

A Multi-fidelity Stochastic Simulation Scheme for Estimation of Small Failure Probabilities

Min Li^a, Srinivasan Arunachalam^a, Seymour M.J. Spence^{a,1,*}

^a*Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI 48109, USA*

Abstract

Computing small failure probabilities is often of interest in the reliability analysis of engineering systems. However, this task can be computationally demanding since many evaluations of expensive high-fidelity models are often required. To address this, a multi-fidelity approach is proposed in this work within the setting of stratified sampling. The overall idea is to reduce the required number of high-fidelity model runs by integrating the information provided by different levels of model fidelity while maintaining accuracy in estimating the failure probabilities. More specifically, strata-wise multi-fidelity models are established based on Gaussian process models to efficiently predict the high-fidelity response and the system collapse from the low-fidelity response. Due to the reduced computational cost of the low-fidelity models, the multi-fidelity approach can achieve a significant speedup in estimating small failure probabilities associated with high-fidelity models. The effectiveness and efficiency of the proposed multi-fidelity stochastic simulation scheme are validated through an application to a two-story two-bay steel building under extreme winds.

Keywords: Failure probability, Stochastic simulation, Multi-fidelity modeling, Bayesian nonlinear regression, Wind engineering

1. Introduction

Structural reliability analysis aims to compute the failure probability of a given engineering system (e.g., building structure) considering randomness in the model inputs. For general systems, the failure probability can be estimated through Monte Carlo simulation (MCS). However, to achieve a specified accuracy in estimating a failure probability in the range of 10^{-N} , the required number of samples is approximately 10^{N+2} [1]. As a result, computing small failure probabilities can fast become computationally intensive as it often involves

*Corresponding author

Email addresses: nimli@umich.edu (Min Li), sriarun@umich.edu (Srinivasan Arunachalam), smjs@umich.edu (Seymour M.J. Spence)

¹Tel. +1-734-764-8419, Fax +1-734-764-4292

28 many evaluations of the system which is typically simulated by complex numerical models.
 29 For example, to fully capture the nonlinear behavior of building structures under extreme
 30 winds, high-fidelity finite element models are generally required for capturing aspects such
 31 as material and geometric nonlinearity, including global buckling and material fatigue [2].
 32 To address this, several variance reduction schemes that include, but are not limited to, im-
 33 portance sampling [3], line sampling [4], subset simulation [5], and simulation schemes based
 34 on stratified designs [6, 7, 8] have been proposed in the literature. These techniques aim to
 35 reduce the required number of model evaluations so that the stochastic simulation becomes
 36 efficient. However, when dealing with high-dimensionality (e.g., in the order of thousands)
 37 and the need for simultaneous estimation of multiple small failure probabilities, methods
 38 such as importance sampling and subset simulation can become ineffective. For instance, it
 39 is difficult to identify an effective high-dimensional proposal density for importance sampling,
 40 moreover, when the choice of the form of proposal sampling density adopted is inappropri-
 41 ate, the variability of the estimator cannot be controlled in the presence of a large number
 42 of uncertain parameters [9]. Subset simulation, on the other hand, can quickly turn ineffi-
 43 cient when multiple limit states are considered since each limit state of interest requires an
 44 independent implementation of the subset simulation procedure. In contrast, stratified ran-
 45 dom sampling enables simultaneous estimation of failure probabilities, and has been shown
 46 to be not only efficient but also unbiased, consistent (in the sense that it approaches the
 47 true failure probability for large sample sizes), and asymptotically Gaussian [8]. Another
 48 approach is to increase the computational efficiency of the system model by establishing
 49 low-fidelity versions of the numerical model to approximate the system response/output.
 50 Such low-fidelity models can be created by simplifying the underlying physics, e.g., using
 51 simplified material constitutive models, or establishing reduced-order models or data-driven
 52 surrogate models [10]. Although less computationally demanding, such low-fidelity models
 53 might be inaccurate and may yield biased or distorted estimators if used directly for the
 54 uncertainty propagation [11, 12]. To overcome this difficulty, one solution is to develop a
 55 multi-fidelity model which combines, instead of replaces, the high-fidelity model with the
 56 low-fidelity model. The fundamental idea is to leverage the accuracy of high-fidelity models
 57 and the efficiency of low-fidelity models [13, 14]. Various research efforts have been devoted

58 to multi-fidelity assisted stochastic simulation. One class of multi-fidelity methods that have
 59 been thoroughly investigated is the control variate approach [15], including multi-level Monte
 60 Carlo [16, 17, 18] and multi-fidelity Monte Carlo [13, 19, 20]. The multi-level Monte Carlo
 61 method combines the outputs of a hierarchy of coarser grid solutions using control vari-
 62 ates, where the trade-off between computational efficiency and accuracy is adjusted by a
 63 parameter, e.g., the mesh size [19]. The multi-fidelity Monte Carlo extends the multi-level
 64 Monte Carlo by allowing arbitrary low-fidelity models that are not limited to coarse grid dis-
 65 cretizations. Multi-fidelity Monte Carlo has also been examined for seismic risk assessment
 66 applications where the quantity of interest is the expected value of risk consequence measures
 67 [21]. However, one drawback of multi-fidelity Monte Carlo is that it requires a linear depen-
 68 dency between the model response and can only provide asymptotic error estimates [22].
 69 Another category of multi-fidelity methods is based on the use of surrogate models. The idea
 70 underpinning these approaches is to integrate information from the low-fidelity models to
 71 improve the prediction accuracy of a single-fidelity surrogate trained on scarce high-fidelity
 72 data [23]. Once trained, this fast-to-evaluate multi-fidelity surrogate model will be used in
 73 place of the expensive high-fidelity model for the subsequent stochastic simulation. One of
 74 the most popular multi-fidelity surrogate models is the multi-fidelity Gaussian process model
 75 (i.e., Cokriging) [24, 25, 26]. Its application in stochastic simulation is being actively studied
 76 [27, 28]. Nevertheless, since the multi-fidelity surrogate models directly map the model input
 77 to the model output, they often suffer from the well-known curse of dimensionality [29]. This
 78 limits its application to problems involving high stochastic dimensionality [22]. Bayesian
 79 multi-fidelity Monte Carlo, another category of multi-fidelity stochastic simulation, was first
 80 proposed in [30] with accuracy and efficiency in uncertainty quantification demonstrated on
 81 complex and large-scale bio-mechanical problems [31]. Bayesian multi-fidelity Monte Carlo
 82 directly predicts the quantitative relationship between the low-fidelity model output and the
 83 high-fidelity model output using a non-parametric Bayesian model, which is then used to
 84 efficiently compute the statistics of the high-fidelity model output [30, 31]. However, since
 85 it allows a more flexible dependence between the low-fidelity and high-fidelity model (e.g.,
 86 non-Gaussian), the posterior distribution of the high-fidelity output given the corresponding
 87 low-fidelity output is computationally intractable, and thus advanced stochastic simulation

techniques are required to sample from the posterior distribution. To improve the accuracy of the regression model in a small data scenario, the Bayesian multi-fidelity model was further enhanced by including informative features in the model input [22]. Compared to the multi-fidelity surrogate model, one significant advantage of these Bayesian multi-fidelity Monte Carlo methods is that they can address high-dimensional stochastic simulation problems, as they only operate in the model output space or by only adding small dimensional features [12, 22]. In addition, they can accommodate more complex correlations between the low-fidelity and high-fidelity models, and thus are more general than the multi-fidelity methods based on the control variate.

This work proposes a multi-fidelity stochastic simulation scheme to efficiently compute small failure probabilities of structural systems within the setting of stratified sampling. More specifically, a low-fidelity model with a fast approximation of the system model is first developed. Then a probabilistic relationship between the low-fidelity response and the high-fidelity response is established based on a small number of high-fidelity runs by applying a Gaussian process regression (GPR) model to estimate the exceedance probability conditional on non-collapse. Similarly, a Gaussian process classification (GPC) model is established to predict whether the system collapses given any low-fidelity response. Active learning strategies are developed to intelligently generate training data for constructing the multi-fidelity models. Once calibrated, the multi-fidelity models are used to directly predict the high-fidelity response and system collapse from low-fidelity samples, based on which the probabilities of interest are estimated. The proposed multi-fidelity model offers a distinct advantage over existing methods by directly predicting the high-fidelity response, or system collapse, based on the low-fidelity response. Indeed, in contrast to many commonly adopted multi-fidelity surrogate models that focus on mapping the model input to the model output, the approach in this work operates solely in the model output space. This enables the proposed scheme to effectively handle high-dimensional stochastic simulation problems. Nevertheless, it should be noted that the multi-fidelity stochastic simulation scheme of this work may introduce bias in the estimation of failure probabilities due to the use of Gaussian process models to approximate the high-fidelity response and system collapse.

2. Problem Setting

Consider a high-fidelity structural system that maps input, $\mathbf{x} = [x^{(1)}, x^{(2)}, \dots, x^{(n_x)}] \in \Omega \subset \mathcal{R}^{n_x}$, to outputs, $y_h(\mathbf{x}) \in \mathcal{R}$, where n_x is the input dimension. The model input $\mathbf{x}$ is characterized by a probability density function $p(\mathbf{x})$. The high-fidelity system model is assumed to be based on numerical simulations (e.g., finite element models) and, in order to capture the system behavior with high accuracy, computationally demanding to evaluate. Alternatively, a low-fidelity model $y_l(\mathbf{x}_l)$, which is inexpensive to run, can be established to approximate the high-fidelity model output. Generally, the low-fidelity model may have a different modeling approach from the high-fidelity model, and some variables in the lower-fidelity model might be omitted [32]. Therefore, the model input of the low-fidelity model is assumed as a subset of the model input of the high-fidelity model, i.e., $\mathbf{x}_l = [x^{(1)}, x^{(2)}, \dots, x^{(n_{xl})}] \in \Omega_l \subset \mathcal{R}^{n_{xl}}$ where $n_{xl} \leq n_x$ and $\mathbf{x}_l \subseteq \mathbf{x}$.

The problem of interest to this work is to develop a multi-fidelity stochastic simulation scheme for estimating the failure probabilities of the system, described by the probability of the response, y_h , exceeding a threshold δ , i.e., as $P(y_h > \delta)$. Without loss of generality, the proposed scheme explicitly takes into account both non-collapse samples $\mathcal{D}_{nc}$ and collapse samples $\mathcal{D}_c$. Since $\mathcal{D}_{nc}$ and $\mathcal{D}_c$ are mutually exclusive, based on the total probability theorem, the failure probability can be expressed as:

$$P(y_h > \delta) = P(y_h > \delta | \mathcal{D}_{nc})(1 - P(\mathcal{D}_c)) + P(y_h > \delta | \mathcal{D}_c)P(\mathcal{D}_c) \quad (1)$$

where $P(y_h > \delta | \mathcal{D}_{nc})$ and $P(y_h > \delta | \mathcal{D}_c)$ are the exceedance probabilities conditional on non-collapse, $\mathcal{D}_{nc}$, and collapse, $\mathcal{D}_c$. $P(\mathcal{D}_c)$ is the collapse probability (i.e., probability of system collapse), which can be determined through a combination of indicators such as non-convergence of the numerical model, the deformed shape of the structure at the last converged time step, and the peak roof drift [33]. For the sake of simplicity, when the system collapses, y_h is always considered to exceed δ , i.e., $P(y_h > \delta | \mathcal{D}_c) \equiv 1$. The goal of this work is to therefore estimate probabilities $P(y_h > \delta | \mathcal{D}_{nc})$ and $P(\mathcal{D}_c)$.

