]

the need for replications by introducing latent variables to approximate the underlying statistics, instead of establishing sample-based approximations.

In the field of earthquake engineering, stochastic emulation techniques were recently adopted[32] for describing the EDP distribution, utilizing Gaussian Process (GP) regression (also referred to as kriging) as foundational surrogate model technique. In this instance, the non-deterministic features of the structural simulations originate from the aleatoric component of the seismic hazard variability. The work in ref. [32] builds closely upon the developments in refs. [23, 26] and relies on the concept of partial-replications to achieve computational efficiency, formulating a framework which incorporates both replicated and non-replicated simulation experiments for the emulator development. In particular, the replicated samples are used to obtain variance estimates at the sampled training points, and an auxiliary GP, termed secondary GP, is then introduced to denoise the sample variance observations and interpolate them to construct a continuous mapping of a variance-field across the input space. Then, one can constrain the relative scales of the variance (i.e., the heteroscedastic characteristics) to train a primary GP to approximate the EDP distribution, exploiting both replicated and non-replicated samples.[32] The approach is computationally attractive as, by introducing the point estimate of variance, it avoids complex alternative solutions for approximating the heteroscedastic behavior,[26,33–35] but has two drawbacks: (i) it always requires some minimal replications; and (ii) the information from non-replicated samples is unused during the approximation of the heteroscedastic variance. Both these drawbacks originate from the fact that the variance approximation in the secondary GP leverages estimates of the sample variance which cannot be provided from the non-replicated samples.

To address the two limitations, this work proposes a modified version of the partial-replication-based stochastic emulation approach that fully utilizes the information from both replicated and non-replicated samples. In particular, it enables the use of non-replicated samples in the variance estimation by adopting the techniques discussed in refs. [30, 36] to replace the observations of sample variance with the squared sample deviation from the mean. To estimate this deviation a primitive mean estimate at each sample location is required, obtained by introducing an additional, tertiary, GP model relying on a homoscedasticity assumption. The latter reliance is justified by the results reported in ref. [32], showing that the mean field estimation is relatively robust to the variance approximation model, even when the underlying observations have high heteroscedastic aleatoric variability. Accordingly, the mean of the response is first approximated using the tertiary GP under the simplified assumption of constant aleatoric variance. Then, taking advantage of the knowledge of the mean, the response variance is approximated in terms of sample squared deviations from the mean, as opposed to the original approach that relied on the sample variance estimation. In this way, not only replicated but also non-replicated samples can contribute in developing the secondary GP for the variance-field. This further enables the framework to be used without any replications, allowing the computational budget previously used for the replication to be used to further explore the unexplored part of the input domain. Meanwhile, this workflow inherits an attractive property of the original strategy, that optimization across latent features is not required, as the latter task may create overfitting challenges to the observation data for problems with high levels of aleatoric uncertainty or large number of exploratory variables.[26] The proposed strategy is demonstrated utilizing stochastic ground motion models to describe the seismic hazard, examining three different parameterization settings to explore the impact of the different sources of excitation aleatoric uncertainty.[22]

2 | SEISMIC RISK ASSESSMENT USING SURROGATE MODELS

Risk assessment within performance based earthquake engineering and seismic loss assessment applications needs to address various sources of uncertainty related to the structural model and/or seismic excitation descriptions.[1,3,37–39] The most influential types of uncertainty are associated with the seismic excitation, and can be broadly divided into two classes. The first class pertains to key seismicity characteristics and/or ground motion features. Depending on the approach adopted to describe the seismic hazard, these may pertain to the moment magnitude, rupture location and local soil properties, to underlying excitation characteristics such as arias intensity, significant duration and frequency content, or to properly chosen intensity measures. This class represents the parametric explanatory input variables that may be used to describe the dominant hazard features in the problem formulation, but can only partially explain the structural response variability. The remaining variability, that is the portion that cannot be described using such EVs, is attributed, as discussed in the introduction, to miscellaneous sources of randomness and represents the second class of uncertainty in the hazard description,[20] corresponding to aleatoric randomness. Depending on the modeling approach for describing the seismic excitation, this uncertainty source might not have a parametric description (corresponds to latent features of the excitation model), or might correspond to a high-dimensional stochastic sequence.[32] Within any predictive modeling setting, this

randomness needs to be treated as inherent uncertainty in the model output given the values of the seismic hazard EVs. Additionally, the structural model might have its own uncertain parameters, such as material and geometric properties, that need to be used as EVs when estimating structural response.

Based on this uncertainty description, the objective of the surrogate modeling within the seismic risk assessment context is to create a mapping between all the EVs, that is, seismicity features and structural parameters, and the EDP distribution of a target structure, where the distribution randomness is attributed to the aleatory uncertainties.[15,32,40] To formalize this concept, let us denote the structural response (EDP) by z, the explanatory seismic excitation characteristics by $\mathbf{x}_h$, and the uncertain structural parameters by $\mathbf{x}_s$. The augmented vector of excitation and structural EVs will be denoted by $\mathbf{x} = \{\mathbf{x}_h, \mathbf{x}_s\} \in \mathbb{R}^{n_x}$, with n_x representing its dimension. The goal of the surrogate model is to produce an approximation to the distribution $p(z|\mathbf{x})$, with notation "|" introduced to denote conditional relationship. It should be pointed out that the estimation of this distribution is one of the key components of modern seismic risk assessment practices.[1,3] Once this distribution as a function of $\mathbf{x}$ is accurately approximated, the propagation of all uncertainties associated with $\mathbf{x}$, to accommodate the risk assessment, can be efficiently performed, for example, using standard Monte Carlo Simulation (MCS) techniques.[32,41]

Stochastic emulation techniques establish directly an approximation for the full EDP distribution, $p(z|\mathbf{x})$, instead of approximating, for example, the EDP statistics.[11] Specifically, GP-based emulation is considered in this work for the $p(z|\mathbf{x})$ approximation, representing one of the most popular stochastic emulation techniques.[23,26] The foundation of the GP emulation is the approximation of the chosen response y as a realization of a GP. Instead of approximating directly the EDP, in seismic risk assessment applications it is recommended to approximate its logarithm, setting $y = \ln(z)$ for the GP output.[32] This agrees with the lognormal assumption for $p(z|\mathbf{x})$ frequently utilized in earthquake engineering.[1] The underlying GP has some mean trend function over $\mathbf{x}$-typically expressed by a vector of basis functions $\mathbf{f}(\mathbf{x})$ and their regression coefficients $\boldsymbol{\beta}$—and stationary auto-covariance model denoted as $\tilde{\sigma}^2 R(\mathbf{x}, \mathbf{x}'|\boldsymbol{\theta})$ with kernel parameters $\boldsymbol{\theta}$ and process variance $\tilde{\sigma}^2$.[8,42] Leveraging the so-called nugget[27,43] to establish a regression over the training data and address the aleatoric randomness, the stochastic emulation model yields the following distribution approximation[32]

$$\hat{p}(y|\mathbf{x}) = N(\tilde{y}(\mathbf{x}), \sigma^2(\mathbf{x}) + \tau^2(\mathbf{x})) \tag{1}$$

where $N(a, b)$ stands for a Normal distribution with mean a and variance b, $\tilde{y}(\mathbf{x})$ and $\sigma^2(\mathbf{x})$ represent the predictive mean and variance for the GP and $\tau^2(\mathbf{x})$ the nugget variance. Note that the nugget variable $\varepsilon \sim N(0, \tau^2(\mathbf{x}))$ is a zero mean normal variable statistically independent across the location $\mathbf{x}$,[27,43] that is, ε can be interpreted as white noise with varying scale respective to $\mathbf{x}$. When $\tau^2(\mathbf{x}) = 0$, the GP regression reduces to an exact interpolation which is useful only when one has noiseless observation of (well-posed) deterministic simulators. When $\tau^2(\mathbf{x}) = \tau^2$ in Equation (1) is a constant, the problem reduces to a simpler homoscedastic case. In this case, only one parameter τ^2 needs to be identified to describe the aleatoric variance scale, though it has been shown[32] that this is a poor approximation for describing hysteretic structural response distributions. Therefore, the emphasis is herein on how to address a heteroscedastic nugget variance $\tau^2(\mathbf{x})$ across the domain $\mathbf{x}$. Appendix A presents a generic GP formulation covering all aforementioned variants. The first two implementations, that is, no nugget and homoscedastic nugget, are obtained by selecting as output $v = y = \ln(z)$ and adopting the simplified assumptions reviewed in Table A1.

3 | REVIEW OF STOCHASTIC EMULATION WITH PARTIAL REPLICATIONS

This section briefly reviews the stochastic emulation with partially replicated samples.[32] This approach leverages two GP models: the primary one with heteroscedastic nugget approximates the response $z(\mathbf{x})$, while its nugget variance $\tau^2(\mathbf{x})$ is approximated by the secondary GP.

For formulating the GP metamodel, a database of observations is first developed by performing simulations for N_t different input configurations $\{\mathbf{x}^i; i = 1, \ldots, N_t\}$, representing the GP training points. The training points should cover the entire input domain, denoted as X herein, where the surrogate model is anticipated to be used for, so that any extrapolations are avoided. For each training point, the EDP prediction is repeated a total of n^i times yielding observations $\{y^{i(k)}; k = 1, .., n^i\}$. These represent replications of the output for the same input $\mathbf{x}^i$, that is, $y^{i(k)} = y^{(k)}(\mathbf{x}^i)$, within a stochastic emulation/simulation setting, and are obtained by utilizing different realizations for the aleatoric hazard description for the same selection for the excitation EVs $\mathbf{x}_h^i$. For example, in the context of NLRHA utilizing stochastic ground motion models, these realizations correspond to acceleration time-histories obtained by utilizing different stochastic sequences.[32]

Herein superscripts will be used to distinguish the training points and superscripts in parenthesis will be used to distinguish the replications. Following ref. [32] let us consider a special configuration of partially replicated training set, where first N_r samples have n_r replications ($n^i = n_r$ for $i \leq N_r$) and remaining $N_n = N_t - N_r$ samples are non-replicated samples ($n^i = 1$ for $N_r < i \leq N_t$). The set of training points for which replications are available will be denoted as $\mathbf{X}_p$. The total number of high-fidelity simulations for the stochastic emulation training is $N_{sim} = n_r N_r + N_n$.

For the training inputs with replications ($\mathbf{X}_p$), unbiased estimates for the sample variance are obtained as:

$$(\bar{\tau}^2)^i = \frac{1}{n_r - 1} \sum_{k=1}^{n_r} \left(y^{i(k)} - \bar{y}^i \right)^2, \quad \text{for} \quad i = 1, \ldots, N_r$$

$$\text{where} \quad \bar{y}^i = \frac{1}{n_r} \sum_{k=1}^{n_r} y^{i(k)} \tag{2}$$

These estimates can serve as noisy observations to train a secondary GP model, denoted GP_τ herein, to predict the continuous field of the response variance. The noise in these observations originates from the variability in the sample variances of Equation (2) [sample-based estimation error], and can be addressed using a homoscedastic nugget in the GP formulation.[32] The predictions of GP_τ are utilized to describe the heteroscedastic behavior of the EDP distribution, denoted as $\tilde{\tau}^2(\mathbf{x})$, and scaling the nugget variance as $\tau^2(\mathbf{x}) = \tau_s^2 \tilde{\tau}^2(\mathbf{x})$, where τ_s^2 is a scaling parameter to be calibrated based on the available observations, formulates then the primary GP, denoted as GP_y. The workflow for this stochastic emulation using the partial replications, denoted *PR-SE*, is summarized by the following steps[32]:

Step 1: Estimation of sample variance: For the $\mathbf{X}_p$ training points $\{\mathbf{x}^i; i = 1, .., N_r\}$, the available replications $\{y^{i(k)}; k = 1, .., n^i\}$ are used to obtain the sample-based estimates $(\bar{\tau}^2)^i$ for the aleatoric variance using Equation (2).

Step 2: Approximation of heteroscedastic variance by GP_τ: Train a GP model that approximates the response variance across the input domain using the observation set $\{\mathbf{x}^i, (\bar{\tau}^2)^i; i = 1, .., N_r\}$. Since the variance is strictly positive, following recommendations in ref. [26], its logarithm is actually approximated (though note this is not necessary). This further justifies[32] the use of a homoscedastic nugget that greatly simplifies the GP_τ formulation. The GP_τ follows the Ho-NR formulation (Table A1) discussed in Appendix A utilizing output selection $v = ln(\tau^2)$, number of training points $N_s = N_r$, and provides approximation $\ln(\tilde{\tau}^2(\mathbf{x}))$ given by Equation (A2). Formulation entails the hyper-parameter calibration given by Equation (A6) utilizing observation set for the replicated samples $\{\mathbf{x}^i, (\bar{\tau}^2)^i; i = 1, .., N_r\}$.

Step 3: Estimation of the response by GP_y: Using $\tilde{\tau}^2(\mathbf{x})$ as unscaled variance component, formulate heteroscedastic GP model that approximates the response distribution across the domain of interest utilizing all available observations $\{\mathbf{x}^i, y^{i(k)}; i = 1, \ldots, N_t, k = 1, \ldots, n^i\}$. The GP_y follows the He-R formulation (Table A1) discussed in Appendix A utilizing output selection $v = \ln(z)$, number of training points $N_s = N_t$, heteroscedastic variability $\tilde{c}(\mathbf{x})^2$ corresponding to $\tilde{\tau}^2(\mathbf{x})$ [from Step 2] and provides approximation $\hat{p}(y|\mathbf{x}) = N(\tilde{y}(\mathbf{x}), \sigma^2(\mathbf{x}) + \tau^2(\mathbf{x}))$ with $\tilde{y}(\mathbf{x})$ given by Equation (A3), $\sigma^2(\mathbf{x})$ by Equation (A3) and $\tau^2(\mathbf{x})$ by Equation (A4). Formulation entails the hyper-parameter calibration given by Equation (A7) utilizing observation set composed of both the replicated and non-replicated samples, $\{\mathbf{x}^i, y^{i(k)}; i = 1, \ldots, N_r; k = 1, \ldots, n_r\}$ and $\{\mathbf{x}^i, y^i; i = N_r + 1, \ldots, N_t\}$, respectively.

As discussed in the introduction, drawbacks of this stochastic emulation strategy are that: (i) replication is always needed in Step 1 to obtain the observation of Equation (2) utilized in Step 2 to establish the GP_τ; and (ii) the samples without replications are utilized only in Step 3 for the calibration of GP_y and not in Step 2 for the calibration of GP_τ, even though correct predictions for GP_τ play a critical role in accurately estimating the EDP distribution. Both issues arise from the fact that the variance estimation in Equation (2) hinges on only the the replicated samples. Such discrepancies in the information utilized from the replicated and non-replicated samples, creates trade-offs in the choice of the number and size of replications, that is, given the total computational budget N_{sim}, having more replications restrict the exploration of domain (smaller N_t), while more non-replication samples will lead to poor variance estimation (smaller N_r or n_r). This conflicting benefits have been also investigated by numerical experiments[32] which showed that, given number of total model evaluations, the overall accuracy improves as more replications are included only until the ratio reaches certain point at which it starts to decrease. Presumably, this conflict arises mainly because the non-replicated sample did not contribute to improving the accuracy in the response variance approximation. Furthermore, generating sufficient

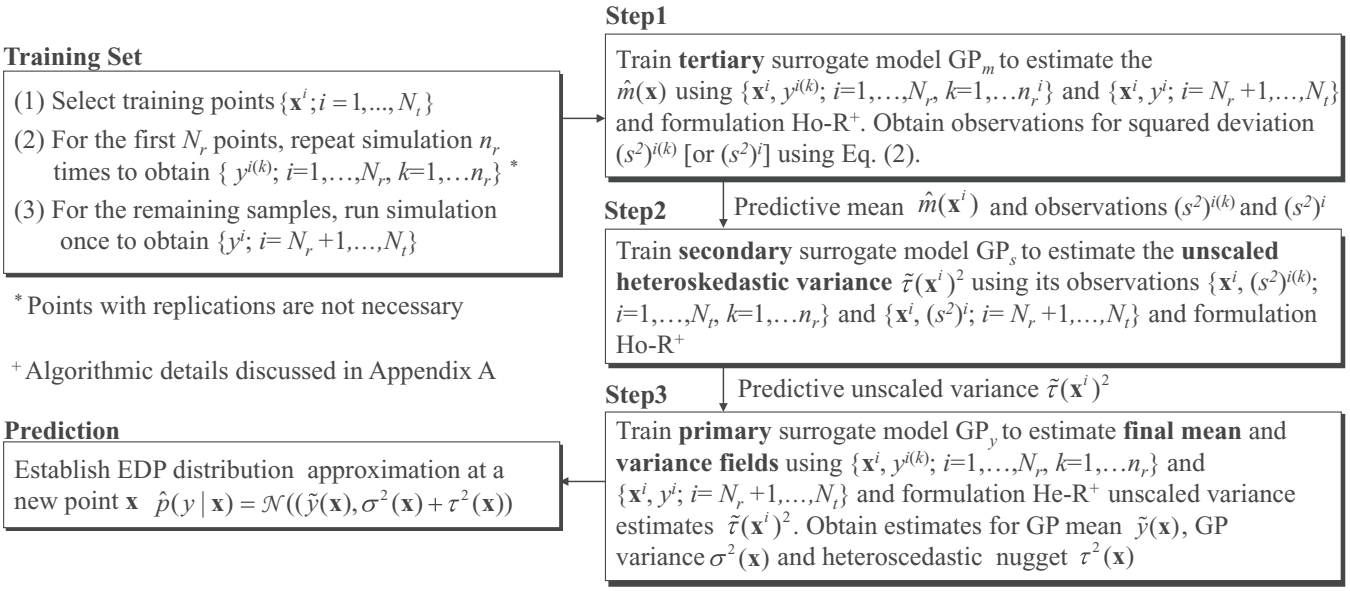

FIGURE 1 Workflow for *ER-SE* algorithm.

replications may not be always feasible. This can occur, for example, when $\mathbf{x}_h$ represents a higher-dimensional intensity measure vector, and one cannot find multiple ground motions for all desired $\mathbf{x}_h$ values, limiting the ability to consider replications for them. To address these challenges an enhanced replication strategy is discussed next, allowing stochastic emulation even for the limiting case without any replications.

