



Machine learning and model driven bayesian uncertainty quantification in suspended nonstructural systems

Zhiyuan Qin ^a, M.Z. Naser ^{a,b,*}

^a School of Civil & Environmental Engineering and Earth Sciences (SCEES), Clemson University, USA

^b Artificial Intelligence Research Institute for Science and Engineering (AIRISE), Clemson University, USA

ARTICLE INFO

Keywords:

Inverse problems
Machine learning
Gaussian process
Blackbox variational inference
Geometric complexity
Suspended nonstructural systems

ABSTRACT

This paper presents a novel framework for the uncertainty quantification of inverse problems often encountered in suspended nonstructural systems. This framework adopts machine learning- and model-driven stochastic Gaussian process model calibration to quantify the uncertainty via a new blackbox variational inference that accounts for geometric complexity through Bayesian inference. The soundness of the proposed framework is validated by examining one of the largest full-scale shaking table tests of suspended nonstructural systems and accompanying simulated (numerical) data. Our findings indicate that the proposed framework is computationally sound and scalable and yields optimal generalizability.

1. Introduction

To obtain accurate predictions, it is necessary to comprehensively consider various degrees of uncertainties [1], such as those arising from measurement uncertainty, model solution error, and model parameters/selection [2]. Generally, there are two types of uncertainty: *aleatoric* and *epistemic* [3]. Aleatoric or stochastic uncertainty describes the randomness observed in the data at hand. On the other hand, epistemic uncertainty, commonly referred to as system uncertainty, results from incomplete or incorrect information, such as limited experimental datasets or biased models.

The complexity of uncertainty quantification models tends to increase with the availability of computational resources. Although this increased complexity can lead to the creation of more accurate models, their associated costs may not be readily justifiable over traditionally reasonable approximations (or surrogates) [4,5]. With the rise of machine learning (ML) [6], surrogates have become attractive solvers [7]. For example, Gaussian Processes (GPs) have gained popularity in the supervised ML community for calibrating computer models and continue to serve as the foundation for contemporary methods [8]. GPs are especially well-suited for this purpose because they offer reliable uncertainty estimates in nonlinear systems, even with little training data. Recent developments have enabled more expressive GPs, despite the initial challenges with their scalability compared to neural networks [9,10].

Derivative-free Bayesian calibration is a statistical approach used to estimate unknown parameters in a mathematical model without requiring knowledge of the underlying analytical derivatives of the model. Traditional methods for derivative-free Bayesian calibration to estimate the posterior distribution, such as Markov chain Monte Carlo (MCMC), typically require a series of iterations—often more than 10^4 steps to reach statistical convergence [11]. Because each forward run can be expensive, conducting a series of runs is computationally unaffordable, rendering MCMC impractical for real-world calibration.

To overcome this challenge, Peterson and Hinton [12] studied Variational Inference (VI) methods, mainly to approximate the posterior probabilities in Bayesian models. The main idea behind variational inference is to seek a class of simple variational distributions to approximate the true posterior distribution through optimization rather than sampling processes. The variational distribution is a collection of hidden variable distributions that adopt KL divergence (Kullback-Leibler) as a measure of inter-distribution similarity and find the variational distribution closest to the true posterior distribution. Compared with Monte Carlo sampling, the variational inference method is faster, simpler, and easier to parallelize and hence, can be more suited for big data and complex models [13]. Variational inference is a popular technique in machine learning but is not as widely used as an MCMC-based sampling technique. In civil engineering, the slow uptake of VI can be attributed to its additional modeling complexity and limited theoretical exploration. Although the traditional mean field VI is

* Corresponding author.

E-mail address: mznaser@clemson.edu (M.Z. Naser).



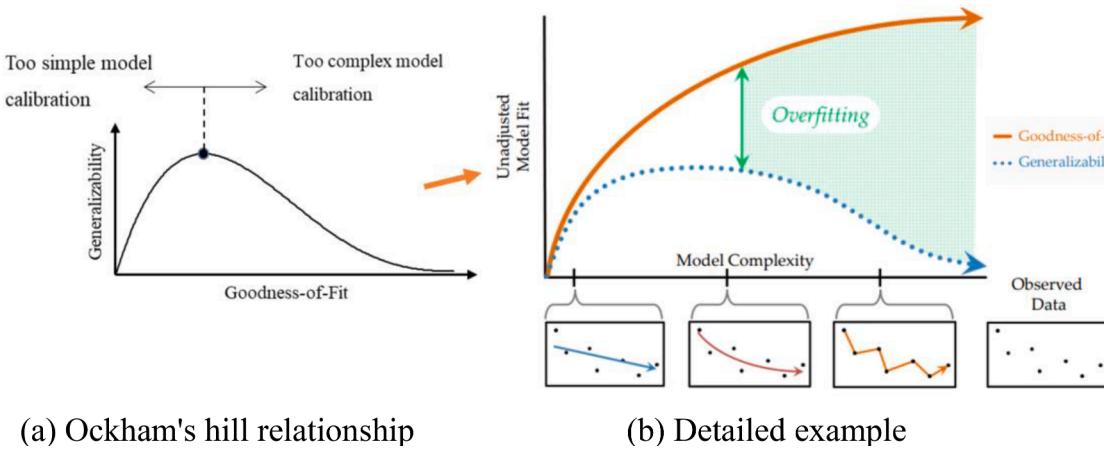


Fig. 1. Interplay among the goodness-of-fit, complexity, and generalizability.

commonly used, it requires complex mathematical derivations and conjugate assumptions, which limit its practical applications [14,15].

On the other hand, blackbox variational inference (BBVI) is a promising and advanced VI technique that remains largely unexplored in civil engineering [16]. BBVI does not require specific model derivations and can be scaled well to large datasets and high-dimensional parameter spaces. In contrast, MCMC-related methods are impractical for large datasets and may not scale well. Therefore, a combination of BBVI with O'Hagan's Bayesian calibration framework [8] can serve as an affordable combination; since such a combination can be easily derived without the need for conjugate assumptions. This enables BBVI to achieve superior results, which is critical for ensuring the reliability and safety of engineering systems.

While Bayesian calibration or inference is common, the model class selection is not. Addressing modeling complexity remains a significant challenge for Bayesian inference applications because integrating metamodeling techniques is not trivial. The challenge here is to establish a fully automated integration process that addresses different degrees of competency for the end user and a wide range of application problems with a certain degree of robustness.

Model calibration involves estimating the best-fit values for a few identifiable calibration parameters from experiments conducted under various control parameter settings. However, these models are inherently incomplete (i.e., systematically biased) and may not fully represent the actual system behavior. Such incompleteness may originate from various fronts, such as the omission of input parameters, interactions between the model input parameters and/or control variables, or assigning incorrect values to the model input parameters considered to be known [8]. This could have led to systematic discrepancies. However, such inherent discrepancies can be identified during model calibration by inferring an independent error model from the experimental data or by blending emulators with physics-based models to explain the omitted relationships between the model input parameters [4,17].