By propagating uncertainties through the high-fidelity model, the exceedance probability conditional on non-collapse (i.e., $P(y_h > \delta | \mathcal{D}_{nc})$) can be directly calculated. For the sake of

144 brevity, let $P_{nc}(y_h > \delta) \equiv P(y_h > \delta | \mathcal{D}_{nc})$. The probability of interest is therefore:

$$P_{nc}(y_h > \delta) = \int \mathbf{I}_f(y_h(\mathbf{x}))p(\mathbf{x})d\mathbf{x}, \quad \mathbf{I}_f(y_h(\mathbf{x})) = \begin{cases} 1 & y_h(\mathbf{x}) > \delta \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

145 where $I_f(\cdot)$ is an indicator function that determines whether the high-fidelity response y_h
 146 given $\mathbf{x}$ exceeds the threshold δ . Similarly, the collapse probability can be estimated as:

$$P(\mathcal{D}_c) = \int \mathbf{I}_c(\mathbf{x})p(\mathbf{x})d\mathbf{x}, \quad \mathbf{I}_c(\mathbf{x}) = \begin{cases} 1 & \text{system collapse} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

147 where $I_c(\cdot)$ is an indicator function that determines whether the system collapses or not.
 148 Since the probabilities in Eq.(2) and Eq.(3) cannot be calculated in closed form for most ap-
 149 plications, they are typically estimated using stochastic simulation techniques such as Monte
 150 Carlo simulation (MCS). The Monte Carlo estimators for the conditional failure probability
 151 and the collapse probability are given by:

$$\hat{P}_{nc}(y_h > \delta) \approx \frac{1}{N - N_c} \sum_{i=1}^{N-N_c} \mathbf{I}_f(y_h(\mathbf{x}_i)) \quad (4)$$

152

$$\hat{P}(\mathcal{D}_c) \approx \frac{1}{N} \sum_{i=1}^N \mathbf{I}_c(\mathbf{x}_i) = \frac{N_c}{N} \quad (5)$$

153 where $\mathbf{x}_i$ is the i th MCS sample generated from $p(\mathbf{x})$, i.e., regions of Ω , N is the total number
 154 of MCS samples, and N_c is the total number of collapse samples. Finally, the total failure
 155 probability can be estimated as:

$$\begin{aligned} \hat{P}(y_h > \delta) &= \hat{P}_{nc}(y_h > \delta)(1 - \hat{P}(\mathcal{D}_c)) + \hat{P}(\mathcal{D}_c) \\ &\approx \frac{1}{N - N_c} \sum_{i=1}^{N-N_c} \mathbf{I}_f(y_h(\mathbf{x}_i)) \left(1 - \frac{1}{N} \sum_{i=1}^N \mathbf{I}_c(\mathbf{x}_i) \right) + \frac{1}{N} \sum_{i=1}^N \mathbf{I}_c(\mathbf{x}_i) \end{aligned} \quad (6)$$

156 For each generated sample $\mathbf{x}_i$, the computationally expensive high-fidelity model needs to be
 157 evaluated to obtain the corresponding model output $y_h(\mathbf{x}_i)$. In particular, for the estimation

of small failure probabilities, a large number of samples is generally required, especially if high accuracy is desired. As a result, MCS-based estimation of $P(y_h > \delta)$ can easily lead to computational demands that exceed available resources.

3. Stratified Sampling-based Monte Carlo Simulation

Stratified sampling is a well-known variance reduction technique. The basic idea is to partition Ω into a set of mutually exclusive and collectively exhaustive events, $\{\Omega_j\}_{1 \leq j \leq L} \ni \bigcup_{j=1}^L \Omega_j = \Omega$ and $\Omega_j \cap \Omega_k = \emptyset$, for $j \neq k$, termed strata. Sample allocation entails prescribing the number of samples, N_j , for each j th stratum while ensuring the satisfaction of the constraint: $\sum_{j=1}^L N_j = N$. As shown in [8], the optimal allocation of samples between the strata can be found by solving a convex optimization problem defined in terms of a preliminary study. In particular, it can be shown that proportional allocation (i.e. $N_j = NP(\Omega_j)$) guarantees variance reduction, however, the efficiency gains are insignificant when the samples from high-probability strata have a small influence on the failure probability.

Once the samples are drawn from the strata, they are weighted and combined to obtain the stratified sampling estimator, $\hat{P}(y_h > \delta)$ as:

$$\hat{P}(y_h > \delta) = \sum_{i=1}^L \hat{P}(y_h > \delta \mid \Omega_j) P(\Omega_j) \quad (7)$$

If carried out in the space of the input random variables, the process of stratification is performed by defining a mutually exclusive and collectively exhaustive partitioning of a partial set of the input uncertainties, $\mathbf{x}$, which are termed stratification variables. This implies that $P(\Omega_j)$ can be readily calculated using $p(\mathbf{x})$. In Eq. (7), $\hat{P}(y_h > \delta \mid \Omega_j)$ is the estimate of the failure probability conditional on Ω_j , and can be obtained by evaluating Eq. (6) in the j th stratum, and therefore as:

$$\begin{aligned} \hat{P}(y_h > \delta \mid \Omega_j) &= \hat{P}_{nc}(y_h > \delta \mid \Omega_j)(1 - \hat{P}(\mathcal{D}_c \mid \Omega_j)) + \hat{P}(\mathcal{D}_c \mid \Omega_j) \\ &\approx \frac{1}{N_j - N_{jc}} \sum_{i=1}^{N_j - N_{jc}} \mathbf{I}_f(y_h(\mathbf{x}_i)) \left(1 - \frac{1}{N_j} \sum_{i=1}^{N_j} \mathbf{I}_c(\mathbf{x}_i) \right) + \frac{1}{N_j} \sum_{i=1}^{N_j} \mathbf{I}_c(\mathbf{x}_i) \end{aligned} \quad (8)$$

179 where $\hat{P}_{nc}(y_h > \delta \mid \Omega_j)$ is the estimate of the exceedance probability conditional on non-
180 collapse and the i th stratum, $\hat{P}(\mathcal{D}_c \mid \Omega_j)$ is the estimate of the collapse probability condi-
181 tional on the i th stratum, while N_{jc} is the number of non-collapse samples drawn from the
182 j th stratum. In particular, the strata-wise sample generation can be performed from the
183 conditional sampling density function defined for the i th stratum as:

$$p_j(\mathbf{x}) = \begin{cases} p(\mathbf{x})/P(\Omega_j) & \mathbf{x} \in \Omega_j \\ 0 & \mathbf{x} \notin \Omega_j \end{cases} \quad (9)$$

184 The gains from stratification are higher when the stratification variables are well-correlated
185 with the target responses such that the strata are internally more homogeneous (in the re-
186 sponse values) than Ω as a whole [8].

187 It is important to note that a k -dimensional stratification with L strata per variable will
188 result in L^k strata. If k and L are high-dimensional, this can lead to significant sampling
189 demands for achieving near-optimal sample allocation and potentially affect the overall effi-
190 ciency in the estimation of the failure probabilities [8, 34]. Therefore, it is recommended to
191 select a limited set of stratification variables that are strongly correlated with the responses
192 of interest [34]. In natural hazards applications, these can take the form of hazard intensity
193 measures such as the maximum wind speed and spectral acceleration. Further details on the
194 choice of stratification variables for reliability problems in natural hazards engineering can be
195 found in [8, 34]. Despite this, stratified sampling offers several advantages in the estimation of
196 small failure probabilities. Indeed, it enables simultaneous estimation of failure probabilities
197 and has been shown to be not only efficient but also unbiased, consistent (in the sense that it
198 approaches the true failure probability for large sample sizes), and asymptotically Gaussian
199 [8].

200 4. A Multi-fidelity Stochastic Simulation Scheme

201 4.1. Overview

202 Although significant gains in efficiency can be obtained through the application of strat-
203 ified sampling, it still requires repeated evaluations of the expensive high-fidelity model. To

address this, a multi-fidelity stochastic simulation scheme within the setting of the stratified sampling framework of Sec. 3 is proposed in this section. More specifically, two multi-fidelity models are developed to predict the conditional exceedance probabilities $\hat{P}_{nc}(y_h > \delta \mid \Omega_j)$ and $\hat{P}(\mathcal{D}_c \mid \Omega_j)$ for each stratum.

4.2. Conditional Non-collapse Exceedance Probability

The first multi-fidelity model aims to construct a quantitative link between the low-fidelity model response y_l and the high-fidelity model response y_h using a probabilistic regression model. The conditional non-collapse exceedance probability, $P_{nc}(y_h > \delta \mid \Omega_j)$, is then estimated from low-fidelity evaluations and the calibrated multi-fidelity model.

To achieve this goal, the conditional non-collapse exceedance probability for the j th stratum is expanded to include the low-fidelity model output [30], and therefore written as:

$$\begin{aligned} P_{nc}(y_h > \delta \mid \Omega_j) &= \int P(y_h > \delta \mid y_l, \Omega_j) p_j(y_l) dy_l \\ &= \int \left(\int I_f(y_h) p_j(y_h \mid y_l) dy_h \right) p_j(y_l) dy_l \end{aligned} \quad (10)$$

where $p_j(y_l)$ is the probability density of the low-fidelity model output conditional on the j th stratum, and $p_j(y_h \mid y_l)$ is the probability density of the high-fidelity model output y_h given the low-fidelity model output y_l evaluated at the same model input $\mathbf{x}$ generated from $p_j(\mathbf{x})$ of Eq. (9). Eq. (10) implies that estimating $P_{nc}(y_h > \delta \mid \Omega_j)$ using the inexpensive low-fidelity model requires two important steps: (i) obtaining $p_j(y_l)$ by propagating uncertainties through the low-fidelity model; and (ii) obtaining $p_j(y_h \mid y_l)$ by constructing a probabilistic relationship between the low-fidelity model response and the high-fidelity model response. Since the low-fidelity model is cheap to evaluate, $p_j(y_l)$ can be efficiently estimated by MCS within the j th stratum. The non-trivial part is to quantify the underlying relationship between y_l and y_h in a probabilistic sense. A probabilistic regression model is used for this purpose in this work. The regression model takes the low-fidelity model output y_l as input and the high-fidelity model output y_h as output and is therefore referred to as a multi-fidelity model. It should be noted that the model outputs y_l and y_h have a noisy relationship instead of a one-to-one mapping. In other words, the same y_l may correspond to different y_h values, and vice versa, since the low-fidelity model simplifies certain properties of the high-fidelity model.

4.2.1. Gaussian Process Regression of the Multi-fidelity Model

In order to infer the conditional probability density $p_j(y_h|y_l)$ from the j th stratum, a probabilistic regression model can be established to predict the underlying relationship between the low-fidelity model output and the high-fidelity model output. To this end, a Gaussian process regression (GPR) [35] is used in this work due to its flexibility in modeling complex functions and its ability to provide a full posterior predictive distribution.

In the current problem, given a specific low-fidelity output y_l , the high-fidelity output, y_h , is non-deterministic due to the noisy relationship between the two models. In contrast to a typical GPR, which aims to approximate a deterministic function output given a set of observations, the GPR in this work is used to predict a probabilistic relationship. The core principle of a typical GPR is to assume the target function y_h as a realization of a regression model $m(y_l)$ and a Gaussian process $\varepsilon(y_l)$. To account for the probabilistic/noisy relationship between y_l and y_h , an additional noise term ϵ is explicitly assumed and added to the typical GPR [36]. The GPR model used in this work is then written as:

$$y_h = m(y_l) + \varepsilon(y_l) + \epsilon \quad (11)$$

where $\varepsilon(y_l) \sim \mathcal{N}(0, k(y_l, y_l'))$, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$

The regression model $m(y_l)$ is selected as a linear combination of basis functions, expressed by $m(y_l) = \mathbf{f}(y_l)^T \boldsymbol{\beta}$, where $\mathbf{f}(y_l)$ is the q -dimensional vector of basis functions and $\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_q]^T$ is the vector of regression coefficients. Typically, polynomials of y_l can be used as the basis functions, e.g., linear or quadratic functions of y_l . After the mean function is set, the property of the GPR model is fully determined by the covariance function, or the kernel, $k(y_l, y_l') = \sigma^2 \Psi(y_l, y_l')$, where σ^2 is the variance and $\Psi(y_l, y_l')$ is the correlation function. A commonly used kernel function is the squared exponential kernel (i.e., Gaussian kernel):

$$k(y_l, y_l') = \sigma^2 \exp\left(-\frac{|y_l - y_l'|^2}{2\theta^2}\right) \quad (12)$$

where θ is the scale correlation parameter (i.e., length-scale) in the correlation function. Regarding the noise term ϵ , a zero-mean Gaussian noise with variance σ_ϵ^2 is typically assumed.

The prior distribution over the GPR-based multi-fidelity model in Eq. (11) is then updated

by a set of given data points, i.e., the calibration of the uncertain model parameters in the mean function, the covariance function, and the noise term including regression coefficients $\boldsymbol{\beta}$, variance σ^2 , length-scale θ , and noise variance σ_ϵ^2 . The calibration of the GPR model first requires the creation of a database of n observations based on the low-fidelity and high-fidelity models. Specifically, it involves evaluations of the high-fidelity model output $\mathbf{Y}_h = \{y_h(\mathbf{x}_t); t = 1, \dots, n_t\}$ for different inputs $\mathbf{X} = \{\mathbf{x}_t; t = 1, \dots, n_t\}$, and evaluations of the low-fidelity model output $\mathbf{Y}_l = \{y_l(\mathbf{x}_t); t = 1, \dots, n_t\}$ for different inputs $\mathbf{X}_l = \{\mathbf{x}_{lt}; t = 1, \dots, n_t\}$. The set $\{\mathbf{Y}_l, \mathbf{Y}_h\}$ is the so-called training set. The selection of the training data will impact the accuracy of the multi-fidelity model. To ensure the model has good accuracy over the design space, a space-filling strategy combined with an active learning strategy is adopted. More details on the selection of the training data will be discussed later. Finally, in order to calibrate the unknown model parameters using the training data, maximum likelihood estimation (MLE) is used to obtain point estimates of the parameters. Due to the Gaussian properties, the likelihood function is given by:

$$L(\boldsymbol{\beta}, \sigma^2, \theta, \sigma_\epsilon^2 | \mathbf{Y}_h) = \frac{1}{(2\pi\sigma^2)^{n_t/2} |\boldsymbol{\Psi} + (\sigma^2/\sigma_\epsilon^2)\mathbf{I}_{n_t}|^{1/2}} \times \exp \left[-\frac{1}{2\sigma^2} (\mathbf{Y}_h - \mathbf{F}\boldsymbol{\beta})^T (\boldsymbol{\Psi} + (\sigma^2/\sigma_\epsilon^2)\mathbf{I}_{n_t})^{-1} (\mathbf{Y}_h - \mathbf{F}\boldsymbol{\beta}) \right] \quad (13)$$

where $\mathbf{F} = [\mathbf{f}(y_{l1}), \mathbf{f}(y_{l2}), \dots, \mathbf{f}(y_{ln_t})]^T$, $\boldsymbol{\Psi}$ is the correlation matrix evaluated over $\mathbf{Y}_l$, and $\mathbf{I}_{n_t}$ is the identity matrix of size $n_t \times n_t$.