4 | STOCHASTIC EMULATION FOR EDP DISTRIBUTION APPROXIMATION WITH ENHANCED REPLICATION OR NO-REPLICATION STRATEGY

4.1 | Algorithmic implementation of the improved stochastic emulation

The objective of the improved stochastic emulation implementation is to utilize all points (even the non-replication points) for the variance-field approximation in Step 2 of the *PR-SE* algorithm discussed in Section 3. Following ref. [30], this is achieved by introducing the *squared sample deviation* to replace the *sample variance* utilized in that step, so that information from both replicated and non-replicated samples can be utilized. The sample deviation is defined as the distance of the response observation from some primitive estimate of the response mean, denoted $\hat{m}(\mathbf{x})$, and its square is expressed as

$$(s^2)^{i(k)} = \left(y^{i(k)} - \hat{m}(\mathbf{x}^i)\right)^2, \quad i = 1,\ldots,N_t \text{ and } k = 1,\ldots,n^i \tag{3}$$

This sample deviation can be used as an alternative to Equation (2) for the variance-field observations, and requires to first establish the prediction of $\hat{m}(\mathbf{x})$. Different smoothing techniques[44–46] can be used to approximate $\hat{m}(\mathbf{x})$, and the one adopted here is to introduce an additional GP with homoskedesity assumption.[30,36] This tertiary GP will be denoted herein as GP_m.

The workflow for this updated stochastic emulation algorithm, denoted *ER-SE*, is shown in Figure 1 and summarized below. Note that the major difference between *PR-SE* and *ER-SE* lies in Steps 1 and 2, while Step 3 is identical for the two implementations.

Step 1: Preliminary estimation of the mean response GP_m and approximation of the sample squared deviation: Train a GP model that establishes a crude approximation of the mean response using a homoscedastic nugget assumption and all available observations $\{\mathbf{x}^i, y^{i(k)}; i=1,\ldots,N_t, k=1,\ldots,n^i\}$. The GP_m follows the Ho-R formulation (Table A1) discussed in Appendix A utilizing output selection $\upsilon = \ln(z)$, number of training points $N_s = N_t$, homoscedastic nugget selection, and provides approximation for the mean across all training points $\hat{m}(\mathbf{x}^i)$, provided by Equation (A2). Formulation entails the hyper-parameter calibration given by Equation (A6) utilizing

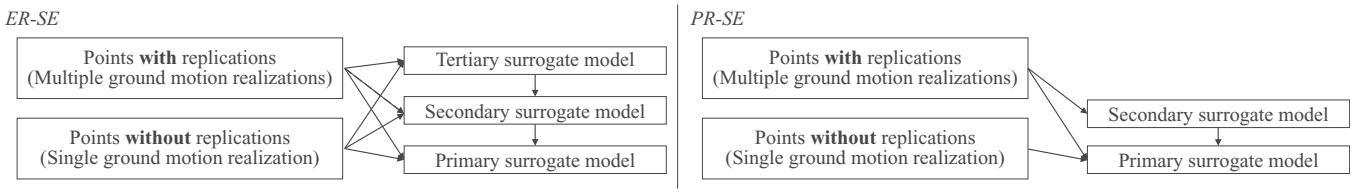

FIGURE 2 Schematic comparison of the flow of information in *ER-SE* and *PR-SE*.

observation set composed of both the replicated and non-replicated samples, $\{\mathbf{x}^i, y^{i(k)}; i = 1, \dots, N_r; k = 1, \dots, n_r\}$ and $\{\mathbf{x}^i, y^i; i = N_r + 1, \dots, N_t\}$, respectively. Then, obtain for each observation the squared deviation from the mean $(s^2)^{i(k)}$ utilizing Equation (3) and the mean field approximation $\hat{m}(\mathbf{x}^i)$.

Step 2: Approximation of heteroscedastic variance by GP$_s$: Train a GP model that approximates the response variance across the input domain using the observation set $\{\mathbf{x}^i, (s^2)^{i(k)}; i = 1, .., N_t, k = 1, \dots, n^i\}$. Similar to the *PR-SE* algorithm (Section 3) the logarithm of the variance can be approximated. The GP$_s$ follows the Ho-R formulation (Table A1) discussed in Appendix A utilizing output selection $\upsilon = \ln(\tau^2)$, number of training points $N_s = N_t$, and provides approximation $\ln(\tilde{\tau}^2(\mathbf{x}))$ given by Equation (A2). Formulation entails the hyper-parameter calibration given by Equation (A6) utilizing observation set composed of both the replicated and non-replicated samples, $\{\mathbf{x}^i, (s^2)^{i(k)}; i = 1, \dots, N_r; k = 1, \dots, n_r\}$ and $\{\mathbf{x}^i, (s^2)^i; i = N_r + 1, \dots, N_t\}$, respectively.

Step 3: Estimation of the response by GP$_y$: Using $\tilde{\tau}^2(\mathbf{x})$ as unscaled variance component, formulate heteroscedastic GP model that approximates the response distribution across the domain of interest utilizing all available observations $\{\mathbf{x}^i, y^{i(k)}; i = 1, \dots, N_t, k = 1, \dots, n^i\}$. Implementation is identical to Step 3 of *PR-SE* algorithm.

The major difference between the variance estimators of the original (GP$_\tau$ in *PR-SE*) and alternative (GP$_s$ in *ER-SE*) stochastic emulation formulations is that the original [utilizing Equation (2) as observations] relies on the pure sample variance, while the modified one [utilizing Equation (3) as observations] treats the mean estimate obtained from the tertiary GP model as the true mean (population mean) and approximates variance as the sample squared deviation from the mean. This difference allows the latter to accommodate the non-replicated samples for approximating both the heteroscedastic behavior (Step 2) and the final EDP distribution (Step 3). It also means that formulation can be adopted without any replications. These characteristics address the two challenges identified in Section 3. Figure 2 summarizes the utilized information in each step of *PR-SE* and *ER-SE*, illustrating that only the latter fully utilizes the information from all samples in all three steps. Note that while *PR-SE* can be used without any non-replicated samples, the best practice suggested in previous literature[32] is to facilitate both replicated and non-replicated samples.

Of course, it should be noted that the proposed method is heuristic in a way that it assumes that the mean estimate obtained from the crude homoscedastic nugget assumption, $\hat{m}(\mathbf{x})$, gives a reasonably reliable estimate. This assumption is justified by past studies,[22,32] that have shown that for seismic applications the mean-field prediction is relatively robust to the choice of the variance model, especially compared to the level of variability observed in the sample replications. This was also observed in more general, non-seismic, applications investigated in ref. [30]. Furthermore, note that if the final estimation of mean $\tilde{y}(\mathbf{x})$ (from Step 3) is significantly different from the initial estimate $\hat{m}(\mathbf{x})$ (from Step 1), then one may iterate Step 2–3 by replacing $\hat{m}(\mathbf{x})$ with the $\tilde{y}(\mathbf{x})$ until the two functions become more similar. Note that, as mentioned in ref. [30], this iteration is rather heuristic and is not guaranteed to converge.

A final remark is warranted for the Gaussianity of the observations and potential prediction bias for the GP$_s$ and GP$_\tau$ surrogate models. The sample variance follows a chi-squared distribution with $k = n^i$ degrees of freedom[46] and therefore the output observation used in GP$_s$ (for *ER-SE*) are expected to be more skewed than the observations used for GP$_\tau$ (for *PR-SE*), deviating more from the underlying Gaussianity assumption. Nevertheless, the mean estimate of each GP is known to be the best linear unbiased estimator (BLUE) for given basis functions and covariance kernel assumption, even when the Gaussianity condition is violated. However, the log-transformation in Step 2 may introduce bias when converting the predictions back to the original space to obtain $\tilde{\tau}^2(\mathbf{x})$. This bias is partially accounted for when estimating $\tau^2(\mathbf{x})$ through the calibration of the variance scaling parameter δ as detailed in Appendix A, that is, the average of variance across the domain is unbiased. This applies to both the *ER-SE* and *PR-SE* formulations, though it is expected to be of a greater utility for the case with larger potential bias *ER-SE*, due to the larger deviation of the observations from the underlying Gaussianity assumption.

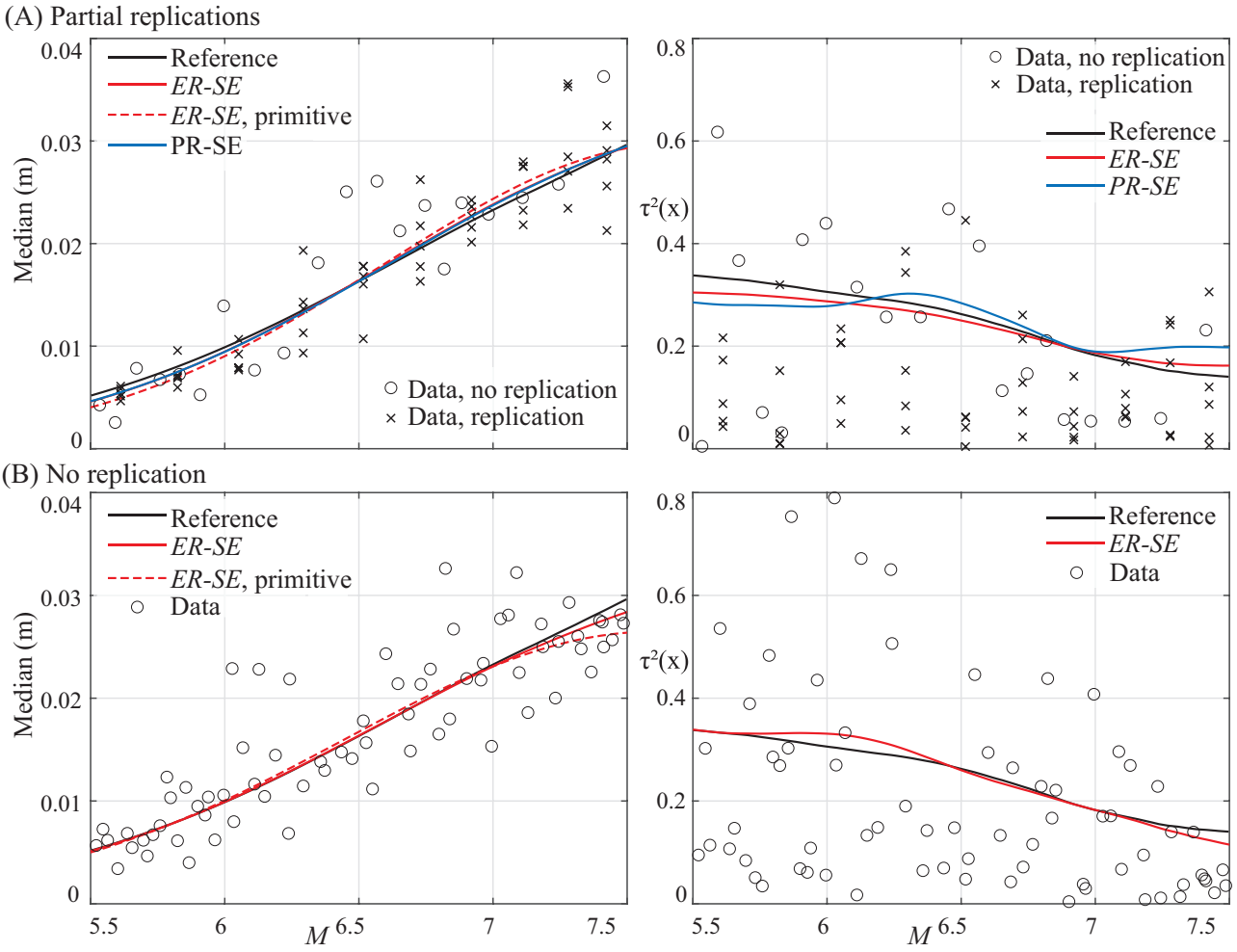

FIGURE 3 Displacement EDP statistics (left column for median and right column for logarithmic standard deviation) for implementations with replications (top row) and without replications (bottom row). Both the reference results and the approximation for *PR-SE* and *ER-SE* are shown. EDP, engineering demand parameter.

4.2 | Illustration for single degree of freedom application

To illustrate the key differences in the *PR-SE* and *ER-SE* implementations NLRHA for a single degree of freedom (SDoF) oscillator is utilized. The SDoF has a modal damping equal to 3% and a restoring force that corresponds to an elastic-perfectly plastic hysteresis model with yield displacement $\delta_y = 2$ cm and pre-to-post yield ratio of 0.3. For describing the acceleration time-history seismic excitation a stochastic point source model[47] is utilized, with characteristics identical to the ones adopted in study.[11] This model provides sample realizations of the excitation using as inputs the moment magnitude M, and the rupture distance, r_{rup}. The range for M is considered to be [5.5 7.5], while r_{rup} is taken to be 20 km. For the EDPs of interest, both the peak SDoF displacement and peak absolute acceleration are considered and results are presented for the median and logarithmic standard deviation. Reference estimates for these statistics are obtained using a large number of $n_r = 300$ replications to accomplish high accuracy for the statistical estimates. The predictions for the EDP distribution statistics using *PR-SE* and *ER-SE* for a total of $N_{sim} = 70$ simulations are examined. For *PR-SE,* 71.4% of the simulations corresponds to replications with $n_r = 5$, leading to $N_t = 20$, $N_r = 10$. For *ER-SE,* two setups are considered, the first one is identical to the one examined for *PR-SE* and the second uses no replications. Training points are obtained using space-filling sampling form M, yielding independent set of samples for the two *PR-SE* variants. Figure 3 presents the estimates of the median (right column) and the logarithmic standard deviation (left column) for the displacement EDP for the *PR-SE* and *ER-SE* implementation with replications (top row) or the *ER-SE* implementation with no replications (bottom row). The observation data utilized for the surrogate model development are also shown, with training points corresponding to replications depicted separately. Additionally, the primitive estimation of median obtained after Step

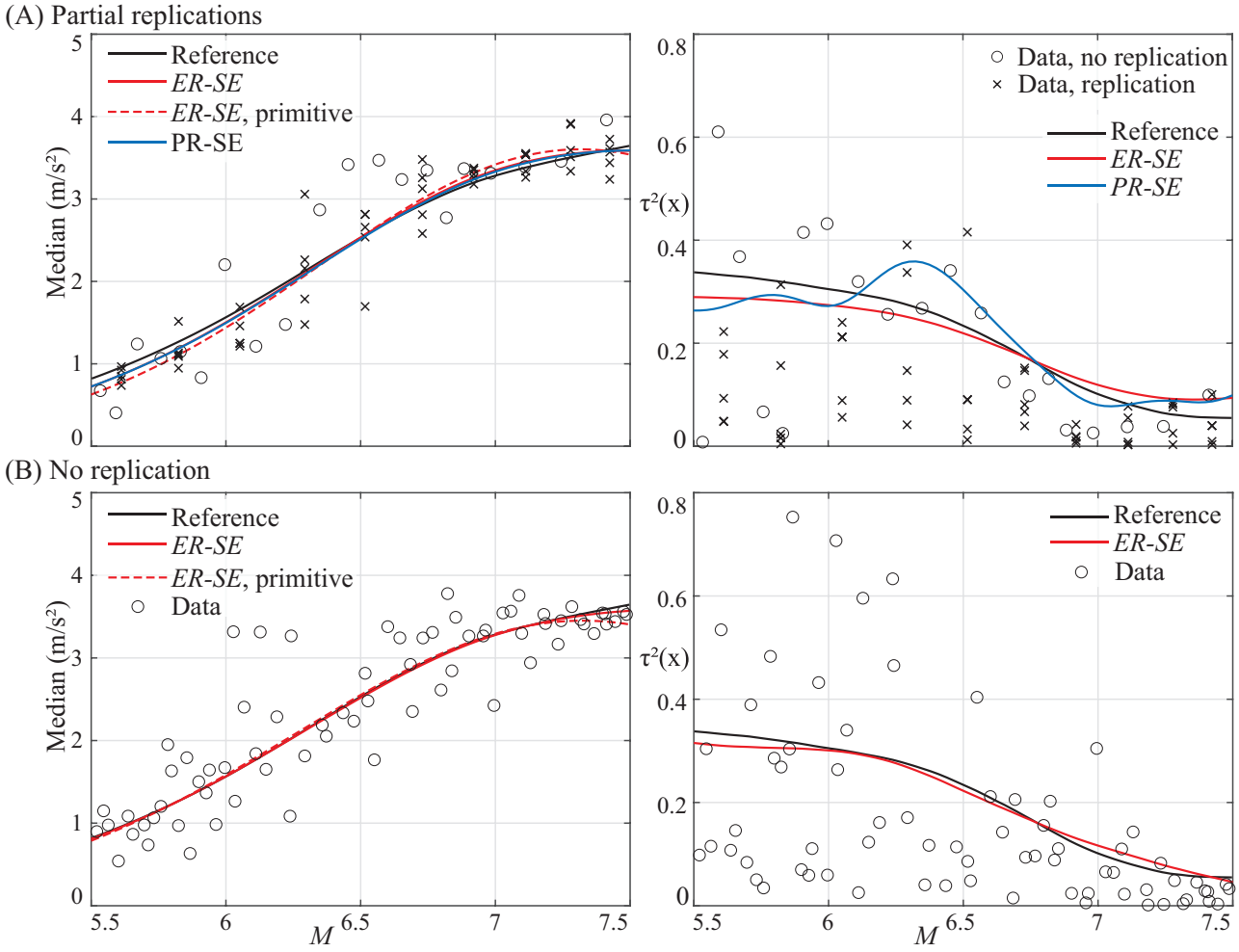

FIGURE 4 Acceleration EDP statistics (left column for median and right column for logarithmic standard deviation) for implementations with replications (top row) and without replications (bottom row). Both the reference results and the approximation for *PR-SE* and *ER-SE* are shown. EDP, engineering demand parameter.

1 of *PR-SE* implementation (under homoscedastic assumption) is shown as the dotted line. Figure 4 presents the same results for the acceleration EDP. Figure 5 presents the curves corresponding to the median and median ± one logarithmic standard deviation for both EDPs (column of figure) for the same surrogate model cases presented earlier in Figures 3 and 4. Figure 5 facilitates ultimately a more direct comparison for the distribution results based on the individual statistics reported in Figures 3 and 4.

Results verify first of all the trends reported in ref. [32], indicating significant aleatoric uncertainty in the hazard description and heteroscedastic behavior, with level of variance in EDPs strongly dependent on seismic intensity. All implementations yield good accuracy for the median EDP predictions, but big differences can be observed for the logarithmic standard deviation estimates (right column of Figures 3 and 4), and, ultimately, the overall probabilistic characterization of the aleatoric response variability (Figure 5). With respect to the predictions for the EDP variability (logarithmic standard deviation) *ER-SE* slightly outperforms *PR-SE* under the partial replication setting, while the accuracy substantially improves for the *ER-SE* implementation without any replications. The comparisons for the partial replication setting show that, as expected, the enrichment of the sample coverage allowed by the use of non-replicated samples for the secondary surrogate model in *ER-SE* can offer tangible benefits. The comparison for the setting with no replications demonstrates that this coverage benefit can be further extended by removing replications altogether. More importantly, comparisons for both the EDP median and logarithmic standard deviation, reveal that there is no evident penalty in avoiding (or not being able – as discussed earlier) to use replications.