In model calibration, the goodness-of-fit of a model to the experimental data measures how well the model captures the data observed during calibration. However, a good fit is a necessary (but not sufficient) condition, as it is possible to calibrate physics-based models to different sets of calibration parameter values to fit a finite set of experiments reasonably well because of the inevitable compensations between various sources of errors and uncertainties [18,19]. Unlike goodness of fit, generalizability is defined as the ability of a model to represent the reality of interest in all settings, including those in which experiments are not available [20]. The generalizability of a calibrated model is critical because computer models are most often calibrated to predict conditions for which experiments are unavailable.

The complexity of a model calibration campaign results in an

Ockham hill relationship between good fitness and a finite number of noisy measurements in the tested settings and the generalizability of the model predictions in the untested settings [20] (See Fig. 1(a)). A model calibration campaign that lacks complexity could lose valuable information that could have been inferred from the data. However, a highly complex model is likely to fit noisy measurements, seemingly improving the goodness of fit while degrading its generalizability. In an extreme scenario, a calibration campaign that produces a model that perfectly matches all possible outcomes (i.e., infinite flexibility) will yield an uninformative tool that is impossible to falsify and one with little or no generalizability. For physics-based models, the inherent functional structure of the model imposes a differential ability to fit patterned data and prevents us from reaching a hypothetical infinite flexibility. The differential ability of a model to fit data is referred to as "selectivity" [21].

In Fig. 1(b), we present a detailed illustration of the interplay between the model's goodness-of-fit, generalizability, and complexity. Cutting et al. [21] recognized that the number of parameters alone is an insufficient indicator of model complexity and advocated for evaluating a model's fitting power or scope to random data. These researchers advocate evaluating the model's scope by comparing its ability to fit actual system data with that of random data using binomial tests. Similarly, complexity has also been defined as the range of data patterns that a model can fit [22].

Pitt et al. [23] and Myung et al. [24] quantified a geometric complexity measure known as the Minimum Description Length (MDL) [22,25]. MDL is based on the understanding that the more data are compressed, the more information about the underlying regularities governing the process of interest is learned [24]. This metric considers experimental data as a code or description to be compressed by the model. The MDL also evaluates the model's ability to compress a dataset by extracting the necessary information from the data without random noise. Therefore, MDL enables selecting a model with the shortest description code (length) in the data [25]. In Bayesian inference problems, most studies overlook generalization and model complexity or use simple criteria such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), or deviance information criterion (DIC). This can result in overfitting and poor generalization. We propose using MDL based on algorithmic information theory, Kolmogorov complexity, and geometric complexity measure of the data space. One of the goals of the present study is to assess the performance of MDL against other metrics in the case of suspended nonstructural systems (SNS).

In recent years, moderate or strong earthquakes have caused significant property loss, interruption of building function, and even threatened life safety owing to damage to suspended nonstructural systems (SNS) [26]. Despite minor damage to the main building structures, their impact on SNS underscores their crucial role in ensuring the

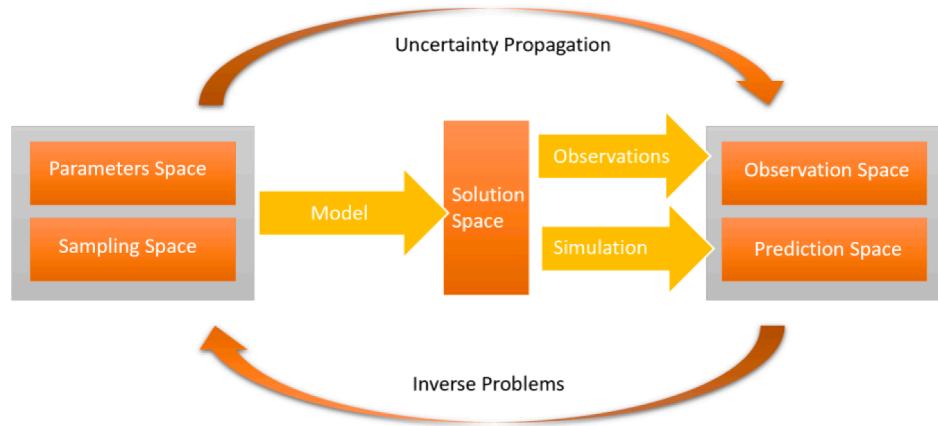


Fig. 2. Classification of the main problems within uncertainty quantification.

resilience of buildings against seismic events. Despite ongoing research efforts, the effects of ultra-large areas, long durations, and long periods under uncertain conditions are still unknown. Fortunately, we recently completed one of the world's largest full-scale suspended nonstructural system (SNS) experiments [26], in which we carefully designed earthquake wave inputs in line with long duration and long periods in supertall buildings, serve as viable candidates for use in uncertainty quantification and inverse problem inference that are a critical gap our study aims to fill.

2. Background and literature review

2.1. Uncertainty quantification (UQ)

The topic of uncertainty quantification (UQ) has garnered significant interest from researchers across various disciplines [1]. Theoretical developments in UQ have drawn on multiple fields, including probability and statistics, functional analysis, and more, resulting in various mathematical, statistical, Bayesian, optimization, and approximation techniques [22]. To better understand the challenges of UQ, Stark et al. [27] distinguished between the propagation of uncertainty (PoU) and its inverse, as shown in Fig. 2.

2.2. Inverse problems (model inference/calibration)

The inverse problem, also known as backward uncertainty propagation (Backward PoU), involves regression and parameter estimation from a Bayesian perspective [28]. This requires determining the nature of the input variable based on known observations, unlike the PoU technique, which determines the distribution features of the Quantities of Interest (QoI) function through the propagation of the input variable in the surrogate mode.

The ill-posed inverse problem is often encountered in parametric regression, but regularization techniques such as Tikhonov regularization [29] can be applied to resolve such unsolved problems. Tarantola [30] offered a Bayesian explanation of inverse problem theory and general methods for model parameter estimation, whereas Engl [31] delved into the theory of regularization methods and their application to inverse problems. Cotter et al. [32] established a framework for Bayesian inverse problems on separable Banach and Hilbert spaces using Gaussian priors and discretized parameter spaces. Dashti [33] employed a wavelet function in L^2 space to address infinite-dimensional nonparametric regression problems.

The measurement of posterior distributions is a key challenge in inverse problems. The Bayesian calibration approach, which uses posterior distributions of parameters, addresses this issue by defining posterior distributions over infinite-dimensional spaces using Kolmogorov's

definition of conditional probability. Lasanen [34] extended the Bayes formula to locally convex Suslin topological linear spaces, and generalizations of the formula as a Radon-Nikodym derivative were discussed in [35]. Maximum posterior estimation is widely used for extracting information from the posterior distribution because it connects the Bayesian technique with traditional regularization functions [35]. The most common approach for sampling the posterior distribution is the MCMC method, which constructs the probability of transition between the prior and posterior distributions. The open literature presents other methods with thorough analyses of Bayesian computation techniques for parametric models [36] and infinite-dimensional spaces [37].