Conditional on the observations $\mathbf{Y}_h$ and the optimal selection of the model parameters, the GPR model gives a prediction at any new low-fidelity model output y_{l*} that follows a Gaussian distribution with mean $\hat{y}_h(y_{l*})$ and variance $\sigma_{\hat{y}_h}^2(y_{l*})$ given by:

$$\hat{y}_h(y_{l*}) = \mathbf{f}(y_{l*})^T \boldsymbol{\beta}^* + \mathbf{k}(y_{l*})^T \mathbf{K}^{-1} (\mathbf{Y}_h - \mathbf{F}\boldsymbol{\beta}^*) \quad (14)$$

274

$$\sigma_{\hat{y}_h}^2(y_{l*}) = k(y_{l*}, y_{l*}) - \mathbf{k}(y_{l*})^T \mathbf{K}^{-1} \mathbf{k}(y_{l*}) + \mathbf{u}(y_{l*})^T (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} \mathbf{u}(y_{l*}) + (\sigma_\epsilon^2)^* \quad (15)$$

where $\mathbf{k}(y_{l*})$ is the n_t -dimensional covariance vector between y_{l*} and each element in $\mathbf{Y}_l$, given by $\mathbf{k}(y_{l*}) = [k(y_{l*}, y_{l1}), k(y_{l*}, y_{l2}), \dots, k(y_{l*}, y_{ln_t})]^T$; $\mathbf{K}$ is the covariance matrix evaluated over $\mathbf{Y}_l$; while $\mathbf{u}(y_{l*}) = \mathbf{F}^T \mathbf{K}^{-1} \mathbf{k}(y_{l*}) - \mathbf{f}(y_{l*})$. Here $\boldsymbol{\beta}^* = (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{F})^{-1} (\mathbf{F}^T \mathbf{K}^{-1} \mathbf{Y}_h)$ corresponds to

the maximum likelihood estimate of the parameter β .

Similar to a noise-free GPR model, the predictive variance of the proposed GPR model contains a part (i.e., the first three terms on the right-hand side of Eq. (15)) that quantifies the uncertainty on the mean prediction. Apart from this, the predictive variance of the proposed GPR model also includes a term $(\sigma_\epsilon^2)^*$, which corresponds to the contribution from the noise assumed in Eq. (11). This term captures the inherent uncertainty of the high-fidelity model output y_h given a low-fidelity model output y_l evaluated at the common model input $\mathbf{x}$.

4.2.2. Homoscedastic and Heteroscedastic Noise

The above GPR-based multi-fidelity model assumes the variance of the noise term ϵ to be constant, i.e., homoscedastic. However, in many cases, the noise could vary with the model input (i.e., y_l in this study). Therefore, a GPR with heteroscedastic noise can be constructed in which the noise ϵ is assumed to follow a zero-mean Gaussian distribution with a non-constant variance $\epsilon \sim \mathcal{N}(0, \sigma^2 \gamma(y_l))$, where $\gamma(y_l)$ is a scaling factor. If prior information about the noise is available, it can be incorporated in selecting $\gamma(y_l)$. Otherwise, to avoid model over-fitting and challenges in optimizing the hyperparameters in $\gamma(y_l)$, a parsimonious selection of $\gamma(y_l)$ is suggested [36]. For this purpose, a second-order polynomial is chosen in this study to describe the scaling factor, i.e., $\gamma(y_l) = \beta_{\gamma 1} + \beta_{\gamma 2} y_l + \beta_{\gamma 3} y_l^2$. In the scaling factor, the coefficients (i.e., hyperparameters) $\beta_{\gamma 1}$, $\beta_{\gamma 2}$, and $\beta_{\gamma 3}$ can be calibrated using MLE, where the likelihood function is Eq. (13) with the variance of the noise term σ_ϵ^2 replaced by $\sigma^2 \gamma(y_l)$. The predictive mean of the heteroscedastic GPR model (i.e., the GPR model with heteroscedastic noise) is the same as the homoscedastic GPR model (i.e., the GPR with homoscedastic noise), given by Eq. (14). The predictive variance of the heteroscedastic GPR model is given by Eq. (15) with $(\sigma_\epsilon^2)^*$ replaced by $(\sigma^2)^*(\beta_{\gamma 1}^* + \beta_{\gamma 2}^* y_l + \beta_{\gamma 3}^* y_l^2)$, where $(\sigma^2)^*$, $\beta_{\gamma 1}^*$, $\beta_{\gamma 2}^*$, $\beta_{\gamma 3}^*$ are the maximum likelihood estimates of the unknown parameters.

4.2.3. Active Learning for Optimal Selection of Training Data

The quality of the estimated conditional exceedance probability largely depends on the predictive accuracy of the constructed GPR model in terms of both the predictive mean and the variance. Typically, to ensure good accuracy over the entire design space (i.e. all

possible low-fidelity model outputs), a selection that evenly fills the design space can be used. Due to the significant computational cost of the high-fidelity model, it is of interest to use a small number of training data to obtain the desired prediction accuracy for approximating the underlying relationship between y_l and y_h . As a result, randomly selecting low-fidelity model output samples to label (i.e., identify for corresponding high-fidelity model runs) is not sufficient, since non-informative samples can be selected while informative samples (e.g., in the region with higher uncertainty in y_l) could be overlooked [37].

To address this issue, an active learning sampling strategy, which enables a more wise allocation of the training data, is proposed. The overall idea of the sampling strategy is to thoroughly explore the region where the prediction uncertainty (i.e., variance), $\sigma_{y_h}^2$, is large. The active learning process begins by constructing a candidate pool comprising strata-wise MCS samples of $\mathbf{x}_l$ (representing the low-fidelity model input). The low-fidelity model is evaluated to obtain corresponding responses of these samples. Subsequently, a learning function (introduced subsequently) is computed for each low-fidelity response to assess its informativeness. Guided by the learning function, the optimal low-fidelity sample is chosen. It is important to note that the low-fidelity model input, $\mathbf{x}_l$, is assumed to be a subset of the high-fidelity model input, $\mathbf{x}_h$. To acquire the corresponding high-fidelity response of the optimal sample, $\mathbf{x}_l$ is augmented by randomly sampling the missing components defining $\mathbf{x}_h$. By evaluating the high-fidelity model for the augmented sample, the corresponding response is obtained. The resulting pairs of low-fidelity and high-fidelity responses are then incorporated into the training set for model refinement. The GPR model is subsequently updated using the expanded training data set. The active learning process follows an iterative approach, where the steps outlined above are repeated until a predefined stopping criterion is satisfied. More specifically, The sampling strategy adopted here is mathematically expressed as:

$$y_l^{**} = \operatorname{argmax}_{y_l \in \mathbf{Y}_{l,candi}} U(y_l) \quad (16)$$

where $U(y_l) = \sigma_{y_h}^2(y_l)$ is the learning function, and $\mathbf{Y}_{l,candi}$ are the strata-wise MCS samples from the low-fidelity model in each active learning iteration, excluding those already selected. The philosophy underlying the active learning strategy is that the region with higher predic-

tion variance is not only where uncertainty exists on the mean prediction, but also where a high level of noise in the underlying relationship between y_l and y_h exists. Allocating training samples to this region will, therefore, be more efficient in producing a globally accurate model. However, a disadvantage of such an active learning scheme is that it can lead to clustering of the added samples. To avoid clustered samples, the following adjusted learning function based on the weighted predictive variance is proposed:

$$U(y_{li}) = \frac{d_{min,i}}{\sum_{i=1}^{N_{candi}} d_{min,i}} \sigma_{\hat{y}_h}^2(y_{li}) \quad (17)$$

where y_{li} is the i th sample from the candidate pool $\mathbf{Y}_{l,candi}$, and $d_{min,i}$ is the minimum distance between the i th candidate sample and all the training samples in $\mathbf{Y}_l$. The adjusted learning function penalizes the candidate samples which are very close to the samples that are already in the training set and thus avoids adding samples with similar y_l values.

Before applying active learning sampling, an initial set of training data is required. In surrogate modeling, Latin Hypercube Sampling (LHS) is often used to this end. However, in the proposed GPR-based multi-fidelity model, the input is the low-fidelity output, y_l , which has an unknown design space. Therefore, a “near” space-filling sampling strategy [22] is applied in which the low-fidelity samples, y_l , are first sorted and divided into a number of bins. Subsequently, an equal number of samples are randomly selected from each bin, therefore, ensuring the initial training data has adequate space-filling properties.

The active learning process continues until a stopping criterion is satisfied. Different types of stopping criteria have been proposed for reliability analysis applications. They can be grouped into learning function-based criterion, accuracy of failure probability-based criterion, and criteria based on the stability of failure probability or limit state surface [38]. In this work, the following stopping criterion based on the stability of the multi-fidelity model is adopted:

$$\Delta = \frac{|\hat{v}_p - \hat{v}_{p-1}|}{\hat{v}_p + \alpha_v} \leq \alpha \quad (18)$$

where $\hat{v}_p$ and $\hat{v}_{p-1}$ are the average predictions over the selected inputs, y_l , in the p th and $(p-1)$ th iterations; α_v is a small value to avoid numerical issue in calculating Δ when $\hat{v}_p$ is zero which can be chosen as 10^{-8} [39]; while α is the convergence threshold. The selected

360 y_l values are uniformly distributed between an interval determined by the minimum and
 361 maximum y_l of the non-collapse MCS samples from each stratum. Eq. (18) is satisfied when
 362 the difference between the averaged prediction in two consecutive iterations is smaller than
 363 α , i.e., the prediction becomes stable. A potential issue is an early termination when the
 364 initial GPR model is inaccurate [38]. One remedy is to track the variation of the prediction
 365 in multiple iterations through the modification:

$$\Delta_t = \frac{|\hat{v}_p - \hat{v}_{p-q}|}{\hat{v}_p + \alpha_v} \leq \alpha \quad (19)$$

366 where for Eq. (19) to be deemed satisfied, it must be satisfied for $q = 1, 2, \dots, N_c$ where N_c is
 367 set by the user. It should be noted that the adaptive sampling stops only when the criterion
 368 in Eq. (19) is satisfied for both the predictive mean $\hat{y}_h$ and standard deviation $\sigma_{\hat{y}_h}$.

369 4.3. Conditional Collapse Probability

370 4.3.1. Overview

371 The estimation of the probability of system collapse conditional on the j th stratum can
 372 be described as a binary classification task. In this task, collapse is labeled as 1 while non-
 373 collapse is labeled as 0. Similar to the estimation of the conditional non-collapse exceedance
 374 probability, it is proposed to efficiently evaluate the low-fidelity model to obtain MCS so-
 375 lutions in each stratum, which are then used to inform through a constructed classification
 376 model whether the system collapses and estimate the stratum conditional collapse probabilit-
 377 ity. The multi-fidelity model takes the low-fidelity response, y_l , as the model input while
 378 providing as output the outcome of the indicator function $\mathbf{I}_c = 1$ or 0. To associate the low-
 379 fidelity response with system collapse, the conditional collapse probability, $P(\mathcal{D}_c|\Omega_j)$, can be
 380 rewritten as:

$$P(\mathcal{D}_c|\Omega_j) = \int \mathbf{I}_c(y_l)p_j(y_l)dy_l \quad (20)$$

381 For a given sample $\mathbf{x}$, the low-fidelity model is first evaluated to obtain $y_l(\mathbf{x})$. Then, based
 382 on the constructed classification model, whether the system collapses or not (i.e., $\mathbf{I}_c$) can
 383 be quickly predicted from the low-fidelity response $y_l(\mathbf{x})$. The estimator of the conditional
 384 collapse probability remains the same as that reported in Eq. (8).

4.3.2. Gaussian Process Classifier of the Multi-fidelity Model

The multi-fidelity model aims to construct a classifier to predict the indicator function $\mathbf{I}_c$ given any low-fidelity response y_l . To this end, a Gaussian process classification (GPC) model is used here. In contrast to other classification models, such as support vector machines, GPC provides a full predictive distribution that can be further used for active learning [40]. The idea behind the GPC is to assume a Gaussian process prior over a latent function, $f(y_l)$, and then transform it through a logit or probit function to obtain the prior over the target function $\mathbf{I}_c$ [35].