Finally, it is important to note that even under homoscedasticity assumption (*PR-SE* primitive case) the median EDP predictions appear to be quite accurate, especially considering the sample variability arising from high aleatoric

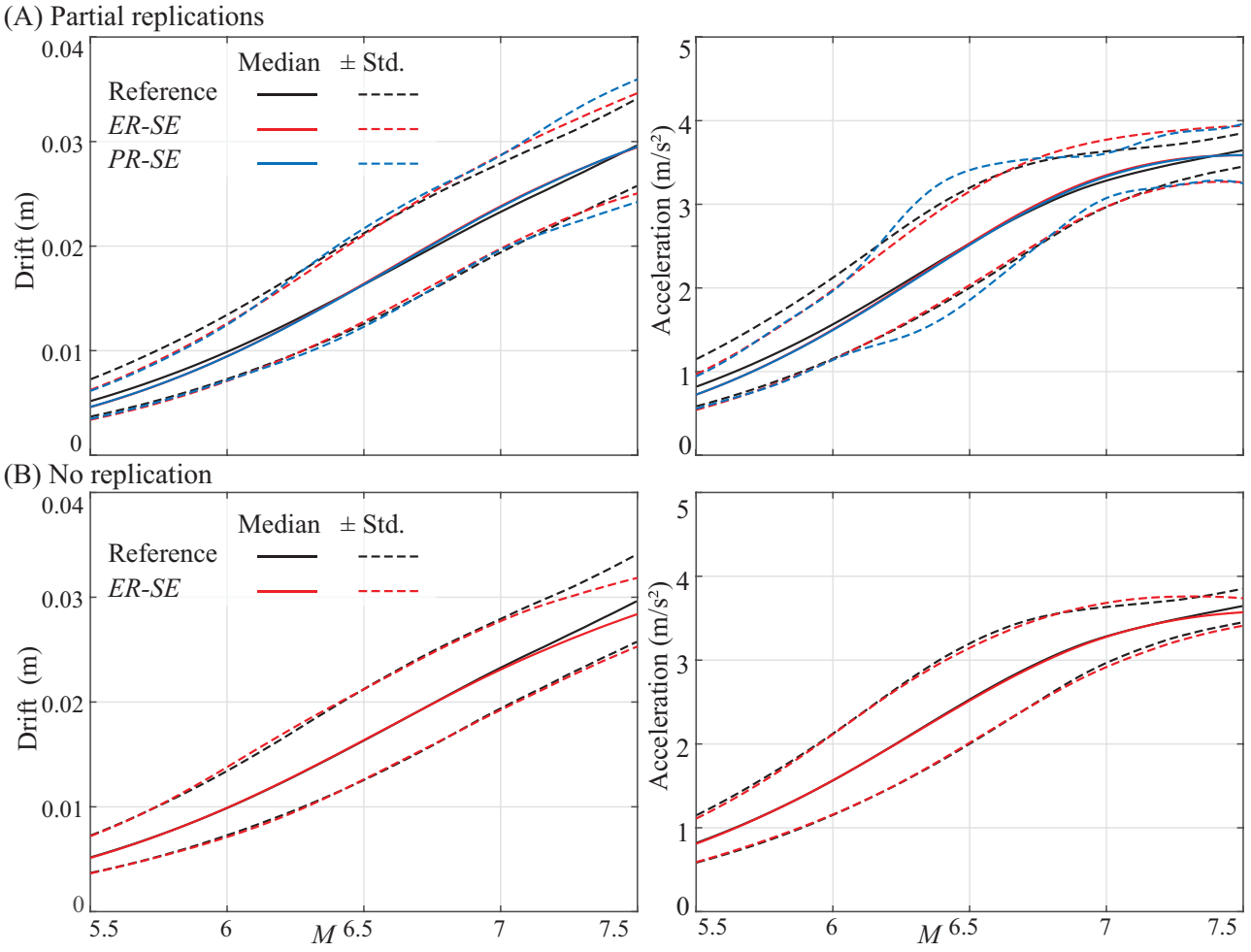

FIGURE 5 EDP response (left column for displacement and right column for standard deviation) for implementations with replications (top row) and without replications (bottom row). Both the reference results and the approximation for *PR-SE* and *ER-SE* are shown for the median and the median ± one logarithmic standard deviation. EDP, engineering demand parameter.

uncertainty. This observation aligns well with the core assumption taken in *PR-SE* algorithm that allows the replacement of $m(\mathbf{x}^i)$ with $\hat{m}(\mathbf{x}^i)$ in Equation (3). Of course, the final *PR-SE* predictions improve upon this primitive mean, indicating that all suggested steps in the *PR-SE* implementation are essential components.

5 | ILLUSTRATIVE EXAMPLE

The illustrative example revisits the examples discussed in refs. [22, 32] to allow a direct comparison between the original (*PR-SE*) and updated (*ER-SE*) stochastic emulation algorithms across different settings for quantifying seismic hazard and the underlying aleatoric variability. The example considers the approximation of the distribution for peak drift and absolute floor acceleration EDPs for a three-story concrete moment resisting frame (MRF), with seismic excitation described through different stochastic ground motion models.

5.1 | Structural model

The model for the three-story concrete MRF building is described in detail in ref. [48]. The uncertain structural model parameters[22,32] include the damping ratio ζ_s and the characteristics of the constitutive laws for the concrete and steel materials: (1) the maximum compressive stress f_c and the strain at maximum stress ε_c for concrete, and (2) the modulus of

TABLE 1 Structural parameter input domain for the surrogate model development and the associated probability distributions used in the risk assessment.

Input parameter	Input domain for surrogate model formulation	Probability model for risk assessment		
		Distribution	Median	Coefficient of variation
f_{c1}, f_{c2}, f_{c3}	[17,000 46,100] (kPa)	Lognormal	28,000 (kPa)	20%
$\varepsilon_{c1}, \varepsilon_{c2}, \varepsilon_{c3}$	[0.0012 0.0033] (-)	Lognormal	0.002 (-)	20%
E_s	$[1.8\ 2.3]\cdot 10^8$ (kPa)	Lognormal	$2\cdot 10^8$ (kPa)	5%
f_y	[430,000 700,000] (kPa)	Lognormal	550,000 (kPa)	10%
a	[0.006 0.017] (-)	Lognormal	0.01 (-)	20%
ζ_s	[3 8.2] (%)	Lognormal	5 (%)	20%

elasticity E_s, the yielding stress f_y and the strain hardening ratio a for steel. Common characteristics across all floors are assumed for the steel and different ones for each floor for the concrete, following standard realistic construction applications, with f_{c1}, f_{c2} and f_{c3} denoting the compressive strength for the first, second and third floors, respectively, and ε_{c1}, ε_{c2} and ε_{c3} denoting the strain at maximum stress for the same floors. This leads to $\mathbf{x}_s = [f_{c1}, f_{c2}, f_{c3}, \varepsilon_{c1}, \varepsilon_{c2}, \varepsilon_{c3}, E_s, f_y, a, \zeta_s]$. The input domain for each of these parameters is reviewed in Table 1 and is chosen so that it extends at least 2.5 standard deviations from the median of the probability distribution—also shown in Table 1—which is used later on in the seismic risk assessment implementation (Section 5.5). This guarantees that the domain of the surrogate model formulation is broad enough for the intended purposes of its use.

5.2 | Excitation models and aleatoric uncertainty description

The seismic excitation is described using a stochastic ground motion model, established[47,49,50] by modulating a high-dimensional stochastic sequence through functions that offer the desired frequency and time-domain characteristics for the earthquake acceleration time-history (excitation). These functions are parameterized through a group of *functional* variables (defining the ground motion model features), such as ground motion duration and arias intensity. For the probabilistic hazard quantification these variables are subsequently connected through predictive relationships to seismicity and local site characteristics, such as the moment magnitude, M, the rupture distance, r_{rup}, the fault type, and the shear wave velocity for the top 30 meters of soil V_{s30}. Notations $\mathbf{x}_g$ and $\mathbf{x}_m$ will be used herein to describe, respectively, the vectors describing (i) the ground motion features and (ii) the seismicity characteristics. The predicting models relating $\mathbf{x}_g$ to $\mathbf{x}_m$ can be purely deterministic[47] or correspond to a probabilistic mapping $p(\mathbf{x}_g|\mathbf{x}_m)$.[49,50]

For selecting the EVs, $\mathbf{x}_h$, used to describe the seismic hazard in surrogate modeling context, three alternative options exist.[22] These result in different definitions for the input $\mathbf{x}_h$ (and subsequently for $\mathbf{x}$), and to different selections for the source of aleatoric variability. As discussed in detail in ref. [22] the alternatives should be carefully evaluated in each application in order to limit the dimensionality of $\mathbf{x}_h$ while maintaining a lognormal distribution under the influence of the aleatoric variability. The latter is critical for accommodating the underlying Gaussian assumption (for addressing aleatoric variability) invoked within the GP-based stochastic emulation strategy. The first option, denoted EV_1, is to choose $\mathbf{x}_h = \mathbf{x}_m$, that is, EVs consisting of only the seismicity characteristics, with aleatoric uncertainty in the hazard description corresponding to both the stochastic sequence utilized in the ground motion model and the variability in the probabilistic mapping $p(\mathbf{x}_g|\mathbf{x}_m)$. This option reduces the dimensionality of the input but might face challenges related to the underlying assumption of Gaussian aleatoric distribution. The second option, denoted EV_2, is to choose $\mathbf{x}_h = \mathbf{x}_g$, that is, EVs consisting of only the ground motion features, with aleatoric uncertainty originating strictly from the stochastic sequence. Note that if $\mathbf{x}_g$ is known then the stochastic ground motion model is completely defined, and one does not need to additionally know $\mathbf{x}_m$ in order to create a realization of the excitation, in other words $\mathbf{x}_g$ is statistically sufficient to define the EDP distribution $p(z|\mathbf{x}_g, \mathbf{x}_s) = p(z|\mathbf{x}_m, \mathbf{x}_g, \mathbf{x}_s)$. The Gaussian distribution assumption for the aleatoric variability has been shown to consistently hold in this case (with aleatoric variability originating only from stochastic sequence), but, depending on the ground motion characteristics, the dimensionality of the input might prohibitively increase. The final option, denoted EV_3, is to consider a partitioning $\mathbf{x}_g = [\mathbf{x}_g^1 \mathbf{x}_g^2]$ and choose $\mathbf{x}_h = [\mathbf{x}_m \mathbf{x}_g^1]$, that is, EVs composed of the seismicity characteristics and a subset of the ground motion features, with aleatoric uncertainty corresponding to both the stochastic sequence and the variability in the probabilistic mapping for the remaining ground motion features $p(\mathbf{x}_g^2|\mathbf{x}_m, \mathbf{x}_g^1)$. By appropriate

TABLE 2 Different excitation models considered in the numerical examples.

Case	Ground motion model	Appropriate selection of explanatory variables[22,32]	
		Variable composition	Selected variables ($\mathbf{x}_h$)
GM_1	Point source stochastic model[47]	Seismicity parameters (EV_1)	M, r_{rup}
GM_2	Rezaeian and Der Kiureghian[49]	Ground motion parameters (EV_2)	$I_a, D_{5-95}, t_{mid}, \omega_{mid}, \omega', \zeta_d$
GM_3	Vlachos, Papakonstantinou and Deodatis[50]	Seismicity and a subset of ground motion parameters (EV_3)	M, r_{rup}, I_a, Q_d

Abbreviation: EV, explanatory variable.

partitioning of $\mathbf{x}_g$, EV_3 can reduce the dimensionality of the input while also maintaining an underlying Gaussian aleatoric distribution. Further details for these selections, including extensive discussion on the relevant implications when the surrogate model is subsequently used to quantify seismic risk can be found in ref. [22]. Note that if predictive models are deterministic, then only EV_1 exists, with aleatoric uncertainty in this instance corresponding strictly to the stochastic sequence utilized in the ground motion model[32] and Gaussianity assumption illustrating good agreement. To establish comprehensive comparisons three different stochastic ground motion models covering all aforementioned options for the EV selection are considered here, reviewed in Table 2. The assumption of lognormal EDP distribution under the assumed (in each case) aleatoric variability has been shown to provide high statistical accuracy for all three options in previous studies.[22,32]

The first ground motion model, denoted GM_1 corresponds to a point source stochastic model,[47] with deterministic predictive relationships. The ground motion model characteristics identical to the ones adopted in study.[11] EV_1 is adopted for the EVs with both the moment magnitude, M, and the rupture distance, r_{rup}, considered as input parameters, leading to $\mathbf{x}_h = \{M, r_{rup}\}$. The input domain for M is chosen as [5 8] and for r_{rup} as [3 60] (km). This setup is identical to the one considered in ref. [32], and was chosen based on the intended probability models used for the risk assessment in that study.

The second ground motion model, denoted GM_2 corresponds to the record-based stochastic ground motion model developed by Rezaeian and Der Kiureghian.[49] EV_2 is adopted for the EVs with ground motion features corresponding to the arias intensity I_a, the significant duration $D_{5\text{-}95}$, the time corresponding to 50% of the intensity t_{mid} the associated median spectral frequency ω_{mid}, the rate of change for that frequency ω', and the damping ratio ζ_d for the excitation spectrum, leading to $\mathbf{x}_h = \{I_a, D_{5-95}, t_{mid}, \omega_{mid}, \omega', \zeta_d\}$. The predictive relationships for this model, relating $\mathbf{x}_m$ to $\mathbf{x}_g$ consisting of M, r_{rup}, V_{s30} and fault type, can be found in ref. [49]. The input domain for $\mathbf{x}_h$ is defined based on these predictive relationships within the range of model applicability,[49] M in [6 8] and r_{rup} in [10 60] (km), and deterministic values for $V_{s30} = 760$ m/s and strike slip fault. This setup is identical to the one considered in ref. [22], and was chosen based on the intended probability models used for the risk assessment in that study.

The third ground motion model, denoted GM_3 corresponds to the record-based stochastic ground motion model developed by Vlachos, Papakonstantinou and Deodatis.[50] The ground motion model parameter vector $\mathbf{x}_g$ in this case is composed of the arias intensity I_a, the duration of the excitation T_d, the shape c_d and scale d_d parameters for the energy accumulation envelope, and for each of the two modes of a bi-modal excitation power spectrum ($e = 1,2$), the damping ratio $\{\zeta_d^{(e)}\}$, the three parameters for determining the modal participation of each of the modes $\{F_d^{(e)}, \mu_d^{(e)}, \sigma_d^{(e)}\}$ and the three parameters defining the frequency variation $\{Q_d^{(e)}, a_d^{(e)}, b_d^{(e)}\}$ where $Q_d^{(e)}$ denotes the median frequency and set $\{a_d^{(e)}, b_d^{(e)}\}$ defines the energy-based variation. The predictive relationships for this model, relating $\mathbf{x}_m$ to $\mathbf{x}_g$ consisting of M, r_{rup}, and V_{s30}, can be found in ref. [50]. To reduce input dimensionality, EV_3 is adopted for the EVs with $\mathbf{x}_h = \{M, r_{rup}, I_a, Q_d\}$ following recommendations in ref. [22]. The input domain for $\mathbf{x}_h$ is defined as M in [6 8], r_{rup} in [10 60] (km) while for $\{I_a, Q_d\}$ this domain is chosen, similarly to GM_2 based on the underlying predictive relationships.[50] As in the case for GM_2, this setup is identical to the one considered in ref. [22], and was chosen based on the intended probability models used for the risk assessment in that study.

5.3 | Implementation details and validation metrics

The three excitation models discussed in Section 5.2 lead to three different surrogate model formulations. The total number of model parameters, composed of $\mathbf{x}_s$ and $\mathbf{x}_h$, is $n_x = 12$ for GM_1, $n_x = 16$ for GM_2, and $n_x = 14$ for GM_3. The investigation focuses on the comparison of the baseline *PR-SE* approach discussed in Section 3 and the proposed here modification *ER-SE* discussed in Section 4. Since both accommodate partial replications, we further investigated the benefit offered

from replications by exploiting different mixtures of replicated and non-replicated training samples. In particular, two different values of replication size ($n_r = 5$ and 10) are investigated, as well as the varying numbers of the total simulations ($N_{sim} = 1500$, 1000, and 500). Similarly, chosen cases of N_n are $N_n = 200$, 400, 600, 800, 1000, and 1500. Among the listed N_n and N_{sim} values, the pairs that satisfy $N_n < N_{sim}$ are investigated in the original *PR-SE* implementation, and $N_n = N_{sim}$ cases (i.e., no replications) are additionally included for the proposed *ER-SE* implementation, as the latter became applicable after the enhancements offered in Section 4. Note that, for example, when ($N_{sim} = 1500$, $N_n = 200$, $n_r = 10$) is chosen, the number of replicated points was $N_r = (N_{sim}\text{-}N_n)/n_r = 130$. Similarly, when ($N_{sim} = 1500$, $N_n = 1500$) is chosen, regardless of n_r, all the samples in the training sets are non-replication points and N_r is zero. The training points are selected using space filling Latin Hypercube sampling in the input domains X defined in Sections 5.1 and 5.2, while the formulation is repeated 15 times in each case for different realizations of the training samples. Linear basis functions are utilized for the influential input parameters, corresponding to the moment magnitude and/or arias intensity depending on the surrogate model variant (depending on whether these parameters are included in the EVs definition or not), while a generalized exponential correlation function was used for the correlation kernel. Hyper-parameter calibration is performed, as discussed in Appendix A, using a pattern-search algorithm. The value of the nugget was chosen to be larger than 1, reflecting the large degree of aleatoric variability in the seismic hazard description and the additional statistical variability that comes from the estimation of sample values for the mean and squared deviation from the mean using a limited number of replications. The latter is especially true for the no-replication case.