2.3. Surrogate models

A surrogate model approximates a complex system and is typically a mathematical expression or algorithmic description of its inner relationships. Data-driven surrogates, such as artificial neural networks, radial basis functions, support vector machines, and Gaussian processes (GPs), are commonly used nowadays [38].

Gaussian processes (GPs) are nonparametric Bayesian models of interest in this study [39]. According to Rasmussen and Williams [40], any finite function sample follows a multidimensional Gaussian distribution, with the QoI function viewed as a specific sample of a Gaussian process. The Gaussian process is the response function's probability distribution, making it a flexible model for many applications. Rasmussen and Williams [40] explored the relationship between GPs and other models, such as regularization methods, and applied GPs theory to machine learning. Hyperparameters of the GP model are typically estimated using maximum likelihood estimation (MLE) or Bayesian information criterion (BIC) criteria, with posterior distributions of hyperparameters often sampled using MCMC methods in applications sensitive to parameter uncertainty.

Gaussian Processes (GPs) provide a powerful tool to model complex systems due to their ability to capture uncertainty and avoid overfitting. The expressiveness of a kernel function plays a crucial role in determining the complexity of single-layer GP models. Recent advances in GP research have applied these to sparse variational inference, enabling efficient computation for large datasets. The induction point approach and expectation propagation have been employed to infer deep GPs with sparsity [41]. Moreover, a novel inference method based on the induction point using the accurate model as the variational posterior distribution has been proposed to improve model accuracy further [42]. Furthermore, the doubly stochastic variational inference technique has been applied to deep GPs, combining random sampling and stochastic inference to improve the independence assumption and variance estimation of implicit function variables.

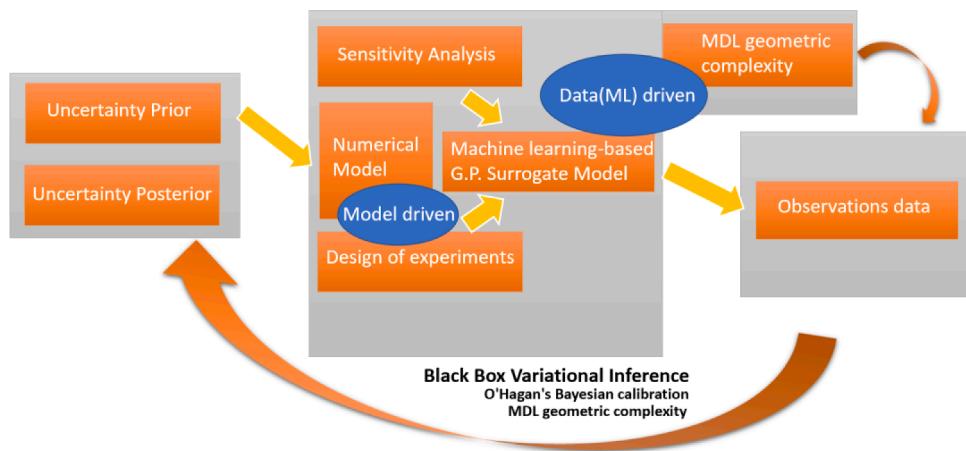


Fig. 3. Framework of machine learning-based data & model driven UQ of inverse problems.

3. A novel framework for machine learning-based data and model driven uncertainty quantification of inverse problems

3.1. Proposed framework

Fig. 3 summarizes our proposed framework. The model-driven approach involves using finite element numerical modeling to simulate data via experiments, providing a large dataset to train the data-driven, machine learning based surrogate model. This approach is used to solve forward problems. The data-driven (machine learning-driven) approach involves sensitivity analysis and training the surrogate model with both full-scale observations (experimental) data and simulated data to solve the inverse problem of Uncertainty Quantification (UQ) for SNS systems, which often have uncertain parameters, including initial and boundary conditions, material properties, and geometry, that can vary in space or time [43,44]. In the new approach, we propose a new blackbox (non-conjugate) variational inference method combined with O'Hagan's Bayesian calibration framework and embed geometric complexity measure MDL model selection to enhance the accuracy, efficiency, and robustness of both forward and inverse problems with generalization capabilities. In the next following sections, we will provide a detailed description of the methods used in our framework.

3.2. Machine learning-based Gaussian process surrogate model

3.2.1. Gaussian process surrogate model

We define the expectation and covariance of a real-valued GP $f(x)$, $m(x)$, and $k(x, x')$. The GP can be expressed as:

$$f(x) \sim GP(m(x), k(x, x')) = GP(\mathbb{E}[f(x)], \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))]) \quad (1)$$

Assuming that each observation is mapped using an unknown function $f(x)$ with independent Gaussian noise in Eq. (2) and the correlation of the observations of the inputs x_p, x_q :

$$y_i = f(x_i) + \epsilon_i, \text{ cov}(y_p, y_q) = k(x_p, x_q) + \sigma^2 \delta_{pq} \quad (2)$$

where δ_{pq} is the Dirac function when $p = q$, $\delta_{pq} = 1$, otherwise δ_{pq} is 0. Then, we can get marginal likelihood $P(y|x)$ directly from the prior $y \sim N(0, K + \sigma_n^2 I)$:

$$\log p(y|x) = -\frac{1}{2} y^T (K + \sigma_n^2 I)^{-1} y - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{n}{2} \log(2\pi) \quad (3)$$

3.2.2. Optimization of hyperparameters in GPs models

Choosing the appropriate hyperparameter is crucial as it determines

the specific form of the kernel function. For example, in the one-dimensional squared exponential covariance function¹:

$$k_y(x_p, x_q) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}(x_p - x_q)^2\right) + \sigma_n^2 \delta_{pq} \quad (4)$$

where, l, σ_f, σ_n are hyperparameters in kernel functions. The model's generalization ability is impacted by various hyperparameters. Here, we use the maximal marginal likelihood method to optimize the hyperparameters, and from Eq. (5), we can attain the logarithmic form of $p(y|x, \theta)$ with hyperparameters represented by the vector θ explicitly. To maximize the marginal likelihood, its partial derivative form for each hyperparameter is:

$$\frac{\partial}{\partial \theta_j} \log p(y|x, \theta) = \frac{1}{2} y^T K^{-1} \frac{\partial K}{\partial \theta_j} K^{-1} y - \frac{1}{2} \text{tr}\left(K^{-1} \frac{\partial K}{\partial \theta_j}\right) = \frac{1}{2} \text{tr}\left((\alpha \alpha^T - K^{-1}) \frac{\partial K}{\partial \theta_j}\right) \quad (5)$$

where, $\text{tr}(\cdot)$ is the trace of the matrix, $\alpha = K^{-1}y$, K^{-1} determines the computational complexity of (5), Once K^{-1} is obtained, it has a gradient complexity of $O(n^2)$ for hyperparameter θ_j .