It should be noted that $f(y_l)$ cannot be observed which, however, is not an issue. The adoption of $f(y_l)$ is only for the convenience of formulating the GPC model. Following GPC convention, a zero-mean Gaussian process is assumed over the latent function, $f(y_l)$, i.e., $f(y_l) \sim \mathcal{N}(0, k(y_l, y_l'))$ where $k(y_l, y_l')$ is again the kernel function with length-scale θ and variance σ^2 . The Bayesian inference of GPC involves firstly calculating the predictive distribution over the latent function f over any new low-fidelity output y_{l*} and subsequently using this distribution to compute the probabilistic prediction over the corresponding indicator function $\mathbf{I}_{c*}$, i.e., $p(\mathbf{I}_{c*} = 1 | \mathbf{Y}_l, \mathcal{I}_c, y_{l*})$. More details on the GPC inference are presented in the Appendix. Finally, the predicted class label, $\hat{\mathbf{I}}_c(y_{l*})$, (i.e., whether or not the system collapses) from the multi-fidelity classification model can be obtained by applying a threshold to the predictive class probability, i.e., $\hat{\mathbf{I}}_c(y_{l*}) = 1$ if $p(\mathbf{I}_{c*} = 1 | \mathbf{Y}_l, \mathcal{I}_c, y_{l*}) > 0.5$ and $\hat{\mathbf{I}}_c(y_{l*}) = 0$, otherwise.

4.3.3. Active Learning for Optimal Selection of Training Data

Since the GPC provides a probabilistic prediction over the class labels, an active learning strategy can be used to adaptively select the training data. The goal is to add samples that can maximize the capability of the classification model in terms of discriminating collapse samples from non-collapse samples, which ultimately reduces the cost of running the high-fidelity model. The most commonly used active learning for classification is to use the uncertainty measure which selects the sample where the classification model is the most uncertain to label [37]. For the GPC model, the sample with posterior predictive class probability, $p(\mathbf{I}_{c*} = 1 | \mathbf{Y}_l, \mathcal{I}_c, y_{l*})$, close to 0.5 should be where the classification model is the

most uncertain. Based on Eq. (26), this learning function is then transformed to $\frac{|\hat{\mu}(y_l)|}{\sqrt{1+\sigma_\mu^2(y_l)}}$ close to 0 [40, 41]. The learning function of active learning for GPC is written subsequently written as:

$$y_l^{**} = \operatorname{argmax}_{y_l \in \mathbf{Y}_{l, \text{candi}}} U(y_l) \quad (21)$$

where $U(y_l) = \frac{|\hat{\mu}(y_l)|}{\sqrt{1+\sigma_\mu^2(y_l)}}$. The active learning process continues until a stopping criterion is reached. A similar stopping criterion as reported in Eq. (19) is used in this work.

4.4. Overall Algorithm

The overall algorithm of the proposed multi-fidelity stochastic simulation scheme is shown in Fig. 1. The algorithm starts with training the classification model (described in Sec. 4.3.2) using the active learning described in Sec. 4.3.3. Once the classification model is calibrated, the algorithm proceeds to the regression part which trains the regression model (described in Sec. 4.2.1) with the active learning strategy described in Sec. 4.2.3. It is noteworthy to point out that the regression model can be pre-trained using the non-collapse samples when the classification model is being trained, or it can start to be trained after the classification model is calibrated. In addition, the trained classification model is used during the training of the regression model to determine if a selected sample is a collapse or non-collapse sample. The high-fidelity model is then only evaluated if the selected sample is predicted as non-collapse. This helps in reducing the computational effort of running the high-fidelity model. To avoid confusion, in the following, the GPR-based multi-fidelity model will be referred to as the multi-fidelity regression model, and the GPC-based multi-fidelity model as the multi-fidelity classification model. It should be noted that this multi-fidelity stochastic simulation scheme is performed within each stratum of the stratification. Once the multi-fidelity classification and regression models are trained, they can be used directly to predict the conditional collapse probability, $\hat{P}(\mathcal{D}_c | \Omega_j)$, and the conditional non-collapse exceedance probability, $\hat{P}_{nc}(y_h > \delta | \Omega_j)$, of Eq. (8). More specifically, N_j MCS samples of $\mathbf{x}$ are first generated from $p_j(\mathbf{x})$, and then the low-fidelity model is evaluated to obtain samples of $y_l(\mathbf{x})$. The next step is to use the trained multi-fidelity models to predict the conditional collapse probability $\hat{P}(\mathcal{D}_c | \Omega_j)$ and the conditional non-collapse exceedance probability, $\hat{P}_{nc}(y_h > \delta | \Omega_j)$, given respectively

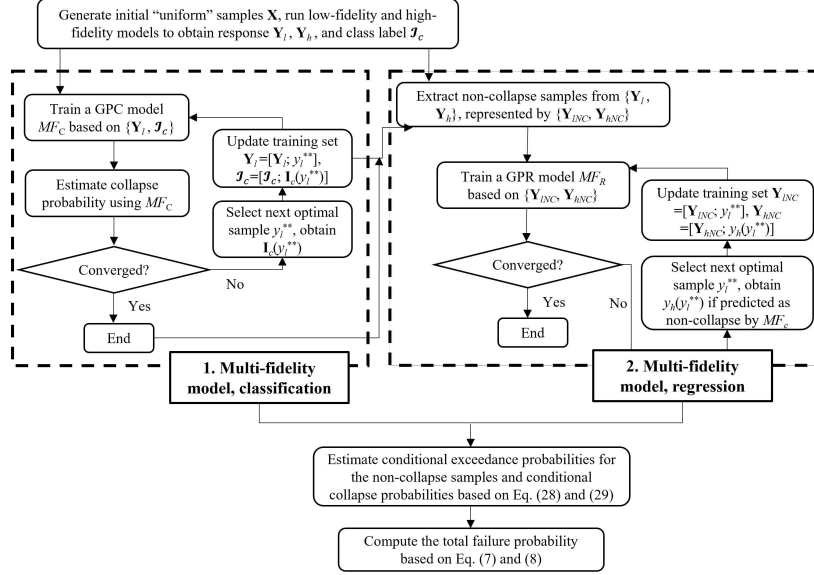


Figure 1: Flowchart of the proposed multi-fidelity stochastic simulation scheme.

by:

$$\hat{P}_{nc}(y_h > \delta \mid \Omega_j) \approx \frac{1}{N_j - \hat{N}_{jc}} \sum_{i=1}^{N - \hat{N}_{jc}} \Phi \left(\frac{\hat{y}_h(y_l(\mathbf{x}_i)) - \delta}{\sigma_{\hat{y}_h}(y_l(\mathbf{x}_i))} \right) \quad (22)$$

$$\hat{P}(\mathcal{D}_c \mid \Omega_j) \approx \frac{1}{N_j} \sum_{i=1}^{N_j} \hat{\mathbf{I}}_c(y_l(\mathbf{x}_i)) = \frac{\hat{N}_{jc}}{N_j} \quad (23)$$

where $\mathbf{x}_i$ is the i th sample, and $\hat{N}_{jc}$ is the estimated number of collapse samples; $\hat{y}_h(y_l(\mathbf{x}_i))$ and $\sigma_{\hat{y}_h}(y_l(\mathbf{x}_i))$ are the predictive mean (Eq. (14)) and standard deviation (Eq. (15)) at $y_l(\mathbf{x}_i)$ from the multi-fidelity regression model; and $\hat{\mathbf{I}}_c(y_l(\mathbf{x}_i))$ is the predicted class label from the multi-fidelity classification model given $y_l(\mathbf{x}_i)$. It follows that the estimator of the conditional failure probability for the j th stratum using the multi-fidelity stochastic simulation scheme is given by Eq. (8) after plugging in Eqs. (22) and (23). After obtaining the conditional failure probabilities for all the strata, the total failure probability can be calculated from Eq. (7).

5. Illustrative Example

To illustrate the proposed multi-fidelity stochastic simulation scheme, it is applied to the two-story two-bay steel frame subject to extreme winds outlined in [42]. The building is assumed to be located in an urban region of Miami, USA, and has a total height of $H = 10$

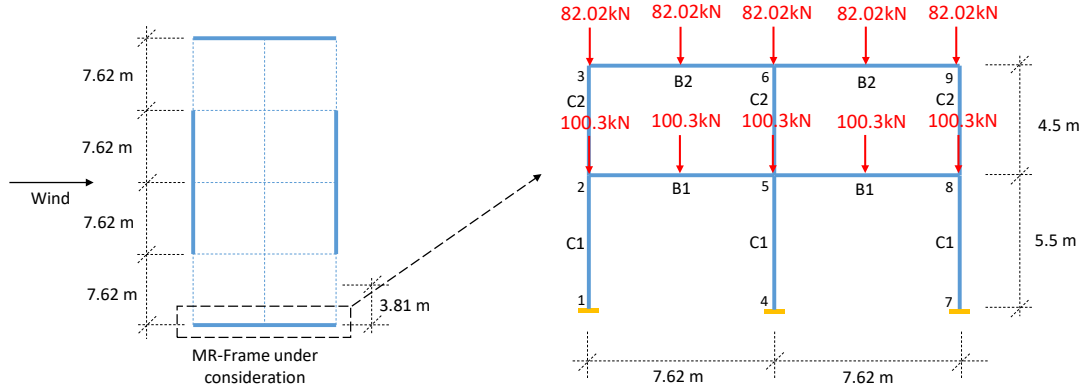


Figure 2: Two-story two-bay steel frame under wind excitation (from [42]).

As illustrated in Fig. 2, the structural system of the building is defined by four moment-resisting (MR) frames (two acting along the short side of the building and two acting in the long direction). Of interest to this work is the performance of the MR frame acting along the short side of the building for which the critical wind direction is perpendicular to the long side of the building. As reported in [42], the MR frame is composed of two unique box sections for columns (C1 and C2) and two unique $W \times 24$ sections for the beams (B1 and B2). The dynamic wind loads are assumed to act at each level of the building. Gravity loads are indicated in Fig. 2. The section sizes were selected to achieve an elastic response under a 700-year return period wind load and satisfy a peak roof displacement limit of $H/300$ under a 25-year return period wind load.

Limit states are considered on the maximum interstory drift ratio and ductility ratio of the members. In particular, the ductility ratio is defined as the maximum strain (during the wind event) over the yield strain of the most critical element in the structure. System collapse is defined as the maximum interstory drift ratio exceeding 5% [43].

5.1. Implementation Details

5.1.1. High-fidelity model

The steel frame shown in Fig. 2 was modeled in OpenSees. Each member was modeled using two nonlinear force-based fiber elements. Five integration points were used for each element and fourteen fibers were used for each section (fiber discretization illustrated in [42]).

Fiber damage resulting from low-cycle fatigue and potential fiber fracture was considered and modeled according to the linear damage accumulation rule and the modified rain flow cycle algorithm [44]. The cyclic behavior of steel was simulated by the Menegotto-Pinto material model [45]. Large displacement effects were considered in the high-fidelity model by adopting a corotational transformation. These settings were implemented to ensure that the high-fidelity model accurately captures the nonlinear behavior, including the collapse, of the structure subjected to extreme winds. For the nonlinear analysis, a Rayleigh damping model was adopted to represent the inherent structural damping [46]. The first two natural frequencies of the structure were approximately $f_1 = 0.51$ Hz and $f_2 = 1.22$ Hz, which ensures a certain degree of dynamic excitation by the wind loads. However, the modal frequencies will be affected by the structural uncertainties described in Sec. 5.1.3, leading to deviations from the above-mentioned values.

5.1.2. Stochastic Wind Model

A local wind climate model, for describing the mean wind speed profile, and a spectral representation model, for describing the fluctuating component of the wind speed, were adopted in defining the wind loads. For the wind speed profile, the widely used power law was adopted, while for the fluctuating component of the wind speeds, the target cross power spectral density matrix proposed by Kaimal was adopted [47] with coherence functions as outlined in [48]. The wind forces were considered to act laterally in the frame's plane at the level of each floor. A quasi-steady model was used to transform the spatio-temporally varying wind speeds to wind load histories acting at the two floors [49, 50]. In particular, the quasi-steady model was calibrated assuming an influence width of 15.24 m for each floor, a quasi-steady pressure coefficient of 1.2, and air density 1.25 kg/m^3 . The stochastic wind loads were simulated with a time step of 0.01 s and total duration 10-minutes. In order to accurately simulate the initial conditions as well as ensure the possibility of directly estimating residual deformation, the first and last two minutes of the loads were linearly ramped up and down.

Table 1: Summary of the basic random variables (CoV = coefficient of variation).