Following ref. [22], to facilitate consistent comparison across the variant implementations, the validation is established with respect to the conditional EDP distribution (or statistics) for a specific seismic scenario and structural configuration $\{\mathbf{x}_s, \mathbf{x}_m\}$, as this distribution can be estimated across all of them, and moreover this is a distribution that is frequently explicitly encountered in seismic loss assessment and performance based earthquake engineering applications.[51,52] To simplify discussions, notation $\mathbf{x}_v = \{\mathbf{x}_s, \mathbf{x}_m\}$ will be also used to describe the input vector used in the validation stage. The surrogate model for explanatory variables EV_1 provides directly the approximation of the distribution for $y = \ln(z)$, but for EV_2 and EV_3, this conditional distribution is obtained through the total probability theorem, respectively, as:

$$\tilde{p}(\ln(z)|\mathbf{x}_v) = \tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_m) = \int \tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_g) p(\mathbf{x}_g|\mathbf{x}_m) d\mathbf{x}_g \tag{4}$$

$$\tilde{p}(\ln(z)|\mathbf{x}_v) = \tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_m) = \int \tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_m, \mathbf{x}_g^1) p(\mathbf{x}_g^1|\mathbf{x}_m) d\mathbf{x}_g^1 \tag{5}$$

where $\tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_g)$ in Equation (4) and $\tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_m, \mathbf{x}_g^1)$ in Equation (5) are the distributions approximated in the respective implementations for selection of $\mathbf{x}_h = \mathbf{x}_g$ (for EV_2) and $\mathbf{x}_h = [\mathbf{x}_m \mathbf{x}_g^1]$ (for EV_3), respectively. Leveraging the computational efficiency of the underlying surrogate model the desired conditional distributions and statistics can be readily calculated using MCS with large number of samples.[22]

Validation is performed using a test-sample setting, considering $N_v = 1000$ samples obtained through Latin Hypercube Sampling (LHS) within X, $\{\mathbf{x}_s^j, \mathbf{x}_m^j; j = 1, \ldots, N_v\} = \{\mathbf{x}_v^j; j = 1, \ldots, N_v\}$. As in refs. [22, 32] the surrogate model predictions are then compared to reference results obtained using Monte Carlo estimates corresponding to $n_r = 200$ replications. To evaluate the overall prediction error in the response distribution, the Kullback–Leibler (KL) divergence measure is used, quantifying the discrepancy between the estimated and reference probability distributions, as in ref. [53]. The averaged KL value over the N_v inputs is utilized as validation metric, given by:

$$AD_{KL} = \frac{1}{N_v} \sum_{j=1}^{N_v} \int_{\mathbb{R}^+} \log \left[\frac{p(\ln(z)|\mathbf{x}_v^j)}{\tilde{p}(\ln(z)|\mathbf{x}_v^j)} \right] p(\ln(z)|\mathbf{x}_v^j) d\ln(z) \tag{6}$$

Details for estimation of the integral in Equation (6), including establishing the Monte Carlo approximation for $p(\mathbf{z}|\mathbf{x}_s, \mathbf{x}_m)$ are included in ref. [22]. The estimation of $\tilde{p}(\ln(z)|\mathbf{x}_v)$ based on the stochastic emulation predictions is discussed in Appendix B. Note that the KL divergence is invariant to transformation for the EDP, so whether it is calculated as difference of the distributions for z or $\ln(z)$ the result is identical. Estimation for $\ln(z)$ as detailed above is preferred, as this avoids any transformation for the stochastic emulation Gaussian predictions. The KL divergence values are dimensionless and have only a lower bound (zero), attained when the compared distributions are identical. To better frame the divergence values, the reference comparisons discussed in ref. [32] may be used. In the case that the two distributions examined are Gaussians and have the same mean, but their standard deviations differ by 0.1, then the KL divergence value is 0.14. On the other hand, if the standard deviation is the same, but the mean differs by 0.1 standard deviations, then the

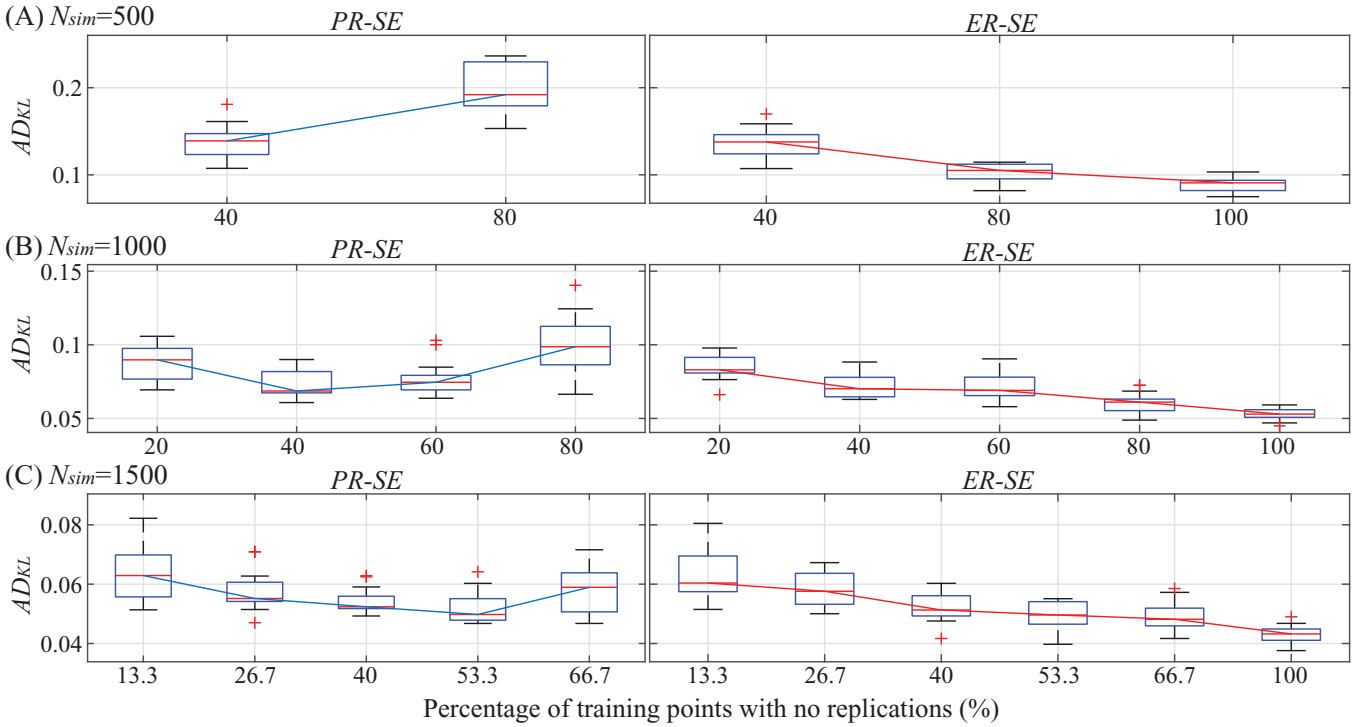

FIGURE 6 Averaged KL divergence across all EDPs for *PR-SE* (left column) and *ER-SE* (right column) implementations for ground motion GM_1 with explanatory variables EV_1. The validation results are shown for different percentage of samples with no replications (abscissas) and for N_{sim} equal to (A) 500, (B) 1000, and (C) 1500, with $n_r = 10$. The results are presented as boxplots, with boxes indicating 25% and 75% sample quantile, red lines the medians, crosses the outliers, and whiskers the last non-outlier samples. EDP, engineering demand parameter; KL, Kullback–Leibler.

KL divergence value is 0.1. If the discrepancy is 0.05 instead of 0.1, the KL divergence values for the two aforementioned cases are 0.07 and 0.051, respectively.

Additionally, the first and second order EDP statistics, corresponding to median and logarithmic standard deviation are compared, utilizing as validation metric the correlation coefficient (*cc*) between surrogate model predictions and reference values. For the logarithmic standard deviation, the *cc* over the test-sample is estimated as:

$$cc_s = \frac{\sum_{j=1}^{N_v}\left(\tilde{s}\left(\mathbf{x}_v^j\right)-\bar{\tilde{s}}\left(\mathbf{x}_v^j\right)\right)\left(s\left(\mathbf{x}_v^j\right)-\bar{s}\left(\mathbf{x}_v^j\right)\right)}{\sqrt{\sum_{j=1}^{N_v}\left(\tilde{s}\left(\mathbf{x}_v^j\right)-\bar{\tilde{s}}\left(\mathbf{x}_v^j\right)\right)^2\sum_{j=1}^{N_v}\left(s\left(\mathbf{x}_v^j\right)-\bar{s}\left(\mathbf{x}_v^j\right)\right)^2}}$$

$$\text{where}\quad \bar{\tilde{s}}\left(\mathbf{x}_v^j\right)=\frac{1}{N_v}\sum_{j=1}^{N_v}\tilde{s}\left(\mathbf{x}_v^j\right)\ \text{and}\ \bar{s}\left(\mathbf{x}_v^j\right)=\frac{1}{N_v}\sum_{j=1}^{N_v}s\left(\mathbf{x}_v^j\right) \tag{7}$$

with $s(\mathbf{x}_v^j)$ corresponding to the reference statistics and $\tilde{s}(\mathbf{x}_v^j)$ corresponding to the emulation-based estimate. Identical equation holds for the *cc* of the median EDP, cc_m. Details for the estimation of the relevant statistical quantities for the different stochastic emulation variants are included in Appendix B.

5.4 | Validation results and Discussion

The validation results are presented in Figures 6–15. Figures 6–12 compare the averaged KL divergence AD_{KL} for different EDP types and ground motion models obtained by *ER-SE* and *PR-SE* implementations. The variability across the 15 different implementation trials (repetitions), corresponding to different random selection of training points, is depicted in these figures in the form of boxplots. The validation results are shown for different percentages of training points with no

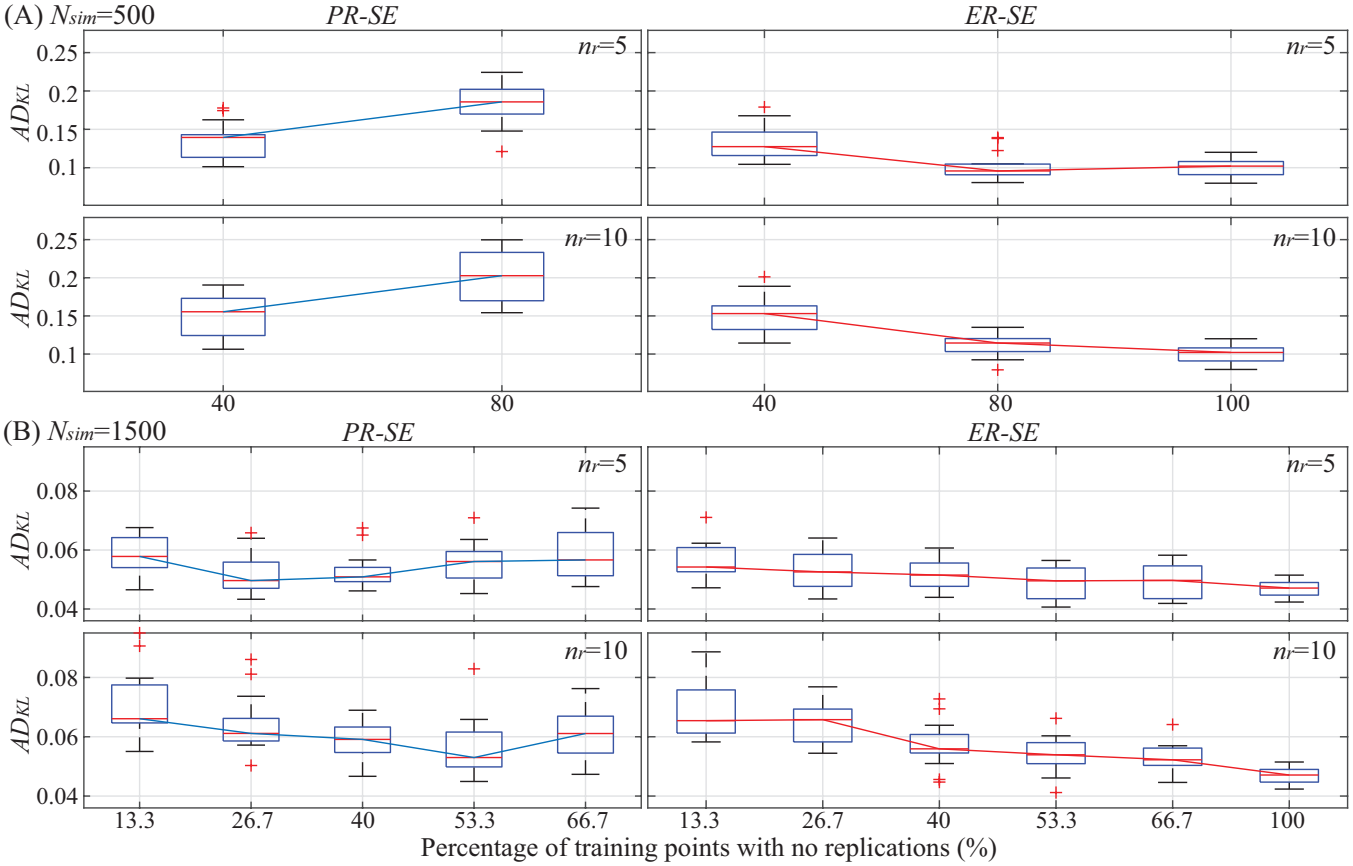

FIGURE 7 Averaged KL divergence across all floors for inter-story drift EDPs for *PR-SE* (left column) and *ER-SE* (right column) implementations for ground motion GM_1 with explanatory variables EV_1. The validation results are shown for different percentage of samples with no replications (abscissas), for N_{sim} equal to (A) 500 and (B) 1500, and for n_r equal to 5 (first row) and 10 (second row). Box plots characteristics are identical to the ones in Figure 6. EDP, engineering demand parameter; KL, Kullback–Leibler.

replications (N_n/N_{sim}×100), where $N_n \in [200, 400, 600, 800, 1000, 1500]$ and $N_{sim} \in [500, 1000, 1500]$ as previously discussed, representing different computational budget allocation scenarios. Note that 100% (i.e., "no replication" implementation) is presented only for *ER-SE* approach because the implementation without replication cannot be accommodated by the previous *PR-SE* implementation. Figure 6 shows the results averaged across all EDPs for GM_1 with EV option EV_1 and for all three considered N_{sim} values for $n_r = 10$. This figure allows an easier comparison of the performance for different cases with respect to the total computational budget, but averages across all EDPs. In contrast, the remaining figures present results separately for the average (across floors) inter-story drift and absolute acceleration EDPs, but only for selective cases for N_{sim}. Figures 7, 9, 11 present results for the drift EDPs for the three ground motion models considered in Table 2, respectively, while Figures 8, 10, 12 present results for the absolute floor acceleration EDPs. Additionally, the effect of replication sizes (n_r) is investigated in Figures 7–12 showing results for both $n_r = 5$ and $n_r = 10$ values. The remaining Figures 13–15 present the *cc* validation metrics for the predicted median cc_m (top row) and logarithmic standard deviation cc_s (bottom row) for both drift (left column) and acceleration (right column) EDPs for different N_{sim} values for both *ER-SE* and *PR-SE* implementations for $n_r = 10$. Each of these figures corresponds to a different ground motion (and EV) selection from the ones reviewed in Table 2. For these figures, only the mean performance across 15 repetitions is shown, and the EDP results are averaged across all floors. To allow easier comparisons the same scale is used in the *y*-axis for the implementations that use the same computational budget (same N_{sim}) but to allow for better resolution across the different variants examined for each N_{sim} that scale is modified across the different N_{sim} cases.

Results in Figure 6–12 show that all observed cases show good prediction accuracy for $N_{sim} = 1500$, with AD_{KL} values being consistently smaller than 0.1. As expected, the prediction accuracy for the target distribution reduces (AD_{KL} increases) as N_{sim} reduces, but still, the majority of results (apart from some cases for the *PR-SE* implementation) are lower than 0.15. The *cc* results in Figures 13–15 show excellent match to the reference for the predicted median response

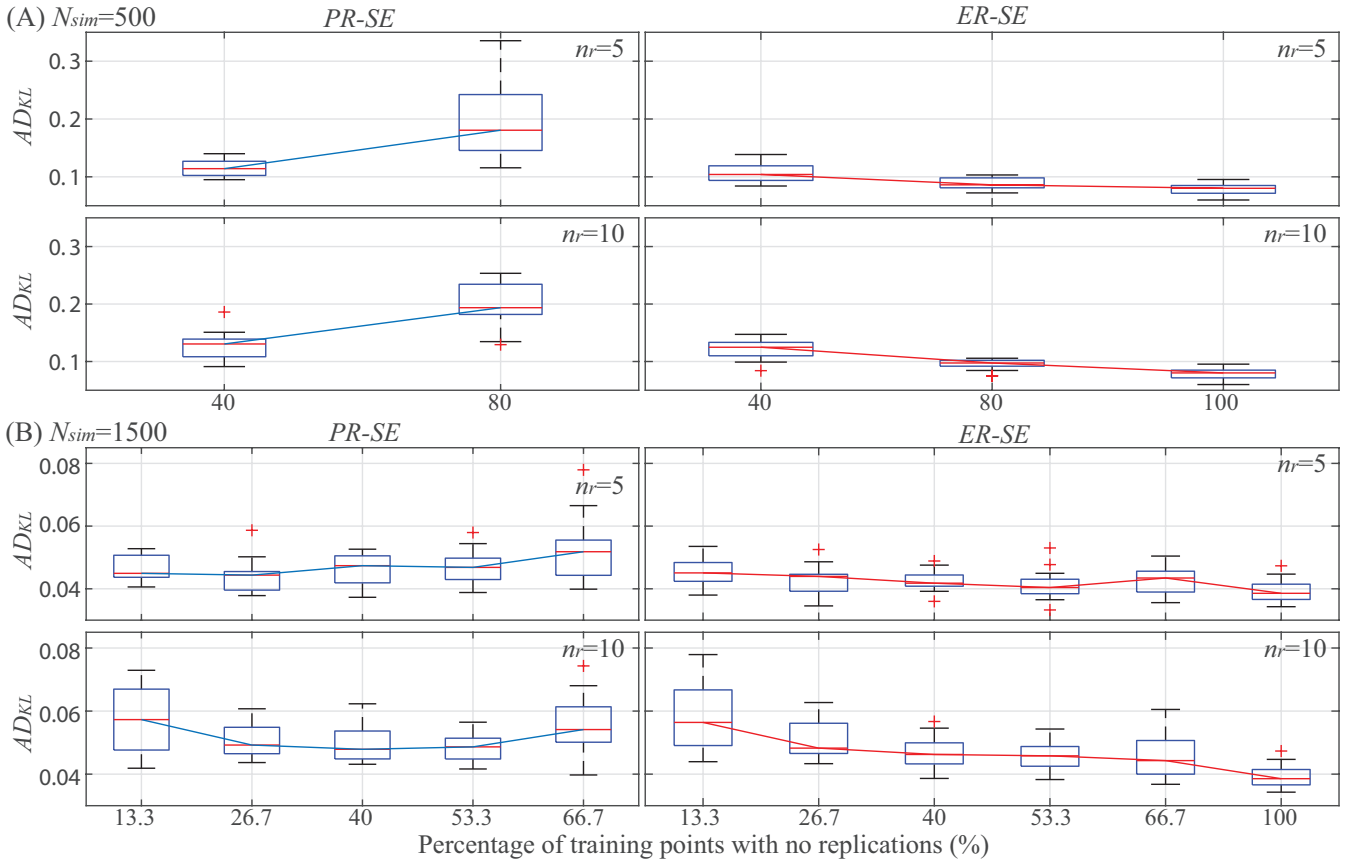

FIGURE 8 Averaged KL divergence across all floors for absolute floor acceleration EDPs for *PR-SE* (left column) and *ER-SE* (right column) implementations for ground motion GM_1 with explanatory variables EV_1. The validation results are shown for different percentage of samples with no replications (abscissas), for N_{sim} equal to (A) 500 and (B) 1500, and for n_r equal to 5 (first row) and 10 (second row). Box plots characteristics are identical to the ones in Figure 6. EDP, engineering demand parameter; KL, Kullback–Leibler.