3.3. Bayesian inference

Bayesian inference provides an approach to the estimation or calibration of a set of parameters Θ in a model (or hypothesis) H for the data D . It is based on a likelihood function derived from a specific probability model of the observed data $L(D|\Theta)$, where Θ is stochastic, it has a prior $\pi(\theta)$. The inference is based on the posterior $\pi(\theta|D)$ as obtained by Bayes' theorem:

$$\Pr(\Theta|D, H) = \frac{\Pr(D|\Theta, H)\Pr(\Theta|H)}{\Pr(D|H)}, \quad \Pr(D|H) = \int \Pr(D|\Theta, H)\Pr(\Theta|H)d\Theta \quad (6)$$

where Θ represents the tensor of uncertain parameters to be estimated, and D represents the tensor of the observations or measurement data to calibrate or estimate our knowledge of Θ . H represents the model or hypothesis which is believed to best represent the available D . $\Pr(D|H) \equiv Z$ is the Bayesian evidence that serves as the normalizing constant of the posterior. The Bayesian evidence factor can be ignored in the estimation process since it is independent of the parameters Θ , and inferences are obtained by computing or sampling the unnormalized posterior [45].

¹ The subsequent study of SNS in this article will use this covariance function after optimization of kernel selection.

3.4. Bayesian calibration

A statistically descriptive model calibration procedure [45] uses Bayesian inference to model the relationship between computer simulations and observed data y , while considering parameter uncertainties, model discrepancy, and observation error.

$$y(x) = \eta(x; t^*) + \delta(x) + \epsilon \quad (7)$$

where, $y(x)$ and $\eta(x; t^*)$ are the observation data and simulation output, respectively. $\delta(x)$ is the discrepancy/bias term accounting for model inadequacy between simulation and physical system at input x . Inadequate, missing physics, and numerical errors in the code could be the cause of this inadequacy. ϵ describes observational data variation, and it is often assumed to have a Gaussian distribution. And t^* represents the true but unknown values of the calibration parameters t .

3.5. Variational inference

In complex, high-dimensional problems, such as those encountered in suspended nonstructural systems, stochastic sampling techniques are used to draw samples from unnormalized distributions when the analytical solution of $Pr(\Theta|D, H)$ is not easily obtained. MCMC schemes, including Gibbs sampling and Metropolis-Hastings, cancel out the Bayesian evidence Z during single-sample computation. MCMC is an irreducible, aperiodic Markov chain that meets detailed balance conditions, and the Kullback-Leibler (KL) divergence monotonically decreases during Markov transition, according to standard theory [46]. MCMC sampling can achieve convergence to the true posterior with infinite samples, but it is slow and unbiased. Variational inference is a popular indirect approximation method [15] that minimizes the distance measure KL divergence between approximate and posterior distributions. This method avoids the calculation of the normalization constant and only requires the joint distribution of observable and latent variable x, z . Compared to MCMC, variational inference is a computationally efficient optimization method for variable posterior distributions [16]. It transforms inference into an optimization problem by selecting an easy-to-handle distribution $q(x)$ that closely approximates the true posterior distribution.

3.5.1. Inference and optimization

In variational inference, the goal is to find an easy-to-handle distribution $q(z)$ that approximates the posterior distribution $p(z|x)$ and calculates the marginal probability $p(x)$ of the observed variable. The inference problem is turned into an optimization problem by minimizing the distance measure between the variational distribution and the posterior distribution. This optimized variational distribution $q(z;\lambda)$ can be used instead of the posterior distribution, and λ is used to represent the variational parameters.

Definition 3.1. For the probability distributions $p(z)$ and $q(z)$, the KL divergence distance between them is defined as:

$$D_{KL}(q(z) \parallel p(z)) = - \int q(z) \log \frac{p(z)}{q(z)} dz \quad (8)$$

3.5.2. Variational lower bound

To minimize KL divergence, special treatment of the "evidence term" and "model conditional" $p(z|x)$ is required. However, MCMC may not efficiently approximate this treatment. Instead, we can use ELBO (Evidence Lower Bound Objective) as a maximized log-likelihood variational lower bound, which is a conservative estimate of the marginal distribution and can indicate how well the data distribution $p(x)$ fits the model.

Definition 3.2. With Jensen inequality, we can briefly derive the

variational lower bound ELBO from the log-likelihood $\log p(x)$, and the distance measured between its logarithmic marginal distribution and the ELBO is the KL divergence distance between variational distributions:

$$\begin{aligned} \log p(x) &= \log \mathbb{E}_{q(z;\lambda)} \left[\frac{p(x, \lambda)}{q(z; \lambda | x)} \right] \\ &= \int q(z; \lambda) \log \left(\frac{p(z, x)}{q(z; \lambda)} \right) dz - \int q(z; \lambda) \log \left(\frac{p(z|x)}{q(z; \lambda)} \right) dz \\ &= ELBO(q) + D_{KL}(q(z; \lambda) \parallel p(x, z)) \\ &\geq \mathbb{E}_{q(z;\lambda)} \log \left[\frac{p(x, z)}{q(z; \lambda | x)} \right] = ELBO(q) \text{ by Jensen's inequality} \end{aligned} \quad (9)$$

Maximizing ELBO is equivalent to minimizing $D_{KL}(q(z|x) \parallel p(z))$, which requires a simple and expressive variational distribution $q(z)$ to approximate the posterior distribution. A common choice is the Mean-Field Distribution. Traditional variational inference requires an explicit calculation of ELBO, which is only possible for directly conjugated variables. However, newer methods do not require this and are preferred in this study [46].

3.5.3. Blackbox (non-conjugate) variational inference

In traditional variational inference, ELBO can only be explicitly calculated and optimized for conditionally conjugated exponential family distributions, leading to unsolvable expectation calculations for most models. To extend variational inference to non-conjugate cases and automate the process, the blackbox Variational Inference (BBVI) [16] is proposed, allowing unbiased estimation of the variational parameter gradient by sampling from the variational distribution. This method is a more general inference method that does not constrain the probability distribution of intermediate variables. ELBO is maximized through updates based on the gradient or stochastic gradient of variational parameters, expressed as the expectation of the variational distribution:

$$\nabla_\lambda L = \mathbb{E}_q[\nabla_\lambda \log q(z|\lambda)(\log p(x, z) - \log q(z|\lambda))] \quad (10)$$

For gradients $\nabla_\lambda L$, stochastic gradient calculation can also be used to sample from the variational distribution and estimate gradients as follows:

$$\nabla_\lambda \hat{L}_{stoch} = \frac{1}{K} \sum_{k=1}^K \nabla_\lambda \log q(z_k|\lambda)(\log p(x, z_k) - \log q(z_k|\lambda)) \quad (11)$$

where $z_k \sim q(z|\lambda)$, the proposed variational inference method provides a black box gradient estimation technique, requiring sampling of K samples from observed and hidden variables instead of explicit calculation of ELBO gradients. This calculation can be implemented using commonly available deep learning methods, which are used in the subsequent SNS analysis.