Parameter	Mean	CoV	Distribution	Reference
E	200 Gpa	0.04	Lognormal	[53, 54]
F_y	380 Mpa	0.06	Lognormal	[53, 54]
b	0.001	0.01	Lognormal	[45, 55]
ϵ_0	0.077	0.161	Lognormal	[45]
R_0	20	0.166	Normal	[45]
a_1	0.01	2	Lognormal	[45]
a_3	0.02	0.5	Lognormal	[45]
δ_1/L	0.056%	0.77	Normal	[56, 57]
ζ	0.015	0.4	Lognormal	[58, 59, 60]

5.1.3. Uncertain Model and Load Parameters

The random variables associated with the structural system and their governing distributions are listed in Table 1. The random variables were divided into three groups: (i) seven material model parameters, uniquely defined for each structural section in the frame, including Young’s modulus E , yield strength F_y , strain hardening ratio b , fatigue material parameter ϵ_0 , elastic-to-plastic transition parameter R_0 , and the two hardening parameters a_1 and a_3 [45]; (ii) damping ratio, ζ , of the first two modes, which are considered to be equal; and (iii) the initial camber, uniquely defined for each column at mid-length, described by a random scale factor, δ_1 , and a random sign, (± 1) based on the first buckling mode.

The uncertainties considered in the wind loads consist of the mean hourly wind speed at the building top, $\bar{v}_H$, and the independent and uniformly distributed phase angles used in the spectral representation model [51]. The wind speed $\bar{v}_H$ is considered to follow a Type-I extreme value distribution with parameters calibrated to the Type-I distribution with parameters estimated through calibration to the wind speeds provided in [52] for Miami, Fl. The total number the random variables associated with structural parameters and wind loads is in the order of 10^5 , illustrating the high-dimensionality of the stochastic simulation problem of this case study.

5.1.4. Stratified Sampling

Since wind speed is the dominant variable affecting the structural performance associated with the considered limit states, it is selected as the stratification variable. Considering that the load effect is approximately proportional to the square of the wind speed, the partitioning of the stratified sampling was based on imposing an equal squared difference in wind speed

Table 2: Sample allocation.

WSI	$\bar{v}_H^{\text{lower}}$ [m/s]	$\bar{v}_H^{\text{upper}}$ [m/s]	n_i
1	0.00	21.08	5,000
2	21.08	29.82	5,000
3	29.82	36.52	5,000
4	36.52	42.17	5,000
5	42.17	47.15	5,000
6	47.15	51.65	5,000
7	51.65	55.78	5,000
8	55.78	∞	5,000

[61, 62, 63, 8]. To ensure the collective completeness of the wind speed intervals (WSIs), the lower bound defining the first WSI is assumed to be zero, while the last WSI from above is assumed to be unbounded. The lower limit of the last WSI was chosen to correspond to an annual exceedance probability of 7×10^{-7} , a specified value in [52] that corresponds to collapse for a risk Category II structure. A uniform sample allocation was adopted in this application with 5000 samples in each stratum. The final sample allocation is shown in Table 2 along with the lower and upper bound wind speeds defining the partition.

5.1.5. Low-fidelity Models

The low-fidelity model adopted an elastic-perfectly plastic material model to simulate the steel therefore significantly simplifying its yielding behavior. In addition, fiber damage due to low-cycle fatigue and potential fiber fracture, and large displacement effects, were ignored. Therefore, uncertain input parameters of the low-fidelity model only include the modulus of Young E , the yield strength F_y , and the damping ratio ζ . To further reduce the computational cost of the low-fidelity model, a larger time step in solving the dynamic equations was adopted. Due to these simplifications, this elastic-perfectly plastic low-fidelity model, denoted *LF2*, is approximately 10 times faster to evaluate than the corresponding high-fidelity model. For comparison purposes, an even more efficient low-fidelity model, which is developed based on modal integration, was also developed. This low-fidelity model ignores the material and geometric non-linearity of the structure and thus only provides a rough approximation of the structural behavior. This elastic model, denoted *LF1*, is approximately 300 times faster to evaluate than the corresponding high-fidelity model.

5.1.6. Calibration of the Multi-fidelity Scheme

To train the multi-fidelity models, only a small subset of the samples from each stratum were selected, for which the high-fidelity model was evaluated. The active learning strategy was used to determine the number and the location of the samples needed for the model training. To set up the initial training set, the “uniform” sampling plan (described in Sec. 4.2.3) was adopted. In particular, $n_{\text{bin}} = 5$ were used and two samples were chosen from each bin to form the initial training set. To monitor the convergence of the multi-fidelity regression model, 25 evenly distributed values were initially picked from the low-fidelity response range in each WSI. It should be noted that the low-fidelity response ranges were defined by the minimum and maximum predicted non-collapse low-fidelity responses from the trained multi-fidelity classification model. Furthermore, the predictive means and standard deviations over the selected low-fidelity responses were averaged to obtain a measure of the overall convergence of the multi-fidelity regression model in each WSI. In this study, the maximum interstory drift ratio was identified as the key structural performance measure and thus was used to determine the training data.

5.2. Results and Discussion

5.2.1. Performance of Multi-fidelity Models

This section reports the results from the multi-fidelity stochastic simulation using the elastic-perfectly-plastic low-fidelity model (*LF2*). To differentiate the multi-fidelity regression models constructed based on the heteroscedastic and homoscedastic Gaussian process model, they are referred to as the HGP-based multi-fidelity model and the GP-based multi-fidelity model. Fig. 3 shows the variation of the averaged predictive means and standard deviations for the maximum interstory drift ratio over the active learning iterations, where the predictions are from the HGP-based multi-fidelity model. From Fig. 3, it can observe that the average predictive means and standard deviations show variation at the beginning of the active learning indicating how the multi-fidelity regression models require additional samples before becoming stable. This is especially true for higher WSIs, e.g., the 6th and 7th WSIs, since $p_j(y_h|y_l)$ is more complex for higher winds due to the significant non-linearity in the system. Based on the variation of the trained multi-fidelity regression models and the

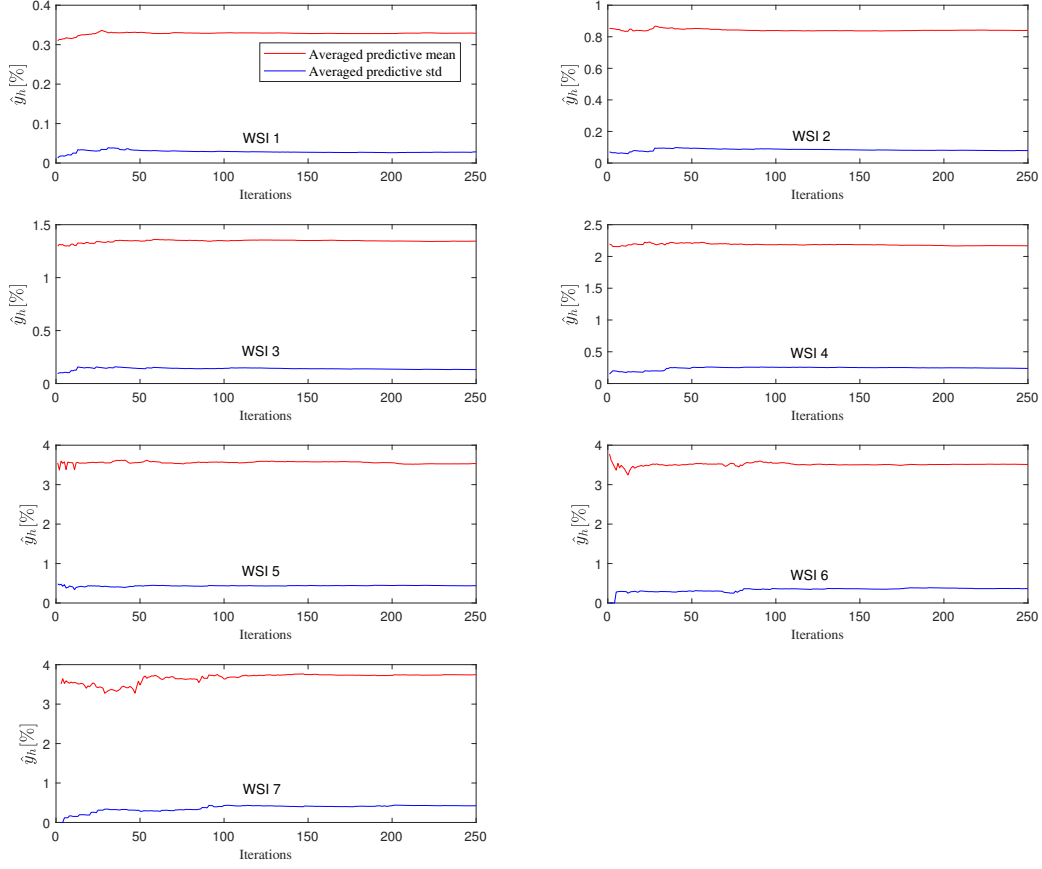


Figure 3: Variation of the HGP-based multi-fidelity models with the number of active learning iterations (std: standard deviation).

convergence criterion defined in Eq. (19), the active learning process stopped at the 101th, 124nd, 93th, 99th, 89nd, 169th, and 220th iteration, respectively, for the 1st to the 7th WSIs.

Regarding the 8th WSI, the multi-fidelity classification model predicted collapse for all samples, and thus no regression model was trained. In summary, the multi-fidelity regression models converge quickly, which ensures the computational efficiency of the multi-fidelity stochastic simulation scheme. For the GP-based multi-fidelity models, the variation of the averaged predictive means and standard deviations for the maximum interstory drift ratio with the iterations showed similar convergence behavior to the HGP-based multi-fidelity model.

Similar to the multi-fidelity regression models, the strata-wise multi-fidelity classification models for predicting the system collapse improve iteratively until the stopping criterion is

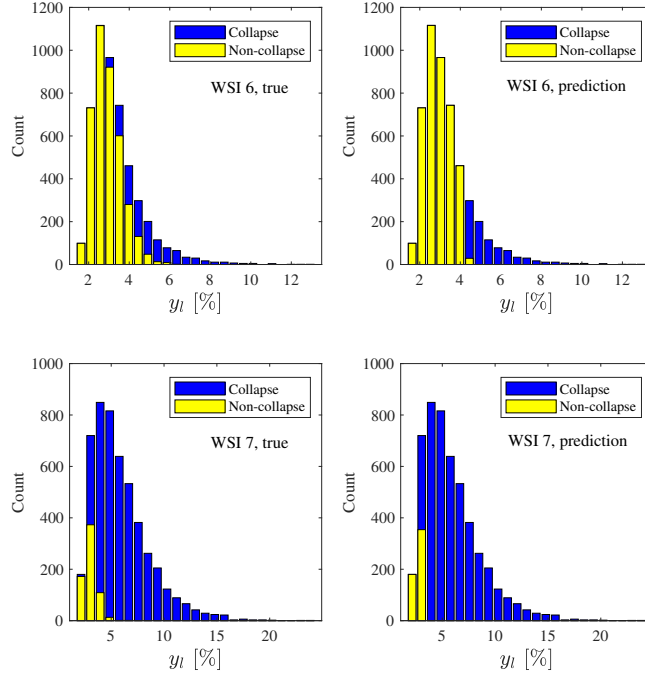


Figure 4: True versus predicted collapse/non-collapse given any low-fidelity maximum interstory drift ratios.

met. The number of iterations for the multi-fidelity classification models determined by the stopping criterion is 21 for the first 4 WSIs, and 32, 68, 84, 68 for the 5th to 8th WSIs, respectively. In order to show the performance of the multi-fidelity classification model, Fig. 4 compares the true and the predicted collapse/non-collapse samples for all low-fidelity maximum interstory drift ratio samples of WSIs 6 and 7. As can be observed from Fig. 4, the predicted collapse/non-collapse matches well with the true collapse/non-collapse, verifying the effectiveness of the established multi-fidelity classification models. Results for the other WSIs are not shown as for these WSI all samples do not collapse, or collapse, therefore facilitating the classification.

Fig. 5 presents the predictive means and the associated uncertainties for any given low-fidelity maximum interstory drift ratio from the strata-wise multi-fidelity regression models together with the training data. It can be seen that the maximum interstory drift ratios from the low-fidelity model correlate well with those from the high-fidelity model when the WSI is low. This is because the frame is still elastic at low wind speeds, and the low-fidelity model should offer similar solutions to the high-fidelity model. The variability of the high-fidelity

response, given a specific low-fidelity response, is mainly attributed to the difference in time step used to solve the dynamic equations in two numerical models. In contrast, the maximum interstory drift ratios are less correlated as wind speeds increase. This is to be expected since high wind speeds generally result in significant non-linear behavior of the frame which the low-fidelity model cannot accurately describe. Overall, the HGP-based multi-fidelity models can fairly accurately capture the mean trend as well as the variability of the high-fidelity response given any low-fidelity response. Similarly, the predictive distributions given by the GP-based multi-fidelity models are shown in Fig. 6. By comparing the HGP-based and GP-based multi-fidelity models, it can be seen that the HGP-based multi-fidelity models can capture the varying uncertainty of y_h given y_l for the first 5 WSIs, while the GP-based multi-fidelity model predicts the uncertainty to be constant as y_l changes. Consequently, the GP-based models can either overestimate or underestimate the uncertainty of y_h given a y_l . For the 6th and 7th WSIs, the HGP-based and GP-based multi-fidelity models provide similar predictive mean and standard deviation, as it is observed that the uncertainty of y_h at any y_l is almost constant.