(values of cc_m close to 1) but larger differences for the logarithmic standard deviation (lower cc_s values), agreeing with the results reported earlier in Figures 3–5 and the identified challenges for accurately approximating the heteroscedastic variance. Comparing the same implementation for drift and acceleration EDPs, prediction of the latter appears to be more challenging. The results for $n_r = 5$ tend to yield better performance, as this choice allows for the use of a larger number of training points. These trends with respect to the impact of n_r, N_{sim} and N_t, on the stochastic emulation performance and the relative greater challenges for predicting the acceleration EDPs agree with the trends reported in previous studies.[22,32] Interested readers are referred to these studies for more detailed discussions and comparisons for these specific aspects of the stochastic emulation implementation.

The more interesting comparison is, of course, with respect to the different implementations *PR-SE* and *ER-SE*. The trends differ across the different ground motion and EV selection cases. These differences appear to be mostly related to the EV selection, with the implementation corresponding to EV_1, and therefore emulator input corresponding directly only to the features with respect to which the validation is performed, exhibiting different trends compared to the implementations using EV_2 or EV_3. The results for EV_1 selection (GM_1) in Figures 6–8 and 13 show that, as the percentage of training points without replications increases, the accuracy of *ER-SE* continues to increase until the percentage reaches 100%, while that of *PR-SE* might start to decay from 40%−60%, depending on what is the value of N_{sim}. This trend is especially evident for smaller number of simulations (smaller total computational budget), demonstrating that the proposed enhancements can support substantially improved computational efficiency. The improvements are consistent for all EDP types and across all replication sizes.

Figure 13, in particular, clearly demonstrates that the reduction of accuracy for *PR-SE* is related to the challenge in predicting the variance. The prediction accuracy for the median consistently increases as the percentage of non-replication points increases while that of the variance decays. The latter can be explained by the reduction of information on the variance that is fed into the secondary GP for *PR-SE*. However, albeit the decay of variance prediction, exploration offered

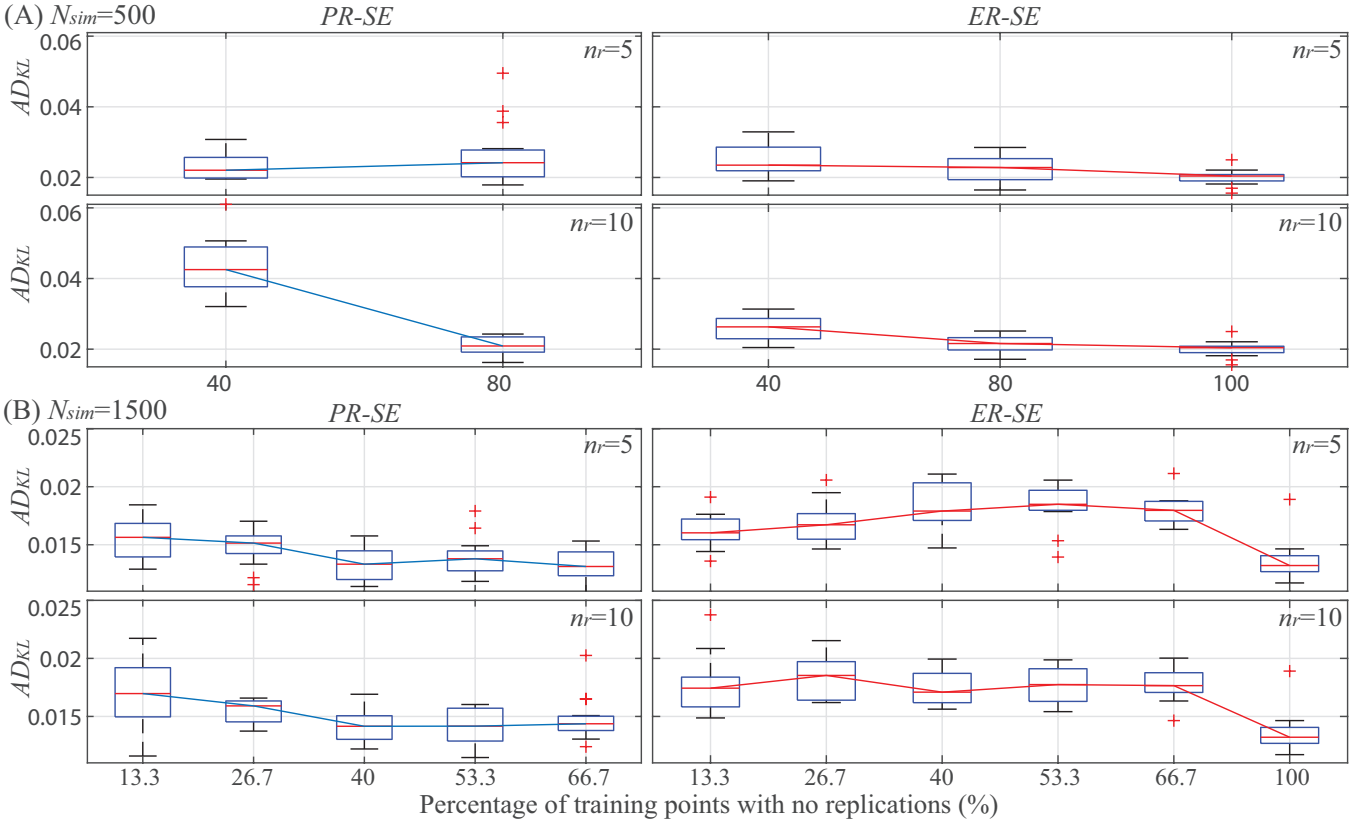

FIGURE 9 Averaged KL divergence across all floors for inter-story drift EDPs for *PR-SE* (left column) and *ER-SE* (right column) implementations for ground motion GM_2 with explanatory variables EV_2. The validation results are shown for different percentage of samples with no replications (abscissas), for N_{sim} equal to (A) 500 and (B) 1500, and for n_r equal to 5 (first row) and 10 (second row). Box plots characteristics are identical to the ones in Figure 6. EDP, engineering demand parameter; KL, Kullback–Leibler.

by non-replication points provides a larger benefit in the median prediction as evidenced by the increasing trend of the accuracy. Still, the case of 80% non-replication shows that a substantial error in the variance estimation can actually lead to poor median prediction. On the other hand, *ER-SE* utilizes fully the information from the training points without replications for training the secondary GP, and this allows *ER-SE* to show a consistent variance prediction accuracy across decreasing values for the percentage of no replication points. The predictions for the median initially follow a similar trend to the *PR-SE* approach, but it continues to increase until it reaches the 100% non-replication case. This is explained by non-degraded performance in the variance prediction. The variance performance is similar when the replication points dominate the training samples, suggesting that addition of non-replication points does not provide any benefit in such cases. However, as the proportion of non-replication points increases, they clearly dominate the variance predictions and restore the level of accuracy observed in the heavily replicated cases. It is important to note that the implementation without any training points with replications (100% for the ratio in the *x*-axis across all figures) shows the best performance, implying that the improved coverage of the training domain achieved by circumventing replications offers meaningful benefits. This agrees with the trends for the SDoF implementation examined in Section 4.2.

The benefits of *ER-SE* become smaller, and consistently appear only for the lower computational budget of $N_{sim} = 500$, for the other two ground motion models and EV selections, that is, GM_2 with EV_2 (Figures 9, 10, and 14) and GM_3 with EV_3 (Figures 11, 12, and 15). For the larger N_{sim} values, cases are observed in which *PR-SE* outperforms the *ER-SE*, though *ER-SE* ends up giving comparable performance to the best instance of *PR-SE* if the total computational budget is assigned to cases with no replications. In particular, Figures 14 and 15 indicate that *PR-SE* and *ER-SE* have comparable performance both for median and variance, and that the performance rank between the two is manifested primarily as sample fluctuations in the variance prediction, rather than meaningful performance trends. The comparisons still indicate a preference towards *ER-SE* over *PR-SE*, though at a smaller degree compared to the previous instance, as only the former can accommodate implementation without any replications for which, according to the dominant trend, we expect the best performance, while circumventing challenges of creating the replications as reported earlier.

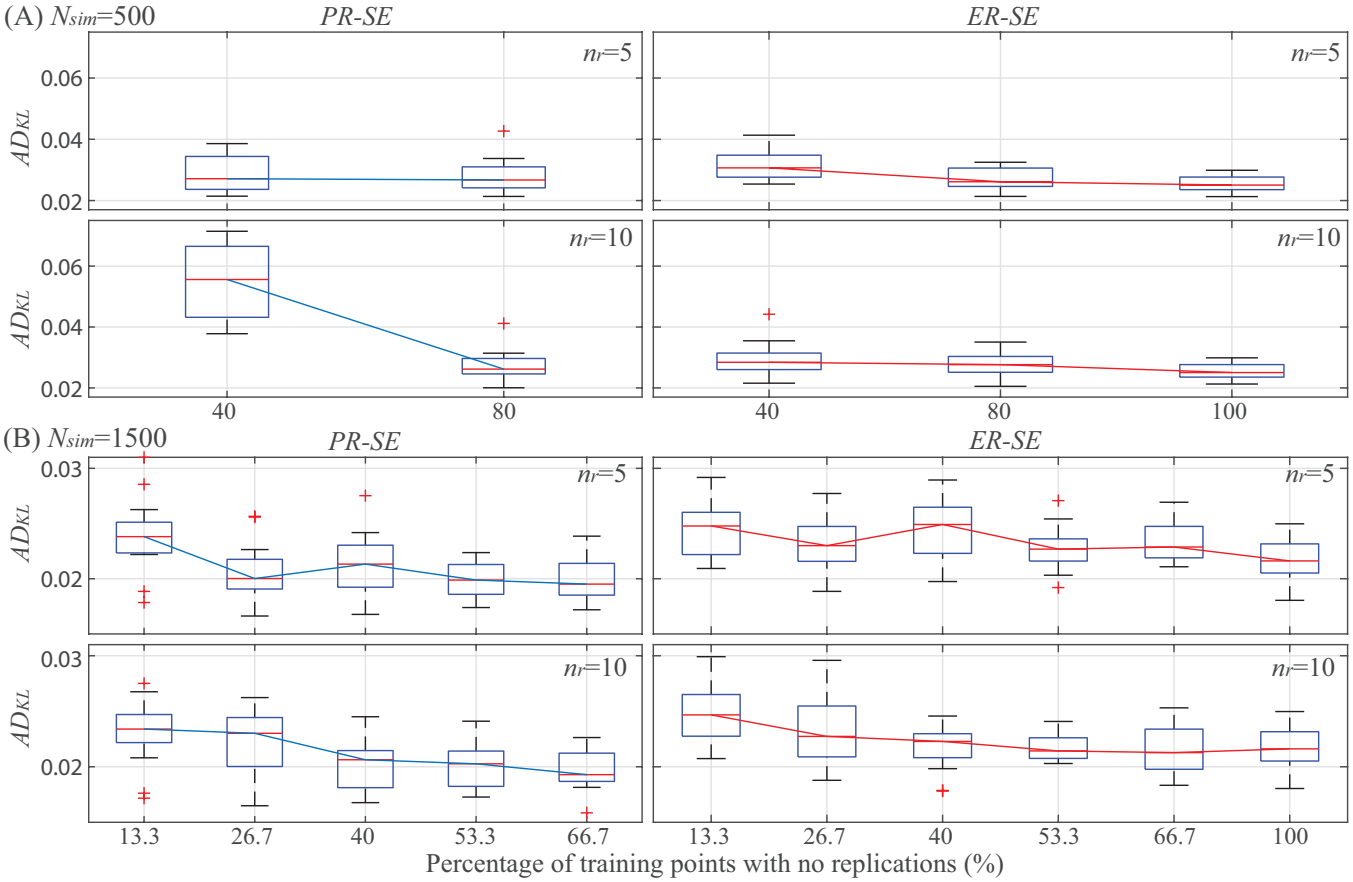

FIGURE 10 Averaged KL divergence across all floors for absolute floor acceleration EDPs for *PR-SE* (left column) and *ER-SE* (right column) implementations for ground motion GM_2 with explanatory variables EV_2. The validation results are shown for different percentage of samples with no replications (abscissas), for N_{sim} equal to (A) 500 and (B) 1500, and for n_r equal to 5 (first row) and 10 (second row). Box plots characteristics are identical to the ones in Figure 6. EDP, engineering demand parameter; KL, Kullback–Leibler.

Looking into more detail on the replication size n_r, for *ER-SE* small only impact is observed, while *PR-SE* appears to prefer smaller replication sizes, $n_r = 5$, especially when training sample size is small, $N_{sim} = 500$. This trend is consistent with the results in refs. [22, 32] and implies that the loss in the domain coverage (fewer training points) for a large replication size has a more negative impact in *PR-SE*, presumably because of the reduction of the training points available for the secondary surrogate model development—something that is avoided in *ER-SE* by leveraging all training points (even the ones without replications) for that development.

5.5 | Implementation and validation for seismic risk estimation

The validated metamodels are now used for the seismic risk assessment of the three-story structure. The setup is identical to the one adopted in refs. [22, 32]. For quantifying the seismic risk, the probability distributions for the structural parameters are the ones presented in Table 1, while for the uncertain hazard parameters the following are adopted: for GM_1 based on ref. [32] the rupture distance, r_{rup}, a lognormal distribution with median 15 km and a coefficient of variation 50% is assumed, whereas for the moment magnitude, M, the Gutenberg–Richter model truncated in interval $[M_{min}, M_{max}] = [5\ 8]$ is utilized, leading to the probability density function $p(M) = b_M e^{-b_M M}/[e^{-b_M M_{\min}} - e^{-b_M M_{\max}}]$. The seismicity parameter for M is taken to be $b_M = 0.9\log_e(10)$. For convenience these probability distributions for the structural and hazard parameters will be denoted as $p(\mathbf{x}_s)$ and $p(\mathbf{x}_m)$, respectively. For GM_2 and GM_3 implementations, and based on ref. [22] the statistics for r_{rup} are adjusted to median 20 km and coefficient of variation 30% while the truncation interval for M to [6 8]. These differences are necessitated by the domain of applicability for each of the three ground motion models.

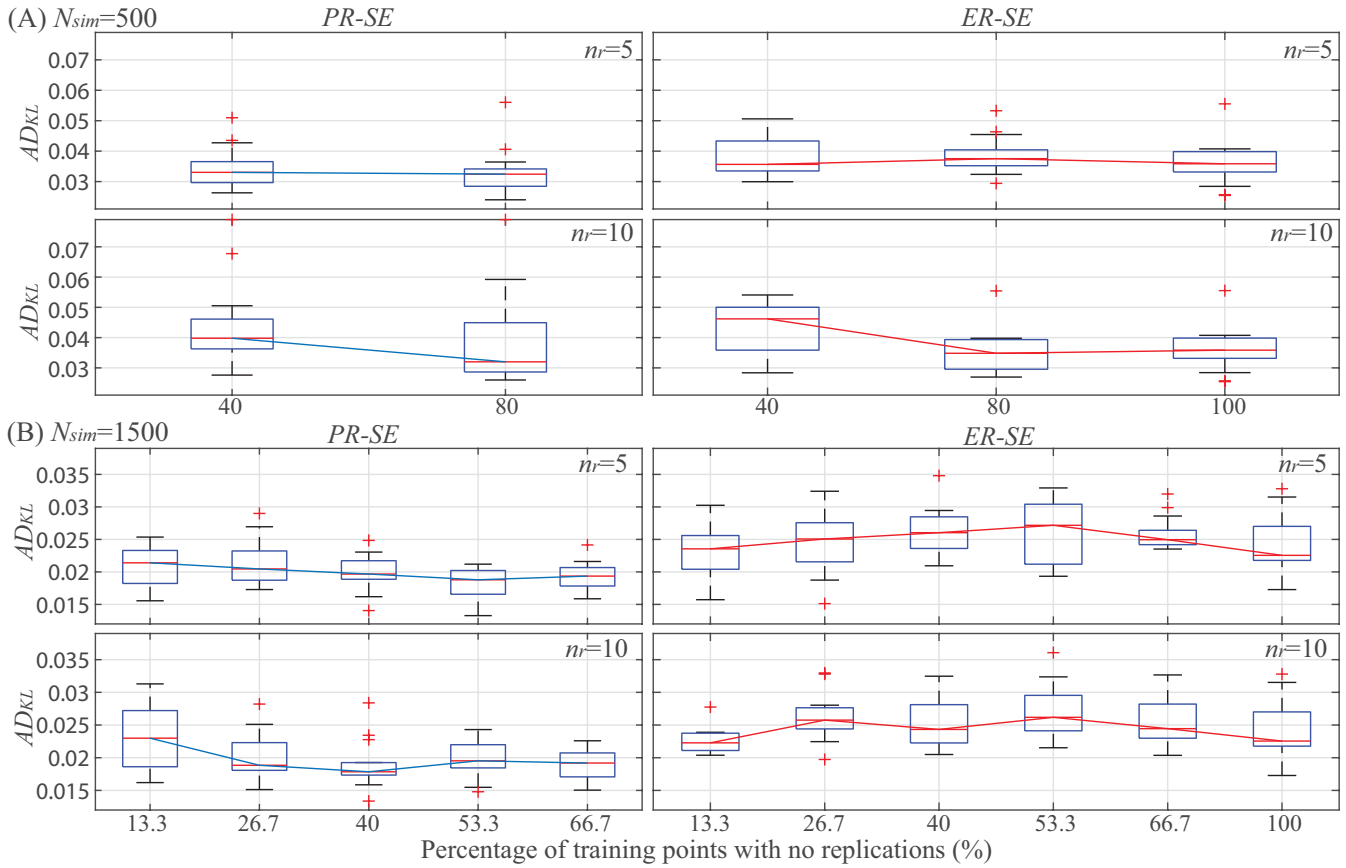

FIGURE 11 Averaged KL divergence across all floors for inter-story drift EDPs for *PR-SE* (left column) and *ER-SE* (right column) implementations for ground motion GM_3 with explanatory variables EV_3. The validation results are shown for different percentage of samples with no replications (abscissas), for N_{sim} equal to (A) 500 and (B) 1500, and for n_r equal to 5 (first row) and 10 (second row). Box plots characteristics are identical to the ones in Figure 6. EDP, engineering demand parameter; KL, Kullback–Leibler.