3.6. Design of experiments

Experimental design [47,48] can be divided into model unknown and model known designs [49]. In the study of SNS systems, we will use the Latin hypercube sampling (LHS) method for sampling, which produces a more uniform distribution in the parameter space than the MC method. Taking the sampling of the random vector $x \sim U([0, 1]^d)$ as an example, the LHS method is divided into the following three steps: *Step 1*: Divide $[0, 1]$ on each dimension into N equal parts, N is the number of samples, and construct N^d small hypercubes, written as $\{c_i\}_{i=1}^{N^d}$, where $i = (i_1, \dots, i_d)$ is a D -dimensional indicator and $\|i\|_{AptCommand2016;1} = \sum_{j=1}^d i_j$; *Step 2*: Select N c_i so that the indicators i^1, i^2 of any two small hypercubes satisfies: $i_j^1 \neq i_j^2, j = 1, \dots,$

Table 1
Widely used model selection metrics.

| Selection Metric | Criterion Equation |
|---|--|
| Akaike information criterion (AIC) | $AIC = -2 * \ln(f(y \hat{\theta})) + 2k$ |
| Bayesian information criterion (BIC) | $BIC = -2 * \ln(f(y \hat{\theta})) + 2k * \ln(n)$ |
| Deviance information criterion (DIC) | $DIC = \overline{D(\theta)} + p_D$ |
| Information-theoretic measure of complexity (ICOMP) | $ICOMP = -\ln(f(y \hat{\theta})) + \frac{k}{2} \ln\left(\frac{\text{trace}[\Omega(\hat{\theta})]}{k}\right) - \frac{1}{2} \ln(\det[\Omega(\hat{\theta})])$ |
| Minimum description length (MDL) | $MDL = -\ln(f(y \hat{\theta})) + \frac{k}{2} \ln\left(\frac{n}{2}\right) + \ln(\int d\theta I(\hat{\theta}))$ |

Note: y = data function; n = sample of size; $\hat{\theta}$ = parameter value that maximizes the likelihood function $f(y|\hat{\theta})$; k = number of parameters; D is the deviance of the likelihood, $D(\hat{\theta}) = -2 * \log(f(y|\hat{\theta}))$; $p_D = \overline{D(\hat{\theta})} - D(\bar{\theta})$, $\overline{D(\theta)}$ is the expectation of $D(\hat{\theta})$ and $\bar{\theta}$ is the expectation of $\hat{\theta}$; Ω = covariance matrix of the parameter estimates; \ln = the natural logarithm of base e.



Fig. 4. View of steel platform and suspended ceiling and cable tray systems [26].

Table 2
Details of motions input to specimen [26].

| Run No. ¹⁾ | Name of input motion | Target acc. of the table (g) | | Duration (s) | PFA of the platform (g) | |
|-----------------------|----------------------|------------------------------|--------|--------------|-------------------------|--------|
| | | X dir. | Y dir. | | X dir. | Y dir. |
| 2 | Sweep | 0.050 | 0 | 100 | 0.071 | – |
| 4 | Sweep | 0 | 0.050 | 100 | – | 0.087 |
| 6 | Sweep | 0.050 | 0.050 | 100 | 0.069 | 0.087 |
| 8 | BCJ-L2 | 0.037 | 0 | 120 | 0.050 | – |
| 10 | BCJ-L2 | 0 | 0.037 | 120 | – | 0.057 |
| 12 ²⁾ | SHW6 (5/128) | 0.089 | 0.070 | 70 | 0.127 | 0.098 |
| 14 ²⁾ | SHW6 (128/128) | 0.149 | 0.132 | 70 | 0.146 | 0.153 |
| 16 ²⁾ | SHW6 (30/30) | 0.405 | 0.377 | 150 | 0.571 | 0.573 |
| 18 | Sweep | 0.150 | 0 | 100 | 0.225 | – |
| 20 | Sweep | 0 | 0.150 | 100 | – | 0.242 |
| 22 | Sweep | 0.150 | 0.150 | 100 | 0.232 | 0.277 |
| 24 | Sweep | 0.250 | 0 | 100 | 0.393 | – |
| 26 | Sweep | 0 | 0.250 | 100 | – | 0.512 |
| 28 | Sweep | 0.350 | 0 | 100 | 0.572 | – |
| 30 | Sweep | 0 | 0.350 | 100 | – | 1.942 |
| 32 | Sweep | 0.500 | 0 | 100 | 1.319 | – |

Notes:

¹⁾ Runs of odd numbers used for white-noise excitation with small magnitude are not listed in the table.

²⁾ During Runs 12 and 14, the floor acceleration responses at the 5th and 128th floors of the 128-story building subjected to the ground motion SHW6 are input, and during Run 16, the floor acceleration response at the 30th floor of the 30-story building is input.

d; Step 3: In each selected small hypercube, a random sample is taken according to a uniform distribution, and its setting is the desired sample set. It should be noted that although the number of constructed small hypercubes N^d increases exponentially with dimension d , the computational complexity of *Step 2* can be reduced to $O(dN)$ through algorithmic optimization.

3.7. Sensitivity analysis and model selection

3.7.1. Sensitivity analysis

Sensitivity Analysis (SA) is a key component of UQ that measures the impact of perturbations on the QoI function $y(x, \theta)$ in terms of mean, variance, distribution, and information entropy. This analysis helps to minimize QoI uncertainty and control variables. Local SA involves approximating the model using a Taylor expansion, while global SA can be done using screening or variance decomposition methods. Screening identifies variables with a significant impact on QoI uncertainty, while variance decomposition quantifies the proportion of each variable's impact. SA is also related to model selection, and variance-based methods are typically more applicable for global sensitivity analysis. In this study, we will use the variance-based sensitivity analysis method described in [49] to analyze subsequent SNS.

3.7.2. Minimum description length model selection

According to Sober [50], informative models are less complex. Kuhn [51] also noted that simpler models are preferable to complex ones. Turney [52] demonstrated that simpler models are more stable (robust) in the face of experimental uncertainty. Model selection criteria (listed in Table 1) aid in comparing alternative calibration campaigns by considering goodness-of-fit and complexity.

MDL [22] is a geometric complexity model selection metric that compresses the experimental data to evaluate models based on their ability to extract necessary information from the data without random noise. The metric favors models with shorter description codes of the data [51,52], allowing for greater learning of underlying regularities governing the process of interest [24]. Using the concept of Normalized Maximum Likelihood, Rissanen further demonstrated the redundancy of the observation model has the asymptotic optimal lower bound of Minimax [53]:

$$\inf_{f_\theta} \sup_{\theta \in \Theta} K - L(q, f_\theta) = \frac{k}{2} \log \frac{n}{2\pi e} + \log \int_K \sqrt{\det I(\theta)} d\theta + O(1) \quad (12)$$

where, θ is a compact subset of the parameter space Θ , k represents the number of parameters entering the model, $I(\theta)$ is the Fisher information matrix of the parameter distribution, and an expectation defines it $I_{ij} = - < \left(\frac{\partial^2 \log(D(\theta))}{\partial \theta_i \partial \theta_j} \right) >_\theta$ over the parametric model $f_\theta(x^n)$. It rigorously shows that model complexity estimates must approach the lower bound progressively and has good generalizability [49] because it does not assume knowledge of the true data distribution or make specific assumptions about the observed data distribution. In Section 5, we will thoroughly compare the MDL approach with other model selection criteria, including AIC, BIC, DIC, and ICOMP.