To further demonstrate the predictive performance of the multi-fidelity regression models, the true conditional distribution $p_j(y_h|y_l)$ was compared with the predictive distribution from the multi-fidelity models. The true conditional distribution $p_j(y_h|y_l)$ is unknown but can be estimated using the non-collapse samples from the 5000 samples generated in each WSI. Fig. 7 and Fig. 8 show the true and predicted $p_j(y_h|y_l)$ for each WSI, where y_l was chosen as the median of the low-fidelity maximum interstory drift ratios of the non-collapse samples in each WSI. The figures show that the predictive mean from the multi-fidelity models agrees well with the mean of the true conditional distribution $p(y_h|y_l)$. Although it can be observed that the predictive variance is overestimated at low WSIs while it is underestimated at high WSIs, the predictive variance can be estimated with relatively high accuracy. In comparison, the GP-based multi-fidelity models overestimate the uncertainty of y_h at a y_l more than the HGP-based multi-fidelity models in the first 4 WSIs, while the difference between the two model types becomes smaller as the wind speed becomes large (e.g., in WSI 6 and 7). Overall,

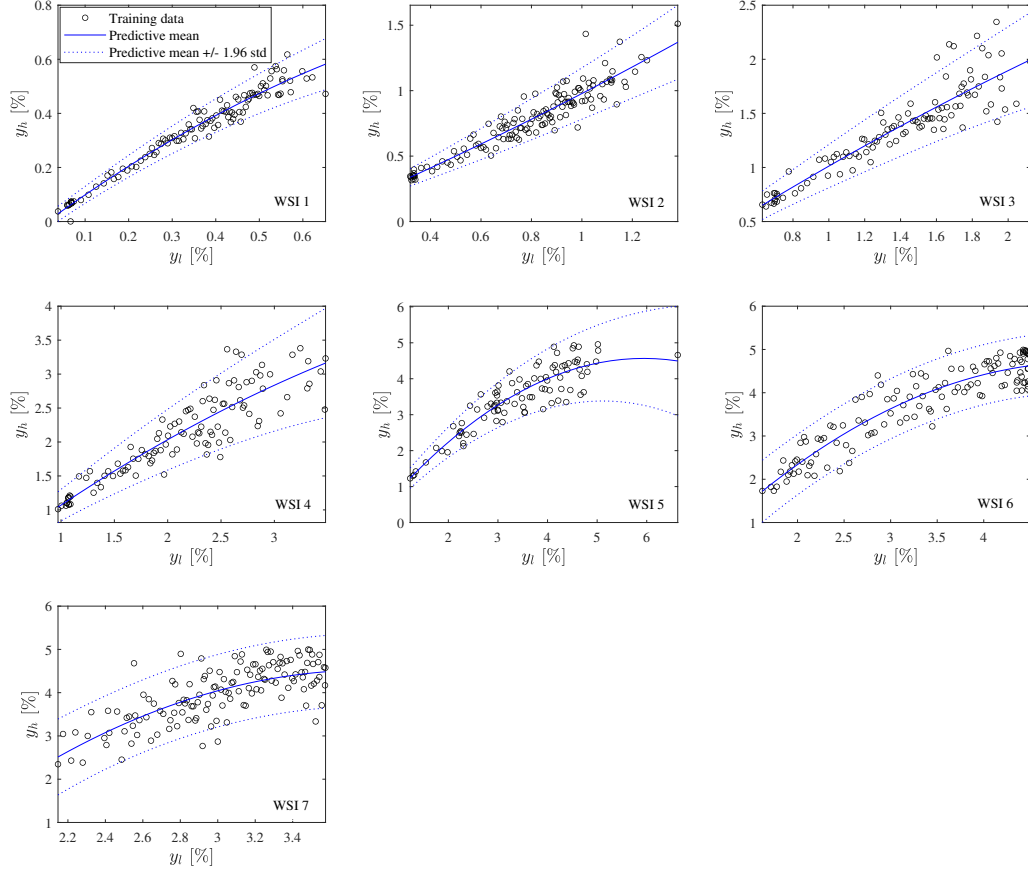


Figure 5: Trained HGP-based multi-fidelity regression models for the maximum interstory drift ratio (std: standard deviation).

the results indicate that with only one to two hundred evaluations of the high-fidelity model in each WSI, the predicted conditional distribution $p_j(y_h|y_l)$ is reasonably accurate.

5.2.2. Estimation of Failure Probability

In this section, the failure probabilities, $\hat{P}(y_h > \delta)$, were estimated using Eq. (7) and Eqs. (22)-(23), where different response limits, $\delta = \{2\%, 3\%, 4\%, 5\%\}$, were considered for the maximum interstory drift ratio. This section includes the results from the multi-fidelity stochastic simulation using both low-fidelity models. Table 3 summarizes the corresponding predicted failure probabilities, where: *LF1-MF-GP* = multi-fidelity model with elastic low-fidelity model and homoscedastic GPR model; *LF1-MF-HGP* = multi-fidelity model

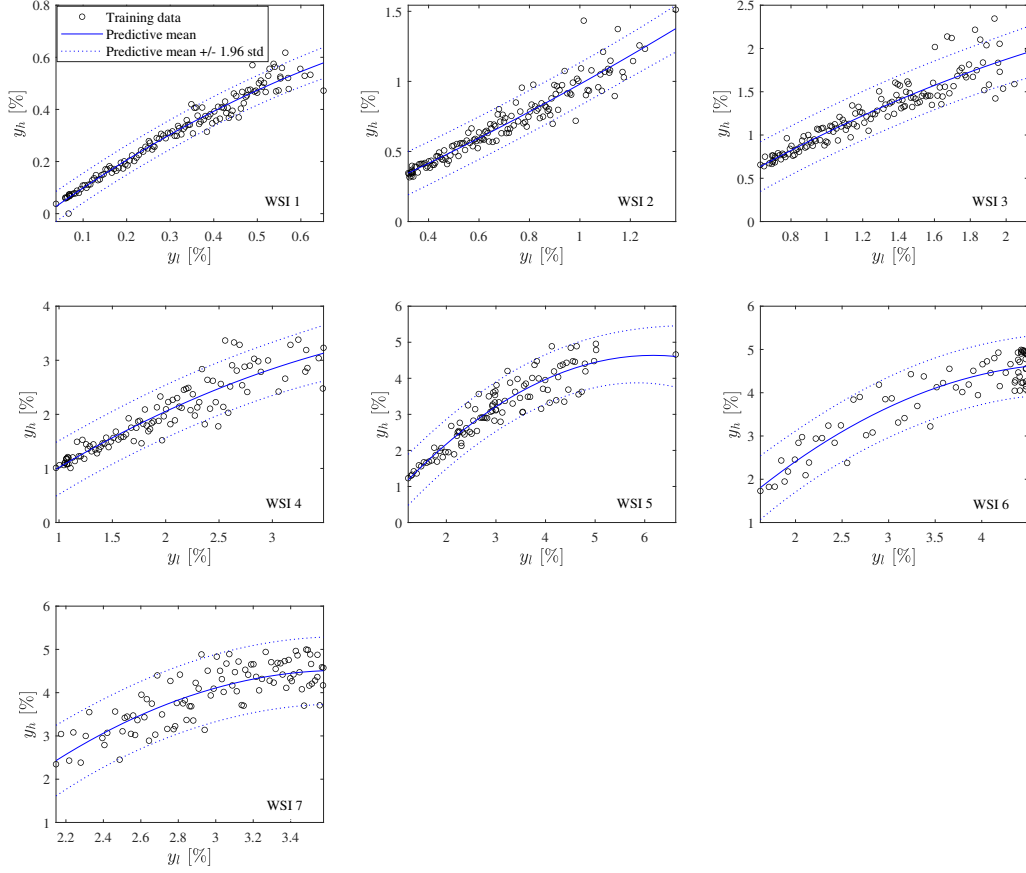


Figure 6: Trained GP-based multi-fidelity regression models for the maximum interstory drift ratio (std: standard deviation).

with elastic low-fidelity model and heteroscedastic GPR model; $LF2-MF-GP$ = multi-fidelity model with elastic-perfectly-plastic low-fidelity model and homoscedastic GPR model; $LF2-MF-HGP$ = multi-fidelity model with elastic-perfectly-plastic low-fidelity model and heteroscedastic GPR model. For comparison purposes, the failure probabilities were also estimated from single-fidelity stochastic simulation, including using $LF1$ = elastic low-fidelity model; $LF2$ = elastic-perfectly-plastic low-fidelity model; HF = high-fidelity model using stratified sampling. The single-fidelity stochastic simulation is based on stratified sampling, where 5000 low-fidelity or high-fidelity samples were generated in each stratum. It is worth mentioning that the estimated failure probabilities from “ HF ” were used as reference results (i.e., ground truth). The use of stratified sampling, instead of a standard Monte Carlo (MC) scheme, for generating the reference solution was simply due to how it enables significant

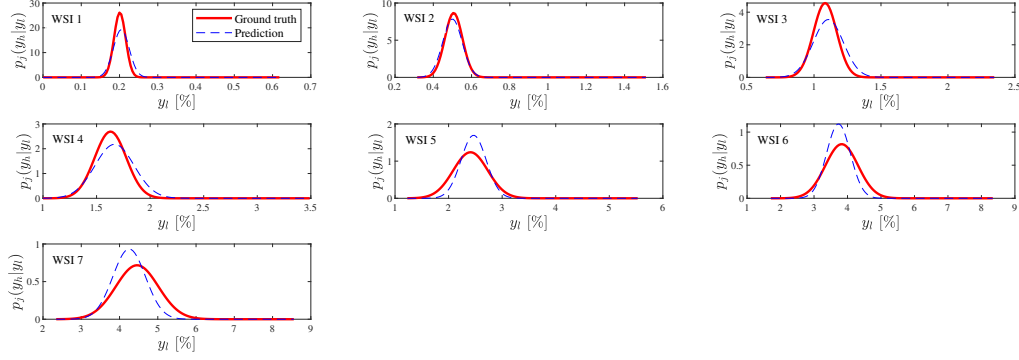


Figure 7: True versus predicted conditional distribution $p_j(y_h|y_l)$ from the HGP-based multi-fidelity model.

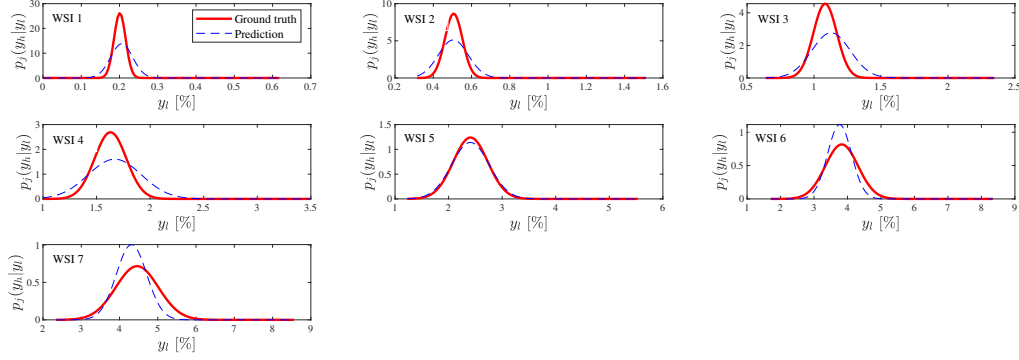


Figure 8: True versus predicted conditional distribution $p_j(y_h|y_l)$ from the GP-based multi-fidelity model.

variance reduction while maintaining the desirable statistical properties of the MC estimator, i.e., unbiased, consistent, and normally distributed estimates of the failure probability, as demonstrated in [8]. Therefore, the use of stratified sampling in generating a reference solution can be considered equivalent to the use of an MC scheme with, however, a significant reduction in the number of samples necessary to estimate the small failure probabilities of interest to this work. To reinforce this, subset simulation was also used to estimate failure probabilities associated with the maximum interstory drift ratio. To ensure similarly small coefficients of variation between the reference solutions estimated from subset simulation and stratified sampling, a total of 33,000 high-fidelity model runs were considered in implementing subset simulation. Specifically, 9 subsets were considered with a level probability of 0.3 and 5,000 samples in each subset. The results from subset simulation are also presented in Table 3 and labeled *HF (SuS)*. As would be expected, the failure probabilities estimated from subset

simulation are closely aligned with those estimated from stratified sampling. In addition, to highlight the computational advantages of the proposed approach, Table 3 also includes the total number of high-fidelity model (# of HF runs) and low-fidelity (# of LF runs) model runs used in the evaluations for each case. To further compare the results, the percentage error: $err = \frac{\hat{P}(y_h > \delta | \Omega_j) - P(y_h > \delta | \Omega_j)}{P(y_h > \delta | \Omega_j)} \times 100\%$, was also defined in terms of the stratified sampling target solution and is shown in Fig. 9.