Risk is quantified as the probability that the EDP will exceed a certain threshold β, corresponding to the complementary cumulative distribution function (CCDF). For estimating this risk, the total probability theorem is utilized to propagate the various sources of uncertainty, while Monte Carlo (MC) simulation is used for the numerical estimation. Utilizing NLRHA for the actual excitation and structural model, the reference seismic risk estimate for the EDP of interest z, and threshold β, is expressed as:

$$P[z > \beta] = \frac{1}{N_h} \sum_{h=1}^{N_h} I[z^h > \beta] \tag{8}$$

where N_h is the total number of samples used for the MC estimation, with z^h representing the hth response sample corresponding to structural configuration $\mathbf{x}_s^h \sim p(\mathbf{x}_s)$, seismic hazard description $\mathbf{x}_m^h \sim p(\mathbf{x}_m)$ and aleatoric hazard description as prescribed by the underlying excitation model. $I[.]$ stands for the indicator function which is one if the quantity within the brackets holds, else it is zero. Utilizing the surrogate model approximation for the EDP distribution, it is straightforward to show that the seismic risk is approximated as follows[22,32]:

$$P[z > \beta] \approx \frac{1}{N_q} \sum_{q=1}^{N_q} \Phi \left[\frac{\tilde{y}(\mathbf{x}^q) - \ln(\beta)}{\sqrt{\sigma^2(\mathbf{x}^q) + \tau^2(\mathbf{x}^q)}} \right] \tag{9}$$

where and N_q is the total number of samples used for the MC estimation, $\Phi[.]$, stands for the standard Gaussian cumulative distribution function, $\mathbf{x}^q$ is the qth sample of the surrogate model input vector, following the appropriate probability distributions based on the definition of this vector (selection of EV): for explanatory variables EV_1 for which $\mathbf{x} = \{\mathbf{x}_s, \mathbf{x}_m\}$, $\mathbf{x}_s^q \sim p(\mathbf{x}_s)$ and $\mathbf{x}_m^q \sim p(\mathbf{x}_m)$; for EV_2 for which $\mathbf{x} = \{\mathbf{x}_s, \mathbf{x}_g\}$, $\mathbf{x}_s^q \sim p(\mathbf{x}_s)$ and $\mathbf{x}_g^q \sim p(\mathbf{x}_g|\mathbf{x}_m)$ with $\mathbf{x}_m^q \sim p(\mathbf{x}_m)$; while for

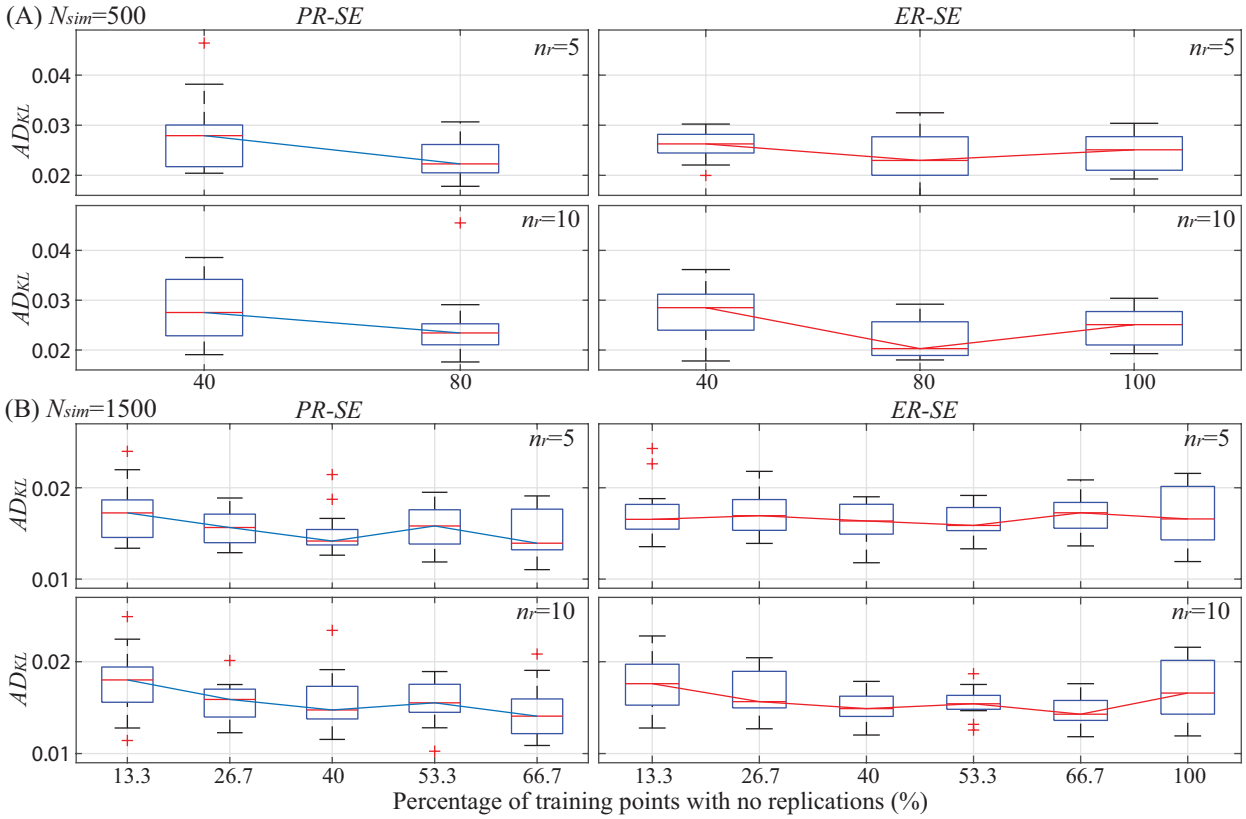

FIGURE 12 Averaged KL divergence across all floors for absolute floor acceleration EDPs for *PR-SE* (left column) and *ER-SE* (right column) implementations for ground motion GM_3 with explanatory variables EV_3. The validation results are shown for different percentage of samples with no replications (abscissas), for N_{sim} equal to (A) 500 and (B) 1500, and for n_r equal to 5 (first row) and 10 (second row). Box plots characteristics are identical to the ones in Figure 6. EDP, engineering demand parameter; KL, Kullback–Leibler.

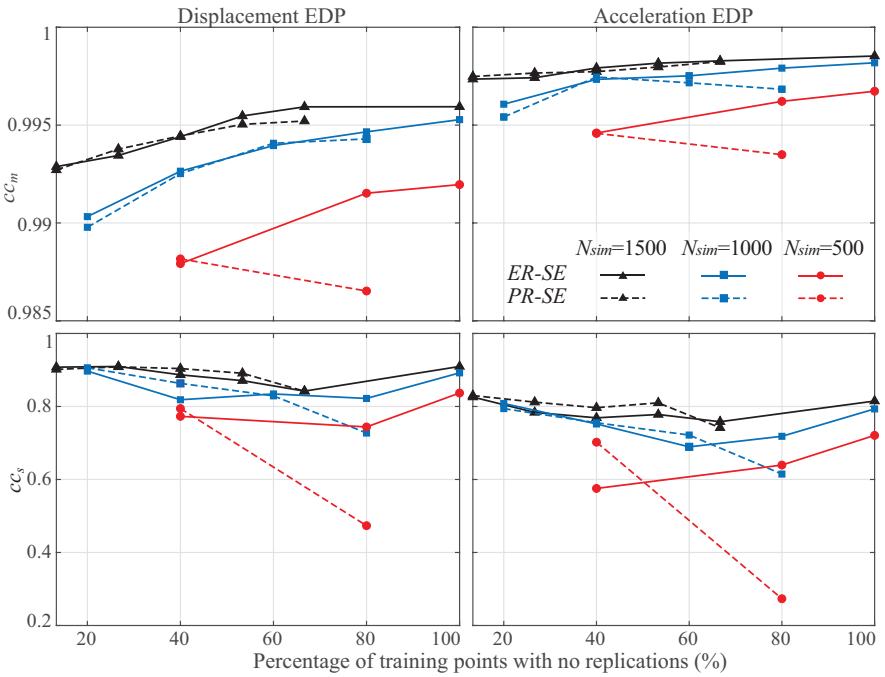

FIGURE 13 Average (across floors) correlation coefficient for different percentages of samples with no replications (abscissas) for *ER-SE* (solid line) and *PR-SE* (dotted line) implementations for ground motion GM with explanatory variables EV_1. Shown are the results for predicted median (top) and logarithmic standard deviation (bottom) of displacement EDPs (left column) and acceleration EDPs (right column) averaged across all floors. n_r is equal to 10. EDP, engineering demand parameter.

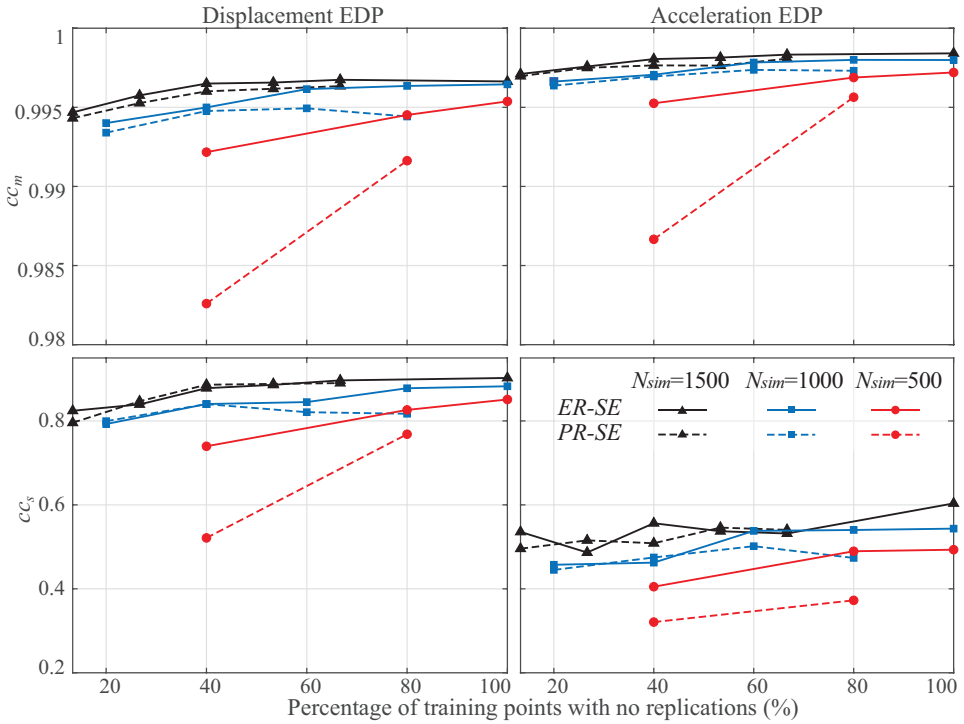

FIGURE 14 Average (across floors) correlation coefficient for different percentages of samples with no replications (abscissas)for *ER-SE* (solid line) and *PR-SE* (dotted line) implementations for ground motion GM_2 with explanatory variables EV_2. Shown are the results for predicted median (top) and logarithmic standard deviation (bottom) of displacement EDPs (left column) and acceleration EDPs (right column) averaged across all floors. n_r is equal to 10. EDP, engineering demand parameter.

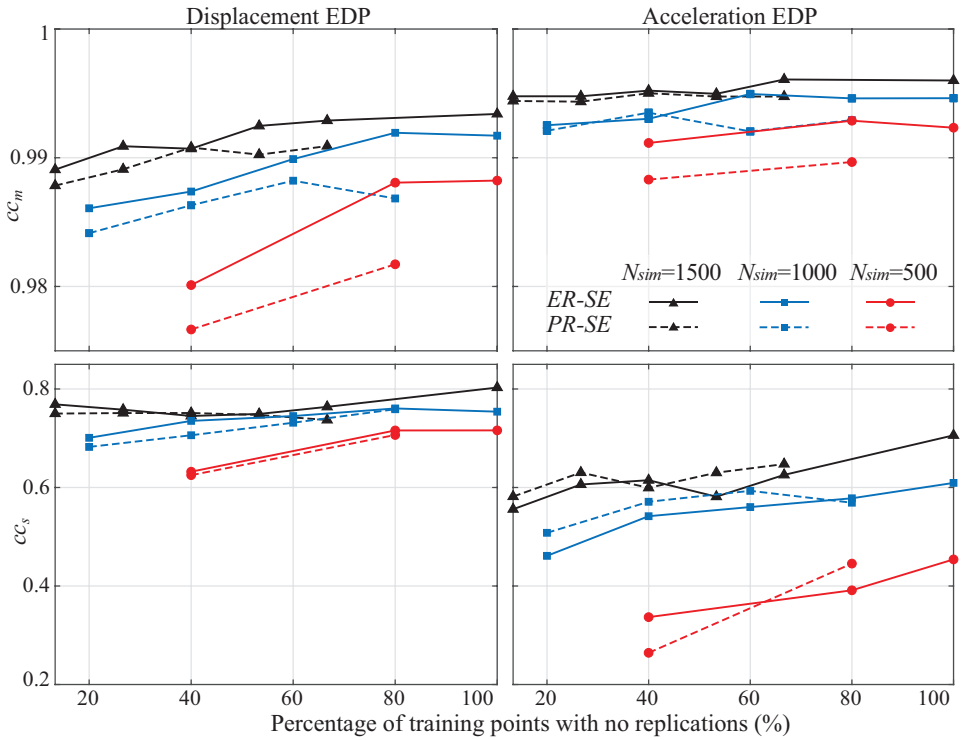

FIGURE 15 Average (across floors) correlation coefficient for different percentages of samples with no replications (abscissas) for *ER-SE* (solid line) and *PR-SE* (dotted line) implementations for ground motion GM_3 with explanatory variables EV_3. Shown are the results for predicted median (top) and logarithmic standard deviation (bottom) of displacement EDPs (left column) and acceleration EDPs (right column) averaged across all floors. n_r is equal to 10. EDP, engineering demand parameter.

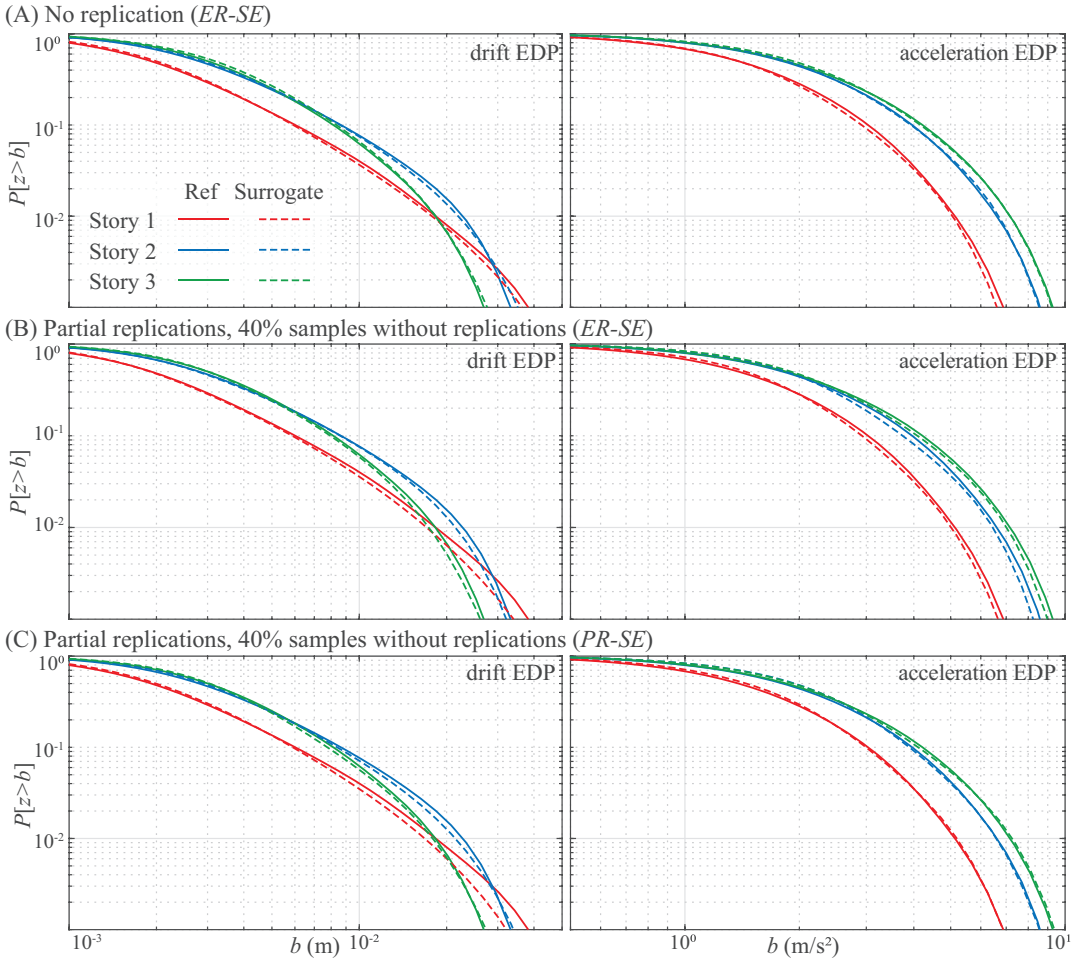

FIGURE 16 Comparison of CCDF curves for the high fidelity model (reference curve) and the surrogate model approximations under the ground motion model GM_1 with explanatory variables EV_1. Shown are the results for selective cases of no replication and partial replication ($N_t = 400$ and $n_r = 10$) implementations with $N_{sim} = 1000$. Results are presented separately for the inter-story displacement EDP (left column) and the absolute floor acceleration EDP (right column). CCDF, complementary cumulative distribution function; EDP, engineering demand parameter.