4. SNS shaking table experiments

4.1. SNS shaking table experiment configurations

To validate our proposed framework, we utilized a recent large-scale shaking table experiment that we conducted in [26], with a brief summary presented herein for completion. Our experiment investigated the seismic effects of nonstructural systems under various seismic input wave conditions [54]. Suspended nonstructural systems, consisting of a suspended ceiling system and a suspended cable tray system, were tested using a steel platform. This platform, the largest of its kind to date, is two

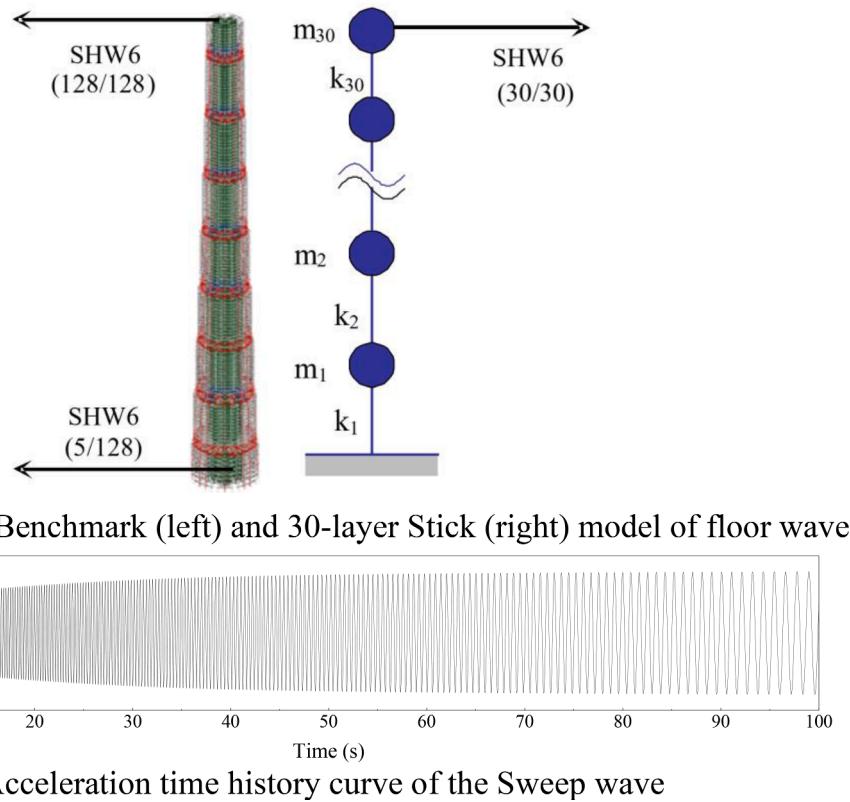


Fig. 5. Schematic of partial motions of input.

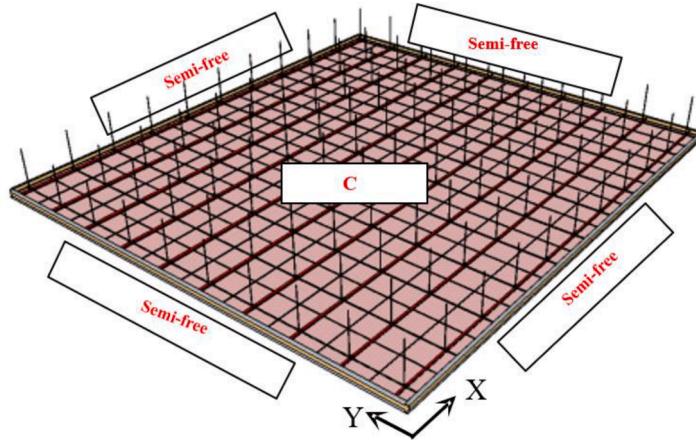


Fig. 6. SCS type C [26].

stories high, measuring 5.40 m in height, 12.84 m in length, and 11.64 m in width. During the test, the systems were suspended from the platform (see Fig. 4).

4.2. Loading protocol

To accurately simulate the long duration and long period seismic input experienced by supertall buildings, the test specimen was loaded with various frequencies after analysis (see Table 2 for a comprehensive list of all the motions tested [26]). Following each run, white noise with a PGA of 0.05 g was input to the specimen to evaluate its dynamic characteristics. The shaking table was subjected to several sets of motions, including sweep waves (Sweep), acceleration responses obtained by stochastic time history analysis at different floors of building

structures, and artificial waves (BCJ-L2).

The sweeping wave in Fig. 5(b) is a sine wave oscillating with a frequency range of 6.0 Hz to 0.8 Hz. The PFA at the floor level of the platform is presented in Table 2. Three input motions were selected for analysis, namely SHW6 (5/128), SHW6 (128/128), and SHW6 (30/30) (see Fig. 5(a)). These motions represent the acceleration responses at the 5th floor and the top of a 128-story supertall structure model and the acceleration responses at the top of a 30-story stick model subjected to the ground motion SHW6 with PGA of 0.1 g, respectively [26]. The natural vibration periods of the first three modes of the 128-story model are 8.94 s, 8.93 s, and 4.48 s, respectively. The corresponding results of the 30-story model are 3.01 s, 1.18 s, and 0.72 s, respectively. The characteristic period of ground motion SHW6 is 0.9 s. The PFA of SHW6 (128/128) is 1.7 times that of SHW6 (5/128). The fundamental period of

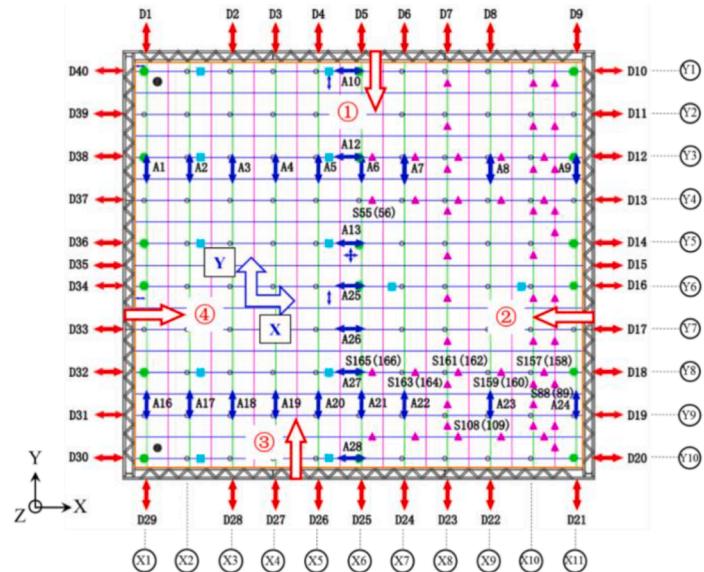


Fig. 7. Instrumentation on the ceiling [26].