From Table 3 and Fig. 9, several interesting findings can be summarized. Firstly, although highly efficient to evaluate, *LF1* provides an extremely biased estimation of the failure probabilities if it is directly used in replace of the high-fidelity model. However, dramatic improvements can be seen using either the *LF1-MF-GP* or *LF1-MF-HGP* models. Secondly, *LF2* offers a better estimation of the failure probabilities as it can describe the structural behavior more realistically than *LF1*. However, this increased accuracy is achieved by sacrificing computational efficiency. Indeed, *LF2* is approximately 30 times slower than *LF1*. Similarly to above, using either *LF2-MF-GP* or *LF2-MF-HGP*, the estimated failure probabilities see a considerable improvement in accuracy. Thirdly, by comparing results from *LF1-MF-GP/HGP* and *LF2-MF-GP/HGP*, it can be observed that the multi-fidelity models based on *LF2* provide a better estimation of the failure probabilities than the models based on *LF1* in most cases. This is expected since *LF2* provides more accurate information about the high-fidelity model than *LF1*. Lastly, the computational advantages of the proposed multi-fidelity framework over existing state-of-the-art approaches, e.g., subset simulation (*HF (SuS)*) or stratified sampling (*HF*), is clearly evident from the total high-fidelity model runs necessary for estimating the failure probabilities using either *HF* or *HF (SuS)* as compared to those necessary using any of the four multi-fidelity model setups. Indeed, in all cases, over 30 times as many high-fidelity model runs are necessary when implementing *HF* or *HF (SuS)* for obtaining failure probability estimates of similar accuracy.

Finally, the difference between the GP-based and HGP-based multi-fidelity models is negligible in general, although the HGP-based multi-fidelity models provide a slightly better estimation of the failure probabilities when the threshold is small. The reason is that the GP-based multi-fidelity models add additional uncertainty to the conditional distribution $p_j(y_h | y_l)$ given small y_l values in WSIs associated with low wind speeds, while the GP and

Table 3: Estimated failure probabilities for the maximum interstory drift ratio.

δ	# of LF runs	# of HF runs	2%	3%	4%	5%
<i>LF1</i>	40,000	0	4.30×10^{-5}	1.86×10^{-6}	1.59×10^{-7}	1.93×10^{-8}
<i>LF1-MF-GP</i>	40,000	918	1.61×10^{-4}	2.81×10^{-5}	8.52×10^{-6}	4.41×10^{-6}
<i>LF1-MF-HGP</i>	40,000	957	1.57×10^{-4}	2.76×10^{-5}	8.53×10^{-6}	4.43×10^{-6}
<i>LF2</i>	40,000	0	1.18×10^{-4}	1.88×10^{-5}	6.55×10^{-6}	3.42×10^{-6}
<i>LF2-MF-GP</i>	40,000	938	1.69×10^{-4}	2.79×10^{-5}	9.63×10^{-6}	4.85×10^{-6}
<i>LF2-MF-HGP</i>	40,000	963	1.58×10^{-4}	2.79×10^{-5}	9.62×10^{-6}	4.89×10^{-6}
<i>HF</i>	0	40,000	1.47×10^{-4}	2.71×10^{-5}	1.01×10^{-5}	5.17×10^{-6}
<i>HF (SuS)</i>	0	33,000	1.79×10^{-4}	3.23×10^{-5}	1.12×10^{-5}	6.13×10^{-6}

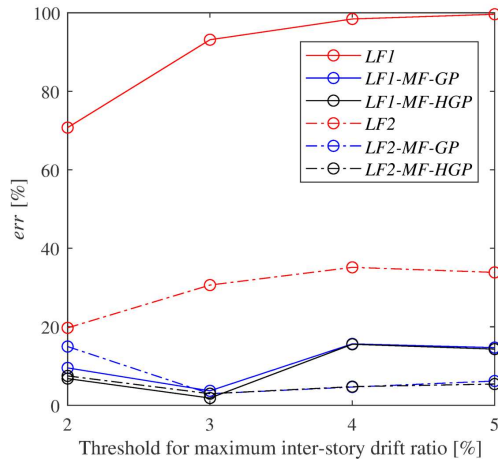


Figure 9: Percentage error of the estimated failure probabilities for maximum interstory drift ratio.

Table 4: Estimated failure probabilities for the ductility ratio.

δ	2%	3%	4%	5%
<i>LF1</i>	2.57×10^{-6}	8.10×10^{-8}	1.93×10^{-8}	1.93×10^{-8}
<i>LF1-MF-GP</i>	1.01×10^{-4}	3.88×10^{-5}	1.85×10^{-5}	9.92×10^{-6}
<i>LF1-MF-HGP</i>	1.00×10^{-4}	3.80×10^{-5}	1.83×10^{-5}	9.98×10^{-6}
<i>LF2</i>	2.21×10^{-5}	5.72×10^{-6}	3.43×10^{-6}	3.42×10^{-6}
<i>LF2-MF-GP</i>	1.19×10^{-4}	4.06×10^{-5}	2.11×10^{-5}	1.19×10^{-5}
<i>LF2-MF-HGP</i>	1.02×10^{-4}	3.85×10^{-5}	1.99×10^{-5}	1.15×10^{-5}
<i>HF</i>	9.65×10^{-5}	3.68×10^{-5}	2.06×10^{-5}	1.25×10^{-5}

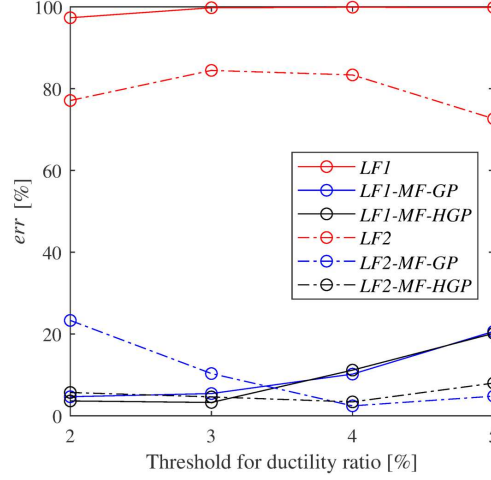


Figure 10: Percentage error of the estimated of failure probabilities for the maximum ductility ratio.

HGP predicted noise/uncertainty convergences to similar values in WSIs with higher wind speeds.

An important property of the proposed multi-fidelity framework is that the samples already evaluated using the high-fidelity model for one limit state can be reused for other limit states. This is illustrated here for the maximum element-level ductility ratio. Table 4 lists the predicted failure probabilities using the multi-fidelity models outlined above while Fig. 10 shows the percentage errors. Similar conclusions to the limit state of the maximum interstory drift ratio can be drawn from these results. Overall, the proposed multi-fidelity stochastic simulation scheme is able to estimate small failure probabilities (e.g., in the order of 10^{-6}) with high accuracy while using only a small number of high-fidelity evaluations.

6. Conclusion

This paper introduced a multi-fidelity stochastic simulation scheme for the estimation of small failure probabilities within the setting of stratified sampling. The core idea is to leverage an inaccurate but inexpensive low-fidelity version of a system model to provide information about the high-fidelity system model and therefore speed up failure probability estimation. For collapse estimation, a Gaussian process classification model, calibrated by a small number of high-fidelity runs, was constructed to predict collapse or non-collapse given any low-fidelity response. For non-collapse samples, the underlying probabilistic relationship between the low-fidelity and high-fidelity response was established through a Gaussian process regression model requiring again a small number of high-fidelity runs for calibration. These multi-fidelity models were established in each stratum, therefore, allowing low-fidelity model runs to provide estimations of failure probabilities with accuracy akin to considering a high-fidelity model. To illustrate the effectiveness of the proposed scheme, it was applied to estimate the failure probabilities of a 2D steel frame subject to extreme winds. The results showed that the multi-fidelity regression models can capture the probabilistic relationship between the low-fidelity and high-fidelity response with only a small number of high-fidelity model evaluations. The advantage of the proposed scheme is that even with a crude low-fidelity model, the corresponding multi-fidelity model can yield a significant improvement in the estimation of the failure probabilities. Moreover, the proposed multi-fidelity scheme is able to simultaneously estimate failure probabilities associated with multiple limit states. While the presented two-story, two-bay steel frame example effectively demonstrated the capabilities of the proposed methodology, the need for further validation in the context of high-rise structures under wind loads is recognized. This validation forms part of ongoing research projects and is included as a future direction to enhance the robustness and applicability of the proposed multi-fidelity stochastic scheme.

Acknowledgements

This research effort was supported by the National Science Foundation (NSF) under Grant No. CMMI-1750339. This support is gratefully acknowledged.

Appendix: Bayesian inference of Gaussian process classification (GPC)

The Bayesian inference of the GPC model again requires a training set $\{\mathbf{Y}_l, \mathcal{I}_c\}$, where the model input is $\mathbf{Y}_l = \{y_l(\mathbf{x}_t); t = 1, \dots, n_t\}$ and the model output (i.e., label) is $\mathcal{I}_c = \{\mathbf{I}_c(\mathbf{x}_t); t = 1, \dots, n_t\}$. The probabilistic prediction over the indicator function, $\mathbf{I}_c$, given any new low-fidelity output y_{l*} can be obtained as:

$$p(\mathbf{I}_{c*} = 1 | \mathbf{Y}_l, \mathcal{I}_c, y_{l*}) = \int p(\mathbf{I}_{c*} = 1 | f_*) p(f_* | \mathbf{Y}_l, \mathcal{I}_c, y_{l*}) df_* \quad (24)$$

where $p(\mathbf{I}_{c*} = 1 | f_*)$ is the prior assumed over the target function $\mathbf{I}_c$, given by $\Phi(f_*)$, and $p(f_* | \mathbf{Y}_l, \mathcal{I}_c, y_{l*})$ is the predictive distribution over f given y_{l*} , expressed as:

$$p(f_* | \mathbf{Y}_l, \mathcal{I}_c, y_{l*}) = \int p(f_* | \mathbf{Y}_l, \mathbf{f}, y_{l*}) p(\mathbf{f} | \mathbf{Y}_l, \mathcal{I}_c) d\mathbf{f} \quad (25)$$

where $p(f_* | \mathbf{Y}_l, \mathbf{f}, y_{l*})$ takes the form of a GPR predictive distribution, and $p(\mathbf{f} | \mathbf{Y}_l, \mathcal{I}_c) = p(\mathbf{f} | \mathbf{Y}_l) p(\mathcal{I}_c | \mathbf{f}) / p(\mathcal{I}_c | \mathbf{Y}_l)$ is the posterior distribution over $\mathbf{f}$.