EV_3 for which $\mathbf{x} = \{\mathbf{x}_s, \mathbf{x}_m, \mathbf{x}_g^1\}$, $\mathbf{x}_s^q \sim p(\mathbf{x}_s)$, $\mathbf{x}_m^q \sim p(\mathbf{x}_m)$ and $\mathbf{x}_g^{1q} \sim p(\mathbf{x}_g^1|\mathbf{x}_m)$. Note that z^h in Equation (8) requires the execution of NLRHAs, while $\tilde{y}(\mathbf{x}^q)$ and $\sigma^2(\mathbf{x}^q) + \tau^2(\mathbf{x}^q)$ in Equation (9) the simple evaluation of the (already calibrated) surrogate model, accommodating substantial computational savings even if $N_q >> N_h$. Similar to refs. [22, 32] to reduce any MC variability in the estimates, both N_h and N_q are assigned a very high value, equal to 5000 samples, and common random numbers are used for the hazard and structural parameters to facilitate the comparison between the curves (establishing a consistent MC error estimation).

The CCDF prediction obtained by Equation (9) is compared with the reference analysis obtained by MCSs using Equation (8), and the results are presented selectively in Figures 16–18. The predictions from both no replication (*ER-SE*) and partial replication (*ER-SE* and *PR-SE*) implementations are presented, separately for each inter-story displacement EDP and the absolute floor acceleration EDP. Figure 16 shows the prediction under ground motion GM_1 and implementation setup is $N_{sim} = 1000$ for no replication case (part A) and additionally $n_r = 10$ and $N_t = 400$ for partial replication cases (parts B and C). Similarly, Figures 17 and 18 respectively present the prediction under ground motion GM_2 and GM_3 for $N_{sim} = 1500$, $n_r = 5$, and $N_t = 600$, where the latter two values apply only to partial replication implementations. Note that results corresponding to higher computational budget are presented for GM_2 and GM_3, since for lower total computational budgets, the accuracy is decreased due to the higher input dimension associated EV_2 and EV_3.

As expected from the high accuracy in the previous validation results, the risk estimates of all the inspected cases show good match with the reference curves. This is an important outcome, indicating that the validation setting can provide

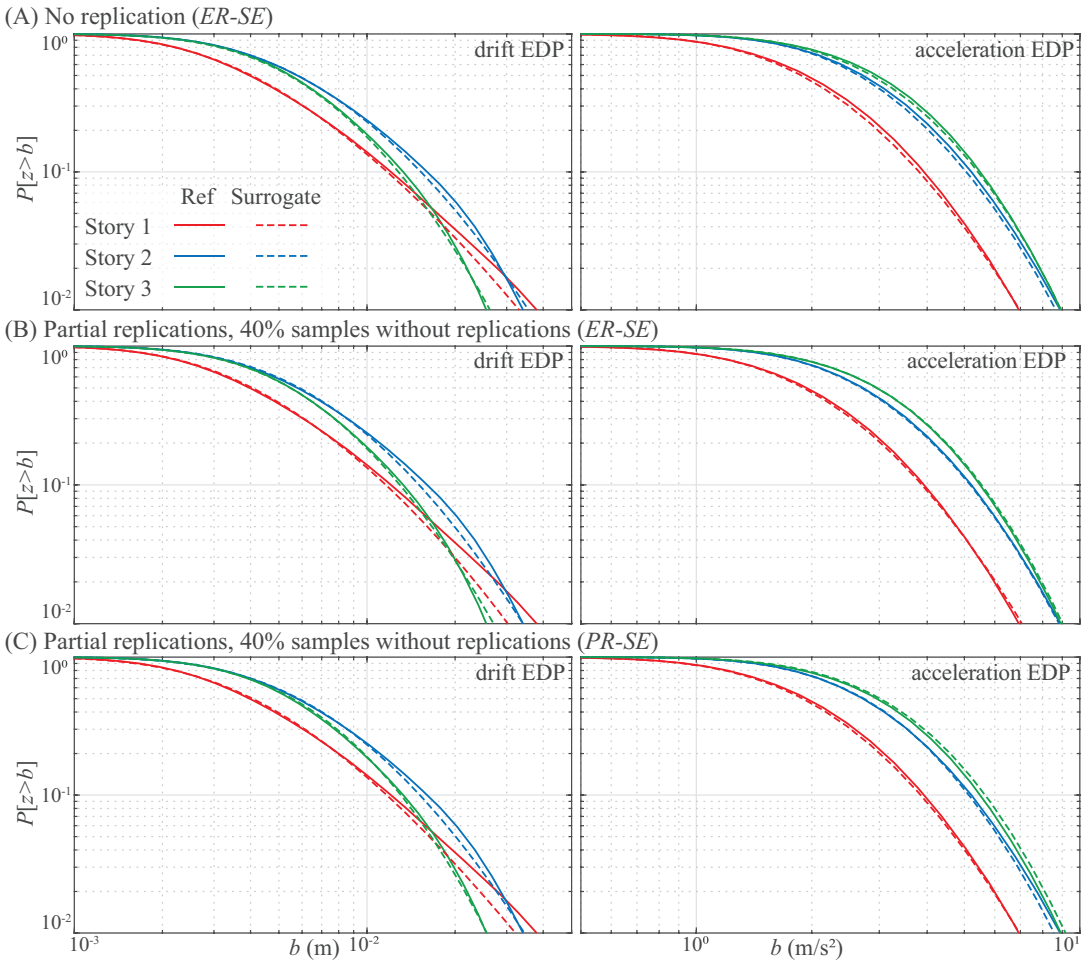

FIGURE 17 Comparison of CCDF curves for the high fidelity model (reference curve) and the surrogate model approximations under the ground motion model GM_2 with explanatory variables EV_2. Shown are the results for selective cases of no replication and partial replication ($N_t = 600$ and $n_r = 5$) implementations with $N_{sim} = 1500$. Results are presented separately for the inter-story displacement EDP (left column) and the absolute floor acceleration EDP (right column). CCDF, complementary cumulative distribution function; EDP, engineering demand parameter.

a faithful indication of the quality of the risk estimates when the trained surrogate model is subsequently used in a risk assessment setting. Minor improvements in the accuracy are observed in the drift EDP by introducing proposed method with no replication strategy. Overall, the risk assessment results agree with the ones reported in previous studies,[22,32] demonstrating the benefits (high accuracy with great computational efficiency) that can be achieved through stochastic emulation strategies within seismic risk and loss estimation.

6 | CONCLUSIONS

This paper offered an advancement of the recently developed partial-replication-based stochastic emulation framework, established by incorporating more holistically information from training points without replications. The original framework leverages a secondary GP to estimate the heteroscedastic variability of the primary GP model, with the latter providing the desired approximation for the EDP response distribution. In this setting, among the partially replicated training data, only the portion with replications is utilized for the inference of the variance-field. The proposed enhancement additionally considers the non-replicated sample portion for this objective, by training the secondary GP based on observations of the deviation of non-replication samples from an approximate (primitive) mean estimate. This primitive mean estimate is obtained leveraging a homoscedastic GP. Compared to the previous partial replication approach, the enhanced approach in this study (i) significantly increases the contribution of no replication samples in estimating the

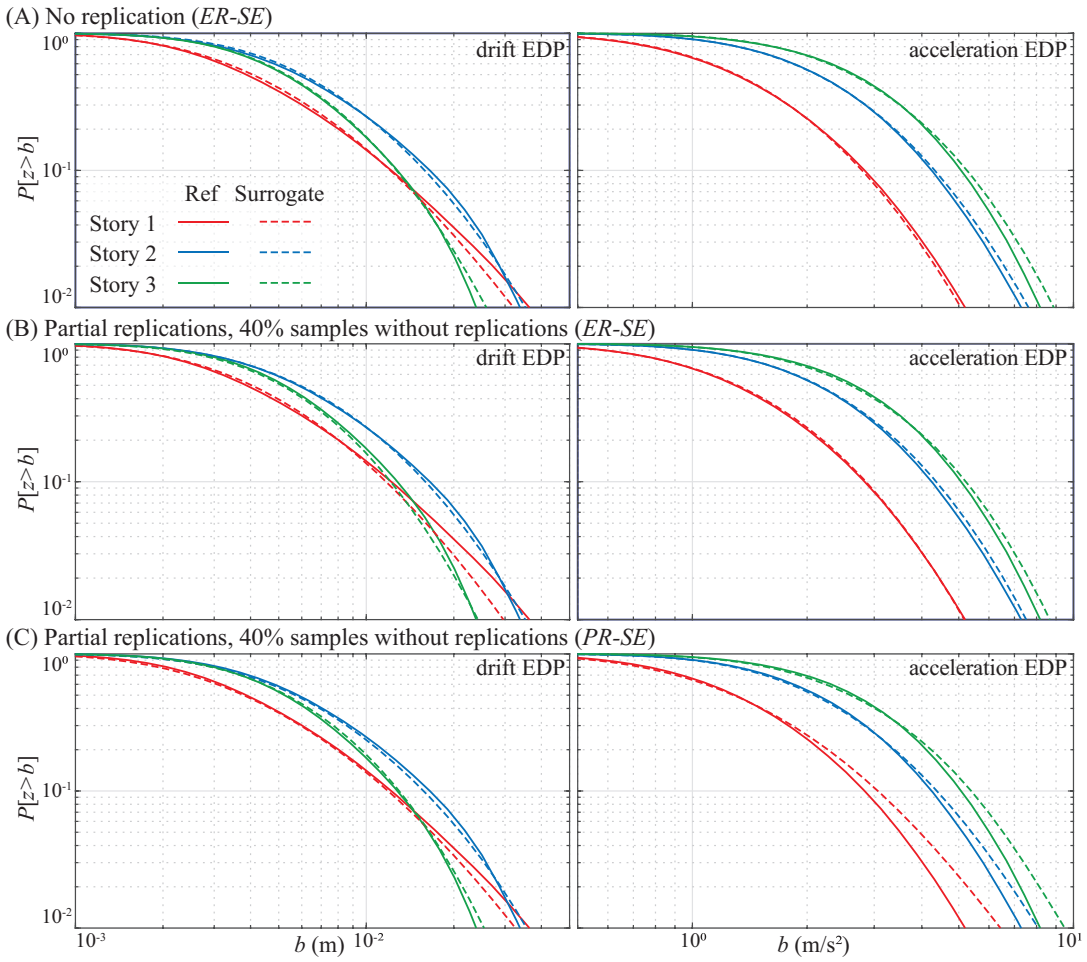

FIGURE 18 Comparison of CCDF curves for the high fidelity model (reference curve) and the surrogate model approximations under the ground motion model GM_3 with explanatory variables EV_3. Shown are the results for selective cases of no replication and partial replication ($N_t = 600$ and $n_r = 5$) implementations with $N_{sim} = 1500$. Results are presented separately for the inter-story displacement EDP (left column) and the absolute floor acceleration EDP (right column). CCDF, complementary cumulative distribution function; EDP, engineering demand parameter.

heteroscedastic variance, while it ultimately (ii) enables an implementation without any replications, accommodating applications for which creating such replications across the entire surrogate model input domain can be challenging. Such merits were demonstrated for earthquake engineering applications.

A comprehensive case study was presented to validate the proposed method, considering three different setups with respect to the aleatoric seismic hazard uncertainty description within NLRHA-based seismic risk assessment. The case study results demonstrated that the complete engagement of non-replication training points can provide a better overall performance. Though differences were reported across the different setups, the enhanced approach shows similar or better performance to the original formulation. Noticeable performance improvement was observed as the proportion of the non-replicated samples increases, especially when the total computational budget (total number of NLRHA simulations) is maintained moderately low. Results indicate that the enrichment of the training domain coverage allowed by the use of non-replicated samples for the secondary surrogate model can offer tangible benefits and that this benefit can be further extended by removing replications altogether. Noticeably, the enhanced approach consistently provided very good performance even for implementation without any replications, demonstrating that it offers an attractive alternative to the existing approach or even a viable solution for applications for which that approach could not be adopted (inability to perform replications). Interesting future research topics include the adaptive selection of simulation experiments for improving prediction accuracy across both the mean and variance stochastic fields and the extension of the stochastic emulation framework to predict the correlation of the EDP responses.

ACKNOWLEDGMENTS

This research was financially supported by the National Science Foundation under Grant CMMI-2131111. This support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

DATA AVAILABILITY STATEMENT

Data and models utilized in this paper are publicly available in DesignSafe data repository (Project PRJ-4192) at https://doi.org/10.17603/ds2-ys6d-bp28.[54]

ORCID

Sang-ri Yi https://orcid.org/0000-0001-7196-127X
Alexandros A. Taflanidis https://orcid.org/0000-0002-9784-7480

REFERENCES

1. Baker JW, Cornell CA. Uncertainty propagation in probabilistic seismic loss estimation. *Struct Saf*. 2008;30(3):236-252.
2. McKenna F. OpenSees: a framework for earthquake engineering simulation. *Comput Sci Eng*. 2011;13(4):58-66.
3. Goulet CA, Haselton CB, Mitrani-Reiser J, et al. Evaluation of the seismic performance of code-conforming reinforced-concrete frame building-From seismic hazard to collapse safety and economic losses. *Earthquake Eng Struct Dyn*. 2007;36(13):1973-1997.
4. McBride K, Sundmacher K. Overview of surrogate modeling in chemical process engineering. *Chem Ing Tech*. 2019;91(3):228-239.
5. Razavi S, Tolson BA, Burn DH. Review of surrogate modeling in water resources. *Water Resour Res*. 2012;48(7):W07401.
6. Alizadeh R, Allen JK, Mistree F. Managing computational complexity using surrogate models: a critical review. *Res Eng Des*. 2020;31:275-298.
7. Song J, Kang W-H, Lee Y-J, Chun J. Structural system reliability: overview of theories and applications to optimization. *ASCE-ASME J Risk Uncertain Eng Syst A: Civ Eng*. 2021;7(2):03121001.
8. Gramacy RB. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. Chapman and Hall/CRC; 2020.
9. Kleijnen JPC. Kriging metamodeling in simulation: a review. *Eur J Oper Res*. 2009;192(3):707-716.
10. Kim J, Song J. Reliability-based design optimization using quantile surrogates by adaptive gaussian process. *J Eng Mech*. 2021;147(5):04021020.
11. Gidaris I, Taflanidis AA, Mavroeidis GP. Kriging metamodeling in seismic risk assessment based on stochastic ground motion models. *Earthq Eng Struct Dyn*. 2015;44(14):2377-2399.
12. Zhang J, Foschi RO. Performance-based design and seismic reliability analysis using designed experiments and neural networks. *Probab Eng Mech*. 2004;19(3):259-267.
13. Buratti N, Ferracuti B, Savoia M. Response surface with random factors for seismic fragility of reinforced concrete frames. *Struct Saf*. 2010;32(1):42-51.
14. Abbiati G, Broccardo M, Abdallah I, Marelli S, Paolacci F. Seismic fragility analysis based on artificial ground motions and surrogate modeling of validated structural simulators. *Earthq Eng Struct Dyn*. 2021;50(9):2314-2333.
15. Zhu X, Broccardo M, Sudret B. Seismic fragility analysis using stochastic polynomial chaos expansions. *Probab Eng Mech*. 2023;72:103413.
16. Gentile R, Galasso C. Gaussian process regression for seismic fragility assessment of building portfolios. *Struct Saf*. 2020;87:101980.
17. Gentile R, Galasso C. Surrogate probabilistic seismic demand modelling of inelastic single-degree-of-freedom systems for efficient earthquake risk applications. *Earthq Eng Struct Dyn*. 2022;51(2):492-511.
18. Kim T, Kwon OS, Song J. Deep learning based seismic response prediction of hysteretic systems having degradation and pinching. *Earthq Eng Struct Dyn*. 2023;52(8):2384-2406.
19. Jalayer F, Beck J, Zareian F. Analyzing the sufficiency of alternative scalar and vector intensity measures of ground shaking based on information theory. *J Eng Mech*. 2012;138(3):307-316.
20. McGuire RK. *Seismic Hazard and Risk Analysis*. Earthquake Engineering Research Institute; 2004.
21. Jalayer F, Beck JL. Effects of two alternative representations of ground-motion uncertainty in probabilistic seismic demand assessment of structures. *Earthquake Eng Struct Dyn*. 2008;37(1):61-79.
22. Kyprioti AP, Taflanidis AA. Addressing the different sources of excitation variability in seismic response distribution estimation using kriging metamodeling. *Earthquake Eng Struct Dyn*. 2022;51(10):2466-2495.
23. Ankenman B, Nelson BL, Staum J. *Stochastic Kriging for Simulation Metamodeling*. IEEE; 2008.
24. Baker E, Barbillon P, Fadikar A, et al. Analyzing stochastic computer models: a review with opportunities. *Stat Sci*. 2022;37(1):64-89.
25. Binois M, Huang J, Gramacy RB, Ludkovski M. Replication or exploration? Sequential design for stochastic simulation experiments. *Technometrics*. 2019;61(1):7-23.
26. Binois M, Gramacy RB, Ludkovski M. Practical heteroscedastic Gaussian process modeling for large simulation experiments. *J Comput Graph Statist*. 2018;27(4):808-821.
27. Yin J, Ng SH, Ng KM. Kriging metamodel with modified nugget-effect: the heteroscedastic variance case. *Comput Ind Eng*. 2011;61(3):760-777.