Notes: \rightarrow displacement transducer; \leftarrow accelerometer; ● strain gage on the threaded rod; ● strain gage on lay-in panel; ▲strain gage on the cross tee and sub cross tee; ■ strain gage on carrying channel.

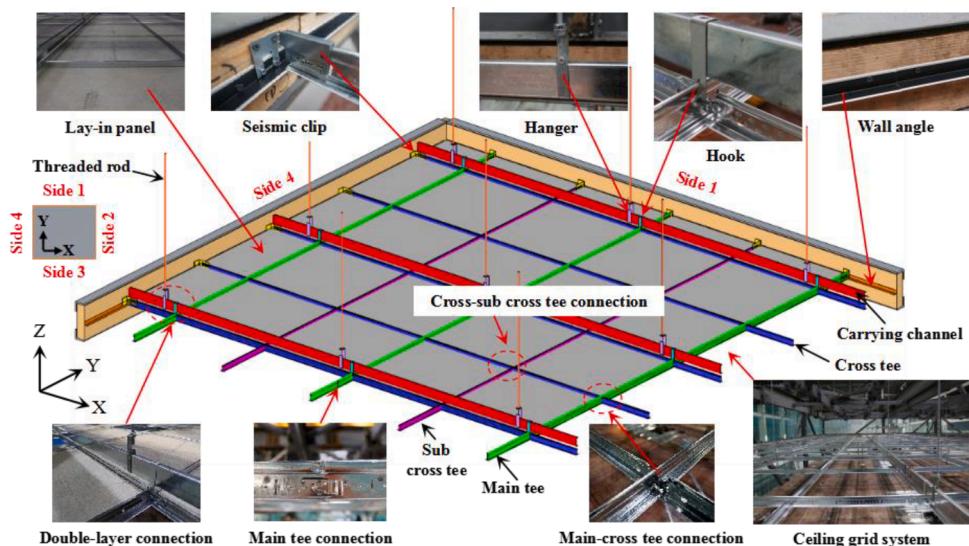


Fig. 8. Composition of double-layer SCS [97].

the 30-story model is closer to the characteristic period of the ground motion SHW6 than that of the 128-story model; therefore, the PFA of SHW6 (30/30) is larger than that of SHW6 (128/128).

4.3. Type C general information

As a case study, we selected one of the SNS experiments, the type C Suspended Ceiling System (SCS), to test our proposed uncertainty quantification methodology. This SCS consists of a single suspended ceiling with a complete semi-free boundary constraint condition, as depicted in Fig. 6. The suspended ceiling covers an area of 150 m² and is currently the largest ceiling worldwide.

4.4. Instrumentation

This specimen was instrumented with 236 instruments, including 30 accelerometers (A1-A30), 40 displacement transducers (D1-D40), and

166 strain gauges (S1-S166) to measure the dynamic response of the ceiling (see Fig. 7). The accelerometers measured the absolute acceleration, whereas the displacement transducer measured the displacement of the ceiling relative to the platform.

4.5. Boundary conditions

The boundary conditions for the SCS are shown in Fig. 8 [26]. The system has a semi-free boundary on sides 1 to 4, with main tees and sub-cross tees attached to the wall angles on side 1 and cross tees attached to the wall angles on side 2. Sides 3 and 4 have the same boundary conditions as sides 1 and 2.

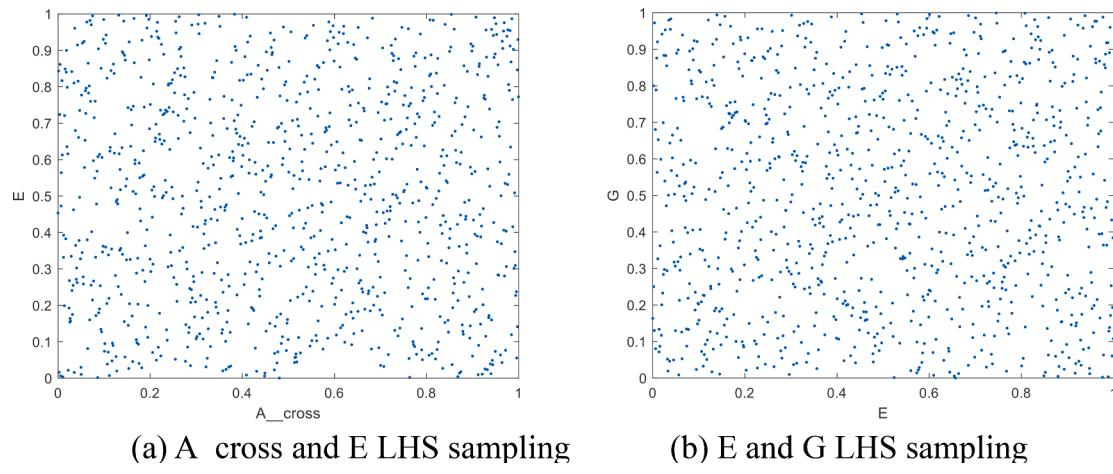


Fig. 9. Latin Hypercube samples results.

Table 3
Inverse problem parameters prior and Observations data prior.

| Inverse problem parameters prior | | |
|----------------------------------|---------------------------------------|--------------------|
| Parameter label | Parameter meaning | Prior distribution |
| 1. A_cross | Cross tee cross-section area | U~(30, 36) |
| 2. E | Elastic modulus | U~(19, 2.3) 10^5 |
| 3. G | Shear modulus | U~(7.4, 8.2) 10^4 |
| 4. J_cross | Cross tee moment of Torsion | U~(4800, 5100) |
| 5. I_y_cross | Y-dir moment of inertia | U~(4000, 4700) |
| 6. I_z_cross | Z-dir moment of inertia | U~(650, 750) |
| 7. K_rod | Threaded rods connection stiffness | U~(0.5, 3) |
| 8. K_friction | Friction elastic stiffness | U~(15, 25) |
| 9. K_cc1 | X-dir Cross tee connections stiffness | U~(750, 850) |
| 10. K_cc2 | Y-dir Cross tee connections stiffness | U~(75, 125) |
| 11. K_cc3 | Z-dir Cross tee connections stiffness | U~(10, 35) |
| 12. A_main | Main tee cross-section area | U~(35, 42) |
| 13. J_main | Cross tee moment of Torsion | U~(7900, 8500) |
| 14. I_y_cross | Y-dir moment of inertia | U~(7250, 7800) |
| 15. K_mc | X-dir Main tee connections stiffness | U~(1400, 1600) |

| Observations data prior | | |
|-------------------------|----------------------|-------------|
| mode1 | SCS frequency mode 1 | (7.5±0.02) |
| mode2 | SCS frequency mode 2 | (8.1±0.006) |
| mode3 | SCS frequency mode 3 | (8.9±0.04) |
| mode4 | SCS frequency mode 4 | (9.3±0.05) |
| mode5 | SCS frequency mode 5 | (9.5±0.01) |
| moded6 | SCS frequency mode 6 | (9.8±0.02) |