Unlike the GPR, the Bayesian inference of the GPC is analytically intractable [64] since a non-Gaussian likelihood $p(\mathcal{I}_c | \mathbf{f})$ (i.e., probit function in this work) is assumed considering that the target function, $\mathbf{I}_c$, produces discrete class labels. Therefore, approximation methods are often used to estimate, $p(\mathbf{f} | \mathbf{Y}_l, \mathcal{I}_c)$, with commonly used ones including the Laplace approximation [65] and expectation propagation [66]. More details about the derivation of the integral in Eq. (25) can be found in [35]. Once $p(\mathbf{f} | \mathbf{Y}_l, \mathcal{I}_c)$ is approximated, the predictive distribution over $\mathbf{f}$ is computed by Eq. (25), with which the GPC probabilistic prediction can be obtained by solving Eq. (24). Depending on the link function, Eq. (24) can be calculated analytically (e.g., when the link function is a probit function) or not (e.g., when the link function is a logit function). If a probit function is used and the Laplace approximation is adopted, the GPC probabilistic prediction is given by:

$$p(\mathbf{I}_{c*} = 1 | \mathbf{Y}_l, \mathcal{I}_c, y_{l*}) = \Phi \left(\frac{\hat{\mu}(y_{l*})}{\sqrt{1 + \sigma_{\hat{\mu}}^2(y_{l*})}} \right) \quad (26)$$

with

$$\hat{\mu}(y_{l*}) = \mathbf{k}_*^T \mathbf{K}^{-1} \mathbf{m} \quad (27)$$

$$\sigma_{\hat{\mu}}^2(y_{l*}) = k(y_{l*}, y_{l*}) - \mathbf{k}_*^T (\mathbf{K} + \mathbf{A}^{-1})^{-1} \mathbf{k}_* \quad (28)$$

where the parameters $\mathbf{m}$ and $\mathbf{A}$ can be found in [35]. In terms of the calibration of the hyperparameters in the GPC model, i.e., length-scale θ and variance σ^2 , MLE is also used, where the likelihood function is given by:

$$p(\mathcal{I}_c | \mathbf{f}) = \prod p(\mathbf{I}_c(y_l(\mathbf{x}_i)) | f(y_l(\mathbf{x}_i))) \quad (29)$$

References

- [1] N. Lelièvre, P. Beaurepaire, C. Mattrand, N. Gayton, AK-MCSi: A kriging-based method to deal with small failure probabilities and time-consuming models, *Structural Safety* 73 (2018) 1–11.
- [2] S. Arunachalam, S. M. J. Spence, Reliability-based collapse assessment of wind-excited steel structures within performance-based wind engineering, *Journal of Structural Engineering* 148 (9) (2022) 04022132.
- [3] R. E. Melchers, Importance sampling in structural systems, *Structural Safety* 6 (1) (1989) 3–10.
- [4] P.-S. Koutsourelakis, H. J. Pradlwarter, G. I. Schueller, Reliability of structures in high dimensions, part I: algorithms and applications, *Probabilistic Engineering Mechanics* 19 (4) (2004) 409–417.
- [5] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, *Probabilistic engineering mechanics* 16 (4) (2001) 263–277.
- [6] M. D. Shields, J. Zhang, The generalization of Latin hypercube sampling, *Reliability Engineering & System Safety* 148 (2016) 96–108.
- [7] M. D. Shields, V. Sundar, Targeted random sampling: a new approach for efficient reliability estimation for complex systems, *International Journal of Reliability and safety* 9 (2-3) (2015) 174–190.
- [8] S. Arunachalam, S. M. J. Spence, An efficient stratified sampling scheme for the simultaneous estimation of small failure probabilities in wind engineering applications, *Structural Safety* 101 (2023) 102310.

- [9] S.-K. Au, J. Beck, Important sampling in high dimensions, *Structural safety* 25 (2) (2003) 139–163.
- [10] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, *Siam Review* 60 (3) (2018) 550–591.
- [11] J. Li, D. Xiu, Evaluation of failure probability via surrogate models, *Journal of Computational Physics* 229 (23) (2010) 8966–8980.
- [12] J. Biehler, M. Mäck, J. Nitzler, M. Hanss, P.-S. Koutsourelakis, W. A. Wall, Multifidelity approaches for uncertainty quantification, *GAMM-Mitteilungen* 42 (2) (2019) e201900008.
- [13] L. W. Ng, K. E. Willcox, Multifidelity approaches for optimization under uncertainty, *International Journal for numerical methods in Engineering* 100 (10) (2014) 746–772.
- [14] B. Peherstorfer, T. Cui, Y. Marzouk, K. Willcox, Multifidelity importance sampling, *Computer Methods in Applied Mechanics and Engineering* 300 (2016) 490–509.
- [15] B. L. Nelson, On control variate estimators, *Computers & Operations Research* 14 (3) (1987) 219–225.
- [16] S. Heinrich, Multilevel Monte Carlo methods, in: *International Conference on Large-Scale Scientific Computing*, Springer, 2001, pp. 58–67.
- [17] M. B. Giles, Multilevel Monte Carlo path simulation, *Operations research* 56 (3) (2008) 607–617.
- [18] K. A. Cliffe, M. B. Giles, R. Scheichl, A. L. Teckentrup, Multilevel Monte Carlo methods and applications to elliptic pdes with random coefficients, *Computing and Visualization in Science* 14 (1) (2011) 3–15.
- [19] B. Peherstorfer, K. Willcox, M. Gunzburger, Optimal model management for multifidelity Monte Carlo estimation, *SIAM Journal on Scientific Computing* 38 (5) (2016) A3163–A3194.
- [20] B. Kramer, A. N. Marques, B. Peherstorfer, U. Villa, K. Willcox, Multifidelity probability estimation via fusion of estimators, *Journal of Computational Physics* 392 (2019) 385–402.
- [21] D. Patsialis, A. Taflanidis, Multi-fidelity Monte Carlo for seismic risk assessment applications, *Structural Safety* 93 (2021) 102129.

- [22] J. Nitzler, J. Biehler, N. Fehn, P.-S. Koutsourelakis, W. A. Wall, A generalized probabilistic learning approach for multi-fidelity uncertainty propagation in complex physical simulations, arXiv preprint arXiv:2001.02892 (2020).
- [23] M. Li, G. Jia, Multifidelity gaussian process model integrating low-and high-fidelity data considering censoring, *Journal of Structural Engineering* 146 (3) (2020) 04019215.
- [24] M. C. Kennedy, A. O’Hagan, Predicting the output from a complex computer code when fast approximations are available, *Biometrika* 87 (1) (2000) 1–13.
- [25] A. Forrester, A. Sóbester, A. Keane, *Engineering design via surrogate modelling: a practical guide*, Wiley, 2008.
- [26] L. Le Gratiet, Bayesian analysis of hierarchical multifidelity codes, *SIAM/ASA Journal on Uncertainty Quantification* 1 (1) (2013) 244–269.
- [27] A. Renganathan, V. Rao, I. Navon, Multifidelity gaussian processes for failure boundary and probability estimation, in: *AIAA SCITECH 2022 Forum*, 2022, p. 0390.
- [28] J. Yi, F. Wu, Q. Zhou, Y. Cheng, H. Ling, J. Liu, An active-learning method based on multi-fidelity kriging model for structural reliability analysis, *Structural and Multidisciplinary Optimization* 63 (1) (2021) 173–195.
- [29] M. Verleysen, D. François, The curse of dimensionality in data mining and time series prediction, in: *International work-conference on artificial neural networks*, Springer, 2005, pp. 758–770.
- [30] P.-S. Koutsourelakis, Accurate uncertainty quantification using inaccurate computational models, *SIAM Journal on Scientific Computing* 31 (5) (2009) 3274–3300.
- [31] J. Biehler, M. W. Gee, W. A. Wall, Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a bayesian multi-fidelity scheme, *Biomechanics and modeling in mechanobiology* 14 (3) (2015) 489–513.
- [32] A. Hebbal, L. Brevault, M. Balesdent, E.-G. Talbi, N. Melab, Multi-fidelity modeling with different input domain definitions using deep gaussian processes, *Structural and Multidisciplinary Optimization* 63 (5) (2021) 2267–2288.
- [33] S. Arunachalam, S. M. J. Spence, A stochastic simulation scheme for the estimation of small failure probabilities in wind engineering applications, *31st European Safety and Reliability Conference*.

- [34] S. Arumachalam, S. M. J. Spence, Generalized stratified sampling for efficient reliability assessment of structures against natural hazards, *Journal of Engineering Mechanics* 149 (7) (2023) 04023042.
- [35] C. E. Rasmussen, Gaussian processes in machine learning, in: *Summer school on machine learning*, Springer, 2003, pp. 63–71.
- [36] A. P. Kyprioti, A. A. Taflanidis, Kriging metamodeling for seismic response distribution estimation, *Earthquake Engineering & Structural Dynamics* 50 (13) (2021) 3550–3576.
- [37] X. Li, Y. Guo, Adaptive active learning for image classification, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 859–866.
- [38] M. Moustapha, S. Marelli, B. Sudret, Active learning for structural reliability: Survey, general framework and benchmark, *Structural Safety* 96 (2022) 102174.
- [39] Z. Meng, Z. Zhang, G. Li, D. Zhang, An active weight learning method for efficient reliability assessment with small failure probability, *Structural and Multidisciplinary Optimization* 61 (3) (2020) 1157–1170.
- [40] A. Kapoor, K. Grauman, R. Urtasun, T. Darrell, Active learning with gaussian processes for object categorization, in: *2007 IEEE 11th international conference on computer vision*, IEEE, 2007, pp. 1–8.
- [41] F. Rodrigues, F. Pereira, B. Ribeiro, Gaussian process classification and active learning with multiple annotators, in: *International conference on machine learning*, PMLR, 2014, pp. 433–441.
- [42] W.-C. Chuang, S. M. J. Spence, Probabilistic performance assessment of inelastic wind excited structures within the setting of distributed plasticity, *Structural Safety* 84 (2020) 101923.
- [43] FEMA 356 prestandard and commentary for the seismic rehabilitation of buildings, ASCE for the Federal Emergency Management Agency (November 2000).
- [44] P. Uriz, Towards earthquake resistant design of concentrically braced steel structures, University of California, Berkeley, 2005.
- [45] E. Karamanci, D. G. Lignos, Computational approach for collapse assessment of concentrically braced frames in seismic regions, *Journal of Structural Engineering* 140 (8) (2014) A4014019.

- [46] F. A. Charney, Unintended consequences of modeling damping in structures, *Journal of structural engineering* 134 (4) (2008) 581–592.
- [47] J. C. Kaimal, J. Wyngaard, Y. Izumi, O. Coté, Spectral characteristics of surface-layer turbulence, *Quarterly Journal of the Royal Meteorological Society* 98 (417) (1972) 563–589.
- [48] A. G. Davenport, The dependence of wind load upon meteorological parameters, in: *Proceedings International Research Seminar “Wind effects on buildings and structures”*, University of Toronto Press, 1967, pp. 19–82.
- [49] P. Tabbuso, S. M. J. Spence, L. Palizzolo, A. Pirrotta, A. Kareem, An efficient framework for the elasto-plastic reliability assessment of uncertain wind excited systems, *Structural Safety* 58 (2016) 69–78.
- [50] W.-C. Chuang, S. M. J. Spence, A performance-based design framework for the integrated collapse and non-collapse assessment of wind excited buildings, *Engineering Structures* 150 (2017) 746–758.
- [51] X. Chen, A. Kareem, Proper orthogonal decomposition-based modeling, analysis, and simulation of dynamic wind load effects on structures, *Journal of Engineering Mechanics* 131 (4) (2005) 325–339.
- [52] ASCE 7-22, Minimum design loads and associated criteria for buildings and other structures, American Society of Civil Engineers (ASCE), Reston, VA (2022).
- [53] F. M. Bartlett, R. J. Dexter, M. D. Graeser, J. J. Jelinek, B. J. Schmidt, T. V. Galambos, Updating standard shape material properties database for design and reliability, *Engineering Journal-American Institute of Steel Construction Inc* 40 (1) (2003) 2–14.
- [54] H. Zhang, B. R. Ellingwood, K. J. Rasmussen, System reliabilities in steel structural frame design by inelastic analysis, *Engineering structures* 81 (2014) 341–348.
- [55] T. V. Galambos, M. K. Ravindra, Properties of steel for use in lrfd, *Journal of the Structural Division* 104 (9) (1978) 1459–1468.
- [56] H. Zhang, S. Shayan, K. J. Rasmussen, B. R. Ellingwood, System-based design of planar steel frames, i: Reliability framework, *Journal of Constructional Steel Research* 123 (2016) 135–143.
- [57] S. Shayan, K. J. Rasmussen, H. Zhang, On the modelling of initial geometric imperfections of steel frames in advanced analysis, *Journal of Constructional Steel Research* 98 (2014) 167–177.

- [58] E. Bernardini, S. M. J. Spence, D.-K. Kwon, A. Kareem, Performance-based design of high-rise buildings for occupant comfort, *Journal of Structural Engineering* 141 (10) (2015) 04014244.
- [59] D. K. Kwon, A. Kareem, R. Stansel, B. R. Ellingwood, Wind load factors for dynamically sensitive structures with uncertainties, *Engineering Structures* 103 (2015) 53–62.
- [60] A. G. Davenport, P. Hill-Carroll, Damping in tall buildings: its variability and treatment in design, in: *Building motion in wind*, ASCE, 1986, pp. 42–57.
- [61] Z. Ouyang, S. M. J. Spence, A performance-based wind engineering framework for envelope systems of engineered buildings subject to directional wind and rain hazards, *Journal of Structural Engineering* 146 (5) (2020) 04020049.
- [62] Z. Ouyang, S. M. J. Spence, Performance-based wind-induced structural and envelope damage assessment of engineered buildings through nonlinear dynamic analysis, *Journal of Wind Engineering and Industrial Aerodynamics* 208 (1) (2021) 104452.
- [63] W.-C. Chuang, S. M. J. Spence, A framework for the efficient reliability assessment of inelastic wind excited structures at dynamic shakedown, *Journal of Wind Engineering and Industrial Aerodynamics* 220 (2022) 104834.
- [64] M. Kuss, C. E. Rasmussen, R. Herbrich, Assessing approximate inference for binary gaussian process classification., *Journal of machine learning research* 6 (10) (2005).
- [65] C. K. Williams, D. Barber, Bayesian classification with gaussian processes, *IEEE Transactions on pattern analysis and machine intelligence* 20 (12) (1998) 1342–1351.
- [66] T. P. Minka, A family of algorithms for approximate bayesian inference, Ph.D. thesis, Massachusetts Institute of Technology (2001).