28. Zhu X, Sudret B. Replication-based emulation of the response distribution of stochastic simulators using generalized lambda distributions. *Int J Uncertain Quantif*. 2020;10(3):249-275.
29. Zhu X, Sudret B. Emulation of stochastic simulators using generalized lambda models. *SIAM/ASA J Uncertain Quantif*. 2021;9(4):1345-1380.
30. Marrel A, Iooss B, Veiga SDa, Ribatet M. Global sensitivity analysis of stochastic computer models with joint metamodels. *Stat Comput*. 2012;22:833-847.
31. Zhu X, Sudret B. Stochastic polynomial chaos expansions to emulate stochastic simulators. *Int J Uncertain Quantif*. 2023;13(2):31-52.
32. Kyprioti AP, Taflanidis AA. Kriging metamodeling for seismic response distribution estimation. *Earthq Eng Struct Dyn*. 2021;50(13):3550-3576.
33. Muñoz-González L, Lázaro-Gredilla M, Figueiras-Vidal AR. *Heteroscedastic Gaussian process regression using expectation propagation. 2011 IEEE International Workshop on Machine Learning for Signal Processing*. IEEE; 2011.
34. Goldberg P, Williams C, Bishop C. Regression with input-dependent noise: a Gaussian process treatment. *Adv Neural Inf Process*. 1997;10:493-499.
35. Nelder JA, Pregibon D. An extended quasi-likelihood function. *Biometrika*. 1987;74(2):221-232.
36. Zhang Q-H, Ni Y-Q. Improved most likely heteroscedastic Gaussian process regression via Bayesian residual moment estimator. *IEEE Trans Signal Process*. 2020;68:3450-3460.
37. Kazantzi A, Vamvatsikos D, Lignos D. Seismic performance of a steel moment-resisting frame subject to strength and ductility uncertainty. *Eng Struct*. 2014;78:69-77.
38. Taflanidis AA, Beck JL. Life-cycle cost optimal design of passive dissipative devices. *Struct Saf*. 2009;31(6):508-522.
39. Gentile R, Galasso C, Pampanin S. Material property uncertainties versus joint structural detailing: relative effect on the seismic fragility of reinforced concrete frames. *J Struct Eng*. 2021;147(4):04021007.
40. Gidaris I, Taflanidis AA, Mavroeidis GP. Surrogate modeling implementation for assessment of seismic risk utilizing stochastic ground motion modeling, In *Proc., 2nd European Conference in Earthquake Engineering and Seismology*. Istanbul, Turkey, 2014.
41. Gidaris I, Taflanidis AA, Mavroeidis GP. Multiobjective design of supplemental seismic protective devices utilizing lifecycle performance criteria. *J Struct Eng*. 2018;144(3):04017225.
42. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press; 2005.
43. Gramacy RB, Lee HK. Cases for the nugget in modeling computer experiments. *Stat Comput*. 2012;22(3):713-722.
44. Cleveland WS, Loader C. Smoothing by local regression: principles and methods. Statistical Theory and Computational Aspects of Smoothing: Proceedings of the COMPSTAT'94 Satellite Meeting held in Semmering, Austria, 27–28 August 1994. Springer; 1996.
45. Li K, Wang S, Liu Y, Song X. An integrated surrogate modeling method for fusing noisy and noise-free data. *J Mech Des*. 2022;144(6):061701.
46. Hastie T, Tibshirani R, Friedman JH, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Vol 2. Springer; 2009.
47. Boore DM. Simulation of ground motion using the stochastic method. *Pure Appl Geophys*. 2003;160:635-676.
48. Patsialis D, Taflanidis A. Reduced order modeling of hysteretic structural response and applications to seismic risk assessment. *Eng Struct*. 2020;209:110135.
49. Rezaeian S, Der Kiureghian A. Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq Eng Struct Dyn*. 2010;39(10):1155-1180.
50. Vlachos C, Papakonstantinou KG, Deodatis G. Predictive model for site specific simulation of ground motions based on earthquake scenarios. *Earthq Eng Struct Dyn*. 2018;47(1):195-218.
51. Haselton CB, Goulet CA, Mitrani-Reiser J, et al. *An Assessment to Benchmark the Seismic Performance of a Code-Conforming Reinforced Concrete Moment-Frame Building*. Pacific Earthquake Engineering Research Center; 2008. PEER 2007/12.
52. Angeles K, Patsialis D, Taflanidis AA, Kijewski-Correa TL, Buccellato A, Vardeman C. Advancing the design of resilient and sustainable buildings: an integrated life-cycle analysis. *J Struct Eng*. 2021;147(3):04020341.
53. Burbea J, Rao CR. Entropy differential metric, distance and divergence measures in probability spaces: a unified approach. *J Multivariate Anal*. 1982;12(4):575-596.
54. Kyprioti AP, Yi S-r, Taflanidis AA. *Peak EDP Distribution from Nonlinear Time-History Analysis for Three Story Concrete Moment Resisting Frame Exposed to Different Stochastic Ground Motion Models*. DesignSafe-CI; 2023. doi:10.17603/ds2-ys6d-bp28v1
55. Bostanabad R, Kearney T, Tao S, Apley DW, Chen W. Leveraging the nugget parameter for efficient Gaussian process modeling. *Int J Numer Methods Eng*. 2018;114(5):501-516.
56. Audet C, Dennis JE, Jr. Analysis of generalized pattern searches. *SIAM J Optim*. 2002;13(3):889-903.

How to cite this article: Yi S, Taflanidis AA. Stochastic emulation with enhanced partial- and no-replication strategies for seismic response distribution estimation. *Earthquake Engng Struct Dyn*. 2024;53:2354–2381. https://doi.org/10.1002/eqe.4115

APPENDIX A: REVIEW OF GP FORMULATION WITH POSSIBLE REPLICATIONS AND HETEROSCEDASTIC NUGGET

This appendix reviews the GP formulation with possible replications and heteroscedastic nugget.[32] It is presented for a generic output v and for input $\mathbf{x}$. For appropriate selections of number of replications, nugget characteristics and output definition, this formulation covers all the GP-variants discussed in this paper.

To formalize concepts, consider N_s training points $\{\mathbf{x}^i; i = 1, \dots, N_s\}$ and corresponding outputs with replications $\{v^{i(k)}; i = 1, \dots, N_s,\ k = 1, \dots .n^i\}$ where n^i denotes the number of replications at sample $\mathbf{x}^i$. The replication outputs correspond to the same input $\mathbf{x}^i$ but different samples for the characteristics describing the aleatoric randomness in the stochastic simulations. Note that the formulation allows for $n^i = 1$ for some or all the training points, corresponding to no replications. We chose a GP with: (a) trend expressed as a linear regression using basis vector $\mathbf{f}(\mathbf{x})$, and (b) correlation function between inputs $R(\mathbf{x}^l, \mathbf{x}^m|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ denotes its hyper-parameters. We also assume a nugget with variance parameterized as $c^2(\mathbf{x}) = c_s^2\tilde{c}(\mathbf{x})^2$ where c_s^2 represents a scaling variable that needs to be calibrated based on observations and $\tilde{c}(\mathbf{x})^2$ represents the heteroscedastic variability that is chosen a priori. The choice for $\tilde{c}(\mathbf{x})^2$ is discussed in detail within the multi-level stochastic emulation formulations discussed in the paper.

Define now through the observation data the input matrix $\mathbf{X} = [\mathbf{x}^1 \ \dots \ \mathbf{x}^{N_s}]^T$ and the mean output vector $\bar{\mathbf{V}} = [\bar{v}^1 \ \dots \ \bar{v}^{N_s}]^T$ in which

$$\bar{v}^i = \frac{1}{n^i}\sum_{k=1}^{n^i} v^{i(k)}, \quad \text{for } i = 1, \dots, N_t \tag{A1}$$

Also, let $\mathbf{F} = [\mathbf{f}(\mathbf{x}^1)\cdots\mathbf{f}(\mathbf{x}^{N_t})]^{\mathrm{T}}$ denote the basis matrix, $\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}^1|\boldsymbol{\theta}) \dots R(\mathbf{x}, \mathbf{x}^{N_s}|\boldsymbol{\theta})]^{\mathrm{T}}$ the N-dimensional correlation vector between $\mathbf{x}$ and each of the elements of $\mathbf{X}$, $\mathbf{R}$ the $N_s \times N_s$ correlation matrix over database $\mathbf{X}$ with the lm-element corresponding to $R(\mathbf{x}^l, \mathbf{x}^m|\mathbf{v})$, $l,m = 1, \dots, N_s$, $\mathbf{A}$ the $N_s \times N_s$ diagonal matrix with i-th element being the inverse of replication size, that is, $1/n_i$, and $\tilde{\mathbf{C}}$ the $N_s \times N_s$ diagonal matrix with $\tilde{c}(\mathbf{x}^i)^2$ at the i-th element. The stochastic emulation establishes approximation $\hat{p}(v|\mathbf{x}) = N(\bar{v}(\mathbf{x}), \sigma^2(\mathbf{x}) + \tau^2(\mathbf{x}))$, and based on maximum likelihood estimation (MLE) the GP predictive mean and variance, and the nugget variance are, respectively[32]:

$$\bar{v}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{r}(\mathbf{x})^{\mathrm{T}}(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})^{-1}(\bar{\mathbf{V}} - \mathbf{F}\beta) \tag{A2}$$

$$\sigma^2(\mathbf{x}) = \tilde{\sigma}^2\left[1 + \boldsymbol{\gamma}(\mathbf{x})^{\mathrm{T}}\{\mathbf{F}^{\mathrm{T}}(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})^{-1}\mathbf{F}\}^{-1}\boldsymbol{\gamma}(\mathbf{x}) - \mathbf{r}(\mathbf{x})^{\mathrm{T}}(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})^{-1}\mathbf{r}(\mathbf{x})\right] \tag{A3}$$

$$c^2(\mathbf{x}) = \left(\delta \cdot \tilde{\sigma}^2\right)\tilde{c}^2(\mathbf{x}) \tag{A4}$$

where $\boldsymbol{\beta} = (\mathbf{F}^{\mathrm{T}}(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})^{-1}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})^{-1}\bar{\mathbf{V}}$ corresponds to the MLE for the trend coefficients, $\boldsymbol{\gamma}(\mathbf{x})$ is an auxiliary variable defined as $\boldsymbol{\gamma}(\mathbf{x}) = \mathbf{F}^{\mathrm{T}}(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})^{-1}\mathbf{r}(\mathbf{x}) - \mathbf{f}(\mathbf{x})$, $\tilde{\sigma}^2$ represents the process variance with MLE estimate given by

$$\tilde{\sigma}^2 = (\bar{\mathbf{V}} - \mathbf{F}\beta)^{\mathrm{T}}(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})^{-1}(\bar{\mathbf{V}} - \mathbf{F}\beta)/N_s \tag{A5}$$

and δ represents a hyper-parameter introduced so that $c_s^2 = \delta\tilde{\sigma}^2$ (i.e., it corresponds to scaling factor between c_s^2 and $\tilde{\sigma}^2$). The MLE calibration of the hyper-parameters, corresponding to $\boldsymbol{\theta}$ and δ leads to optimal solution[32]:

$$[\boldsymbol{\theta}\delta]^* = \arg\min\{\ln(\det(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})) + N_s\ln((\bar{\mathbf{V}} - \mathbf{F}\beta)^{\mathrm{T}}(\mathbf{R} + \delta\mathbf{A}^{-1}\tilde{\mathbf{C}})^{-1}(\bar{\mathbf{V}} - \mathbf{F}\beta))\} \tag{A6}$$

where det(.) stands for the determinant of a matrix. The optimization of Equation (A6) is well known to have multiple local minima and non-smooth characteristics for small values of δ.[55] These challenges, are addressed by performing all numerical optimizations in this study using a pattern-search algorithm.[56] Should be pointed out that the introduction of δ as hyper-parameter simplifies the MLE calibration process,[32] yielding the closed-form solution of the process variance in Equation (A5), removing the need to consider an explicit numerical optimization for it.[27] Additional details for the derivations in Equations (A2)–(A6) can be found in ref. [32].

The general formulation presented above accommodates a heteroscedastic variance and replications for some portion of the training points, and will be referred to as He-R. By utilizing specific assumptions for n^i, δ and $\tilde{c}(\mathbf{x})^2$, the above derivation covers other simplified GP implementations. Utilizing $\tilde{c}(\mathbf{x})^2 = 1$ leads to a homoscedastic nugget (constant

TABLE A1 GP formulations of interest.

Abbreviation	Characteristics	Variance assumption	Replication assumption	Matrices utilized in implementation
He-R	heteroscedastic with replications	$c^2(\mathbf{x}) = c_s^2\tilde{c}(\mathbf{x})^2$	n^i	$\delta\mathbf{A}^{-1}\tilde{\mathbf{C}} = \delta \cdot \text{diag}\{\tilde{c}(\mathbf{x})^2 \cdot n^i\}^+, \bar{\mathbf{V}}$
He-NR	heteroscedastic without replications	$c^2(\mathbf{x}) = c_s^2\tilde{c}(\mathbf{x})^2$	$n^i = 1$	$\delta\mathbf{A}^{-1}\tilde{\mathbf{C}} = \delta \cdot \text{diag}\{\tilde{c}(\mathbf{x})^2\}^+, \bar{\mathbf{V}} = \mathbf{V}$
Ho-R	homoscedastic with replications	$c^2(\mathbf{x}) = c_s^2$	n^i	$\delta\mathbf{A}^{-1}\tilde{\mathbf{C}} = \delta \cdot \text{diag}\{n^i\}^+, \bar{\mathbf{V}}$
Ho-NR	homoscedastic without replications	$c^2(\mathbf{x}) = c_s^2$	$n^i = 1$	$\delta\mathbf{A}^{-1}\tilde{\mathbf{C}} = \delta, \bar{\mathbf{V}} = \mathbf{V}$
No-N	No-Nugget	$c^2(\mathbf{x}) = 0$	$n^i = 1$	$\delta\mathbf{A}^{-1}\tilde{\mathbf{C}} = 0, \bar{\mathbf{V}} = \mathbf{V}$

Note: $^+\text{diag}\{\mathbf{a}\}$ corresponds to a diagonal matrix with the elements of vector $\mathbf{a}$ across its diagonal.

nugget variance). In this case, $\tilde{\mathbf{C}}$ is equal to the identity matrix, $\tilde{\mathbf{C}} = \mathbf{I}$. Additionally, if $\delta = 0$ is adopted, an implementation without any nugget is considered, which would be appropriate only if observations are treated as entirely deterministic (no aleatoric variability). With respect to the replications, setting $n^i = 1$ for all training points yields the no-replication implementation. In this case, $\mathbf{A} = \mathbf{I}$ and $\bar{\mathbf{V}}$ corresponds directly to the available observations for each training points, $\bar{\mathbf{V}} = \mathbf{V}$, as no averaging is performed. Combinations of these assumptions lead to different limiting implementations. Of interest in this study are the following: (a) heteroscedastic variance with no replications, denoted as He-NR; (b) homoscedastic variance with replications, denoted as Ho-R; and (c) homoscedastic variance with no replications Ho-NR. A summary of the different formulations of interest is presented in Table A1, including the implementation without any nugget.

APPENDIX B: REVIEW OF ESTIMATION OF CONDITIONAL STATISTICS USED IN THE STOCHASTIC EMULATION VALIDATION

This Appendix reviews the estimation of the relevant EDP distribution statistics needed for the validation examined in Section 5.3, corresponding to the probability density function, logarithmic mean (needed to obtain the median) and logarithmic standard deviation, all expressed as functions of $\mathbf{x}_\upsilon = \{\mathbf{x}_s, \mathbf{x}_m\}$. For the EVs options, using EV_1 as the input of surrogate model provides directly the estimates needed as $\mathbf{x} = \mathbf{x}_\upsilon$. For EV_2, the expression of the total probability theorem of Equation (4) can be approximated utilizing MCS as:

$$\tilde{p}(\ln(z)|\mathbf{x}_\upsilon) = \tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_m) = \int \tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_g)p(\mathbf{x}_g|\mathbf{x}_m)d\mathbf{x}_g \approx \frac{1}{N_e}\sum_{o=1}^{N_e} N(\tilde{y}(\mathbf{x}^o), \sigma^2(\mathbf{x}^o) + \tau^2(\mathbf{x}^o)) \tag{A7}$$

with samples $\mathbf{x}^o = \{\mathbf{x}_s, \mathbf{x}_g^o; o = 1, \dots, N_e\}$ where $\mathbf{x}_g^o \sim p(\mathbf{x}_g|\mathbf{x}_m)$. The value of N_e is chosen to be equal to 1000 across all estimates in this paper, to establish a small MC variability. Similarly, the conditional logarithmic mean and variance can be estimated based on the total expectation and variance theorems, and can be calculated using MC and samples $\mathbf{x}^o = \{\mathbf{x}_s, \mathbf{x}_g^o; o = 1, \dots, N_e\}$ as:

$$\begin{aligned} \text{E}[\ln(z)|\mathbf{x}_\upsilon] &= \text{E}[\ln(z)|\mathbf{x}_s, \mathbf{x}_m] = \text{E}[\text{E}[\ln(z)|\mathbf{x}_s, \mathbf{x}_g]|\mathbf{x}_m] \approx \frac{1}{N_e}\sum_{o=1}^{N_e}\tilde{y}(\mathbf{x}^o) \\ s(\mathbf{x}_\upsilon) &= \text{Var}[\ln(z)|\mathbf{x}_s, \mathbf{x}_m] = \text{E}[\text{Var}[\ln(z)|\mathbf{x}_s, \mathbf{x}_g]|\mathbf{x}_m] + \text{Var}[\text{E}[\ln(z)|\mathbf{x}_s, \mathbf{x}_g]|\mathbf{x}_m] \\ &\approx \frac{1}{N_e}\sum_{o=1}^{N_e}\left[\sigma^2(\mathbf{x}^o) + \tau^2(\mathbf{x}^o)\right] + \frac{1}{N_e}\sum_{o=1}^{N_e}(\tilde{y}(\mathbf{x}^o) - \text{E}[\ln(z)|\mathbf{x}_s, \mathbf{x}_m])^2 \end{aligned} \tag{A8}$$

For EV_3, the expression of the total probability theorem of Equation (5) can be approximated utilizing MCS as

$$\tilde{p}(\ln(z)|\mathbf{x}_\upsilon) = \tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_m) = \int \tilde{p}(\ln(z)|\mathbf{x}_s, \mathbf{x}_m, \mathbf{x}_g^1)p(\mathbf{x}_g^1|\mathbf{x}_m)d\mathbf{x}_g^1 \approx \frac{1}{N_e}\sum_{o=1}^{N_e} N(\tilde{y}(\mathbf{x}^o), \sigma^2(\mathbf{x}^o) + \tau^2(\mathbf{x}^o)) \tag{A9}$$

with samples defined in this case as $\mathbf{x}^o = \{\mathbf{x}_s, \mathbf{x}_m, \mathbf{x}_g^{1o}; o = 1, \dots, N_e\}$ where $\mathbf{x}_g^{1o} \sim p(\mathbf{x}_g^{1o}|\mathbf{x}_m)$. Note that the MCS expression is the same as for Equation (A7) with only difference the definition of the surrogate model input $\mathbf{x}$ and the corresponding density utilized to obtain samples $\mathbf{x}^o$. Similarly, the total expectation and variance theorems, yields identical MCS estimates for the logarithmic mean and variance as in Equation (9) for the definition of samples $\mathbf{x}^o$ the same as utilized in Equation (A9).