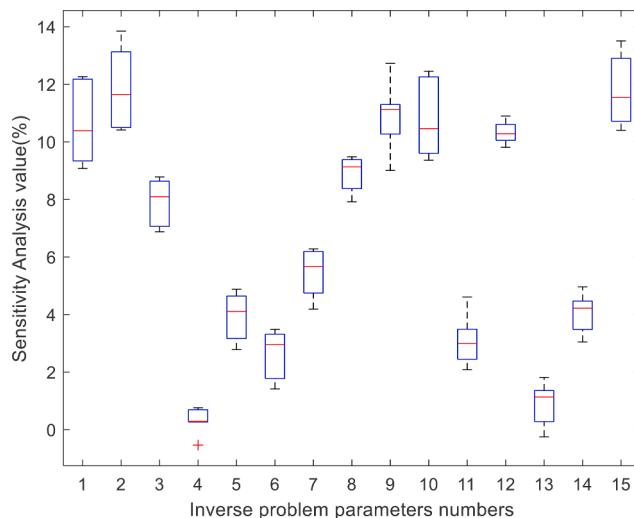


Fig. 10. Sensitivity analysis result.

5. Case study and applications of the proposed approach to suspended nonstructural systems (SNS)

5.1. Training points

As previously mentioned, the LHS method was used to sample combinations of the training points. For each pair of the two sets of training data, we generated 1000 Latin Hypercube samples (Fig. 9 shows two selected two pairs of training point data for illustration). This data is used for forward running (model-driven) the computational numerical model of SNS systems and use the maximin method projected in 2D (input: 0–1 uniform space, output: 0–1 uniform space). This particular design relies on the distance criterion, and the final design is the result of maximizing the minimum distance between the points.

5.2. Priors

The parameters² selected in this work were based on the linear model parameters (see Table 3). The fifteen selected inverse problem parameter names, symbols, prior, and observation prior data are as follows:

5.3. Sensitivity analysis

As discussed in Section 3.7.1, a sensitivity analysis was conducted to identify the key uncertainty parameters (refer to Fig. 10). The findings indicate that the 4th and 13th parameters have a negligible effect (less than 1%) on the results. Consequently, we retained the remaining thirteen inverse problem parameters and specified their prior ranges (see Table 4).

Our sensitivity analysis revealed that specific parameters, such as those associated with main tee connections, cross tee connections, and boundary connections, significantly influence the accuracy of the model's predictions. For instance, we observed that the acceleration prediction results were not precise between 60–61 s during the validation process. This inaccuracy can be attributed to bias and uncertainty in the parameters, which may have been affected by energy dissipation and friction effects in the connection and friction parts, leading to a notable difference in the amplitude. We will provide further insights into these findings in Section 5.7 of this article.

² The reader is to note that in this Bayesian inference problem, the focus of our study is mainly on linear systems and does not involve nonlinearities such as collision nonlinearity and connection nonlinearity problems because the training input is an inverse (calibration) parameter, and the output is the first six modes of frequency response data. The treatment for such nonlinearities will be addressed in a future work.

Table 4

Pre-defined prior box for the thirteen inverse problem parameters.

| Inverse problem parameters prior after sensitivity analysis | | |
|---|---------------------------------------|-------------------------------|
| Parameter label | Parameter meaning | Prior distribution |
| A_{cross} | Cross tee cross-section area | $U \sim (30, 36)$ |
| E | Elastic modulus | $U \sim (1.9, 2.3) * 100,000$ |
| G | Shear modulus | $U \sim (7.4, 8.2) * 10,000$ |
| I_{y_cross} | Y-dir moment of inertia | $U \sim (4000, 4700)$ |
| I_{z_cross} | Z-dir moment of inertia | $U \sim (650, 750)$ |
| K_{rod} | Threaded rods connection stiffness | $U \sim (0.5, 3)$ |
| $K_{friction}$ | Friction elastic stiffness | $U \sim (15, 25)$ |
| K_{cc1} | X-dir Cross tee connections stiffness | $U \sim (750, 850)$ |
| K_{cc2} | Y-dir Cross tee connections stiffness | $U \sim (75, 125)$ |
| K_{cc3} | Z-dir Cross tee connections stiffness | $U \sim (10, 35)$ |
| A_{main} | Main tee cross-section area | $U \sim (35, 42)$ |
| I_{y_cross} | Y-dir moment of inertia | $U \sim (7250, 7800)$ |
| K_{mc} | X-dir Main tee connections stiffness | $U \sim (1400, 1600)$ |

5.4. Data driven machine learning-based Gaussian process surrogate model training

A nonstructural system with n degree of freedom (DOFs) can be described by

$$M\ddot{x}(t) + C\dot{x}(t) + Kx(t) = F(t) \quad (13)$$

where, M , C , and K are the mass, damping, and stiffness matrix, respectively; $\ddot{x}(t)$, $\dot{x}(t)$, and $x(t)$ are the acceleration, velocity, and displacement of the nonstructural system. where $F(t)$ denotes the excitation force applied to the system.

The measured acceleration response can be expressed in a discrete form as

$$y_{obs}(x, t) = d\ddot{x}(t) + e_i \quad (14)$$

where, y_{obs} is the observed or measured response at time instant t . $\ddot{x}(t)$ and e_i are the values of the acceleration response of the system and observation noise. Matrix d is a sensor placement matrix associated with the accelerometer locations.

We now consider Kennedy and O'Hagan's formulation [4], where the data observation (frequency response) is a computational numerical simulation are available, namely $y_{obs}(x)$ and $y_{sim}(x, t)$. The design variable x is assumed to be taking values within a feasible design space

$X \subset \mathbb{R}^D$, and θ is a set of parameters to be calibrated or inferred.

The relationship between data observation and simulation is assumed to be

$$y_{obs}(x, t) = \alpha y_{sim}(x, \theta, t) + \delta(x, t) + \epsilon(x) \quad (15)$$

where $\delta(x)$ is a discrepancy term that is statistically independent of $y(x, \theta, t)$ and $\epsilon(x)$, which accounts for observation noise. The coefficient α satisfies

$$\alpha = \frac{\text{cov}[y_{obs}(x, t), y_{sim}(x, \theta, t)]}{\text{var}[y_{sim}(x, \theta, t)]} \quad (16)$$

which account the further we take $y_{sim}(x, \theta, t)$, $\delta(x)$ to be a Gaussian process with zero mean and variances $\sigma_{sim}^2 \gamma_{sim}(x, x')$ and $\sigma_\delta^2 \gamma_\delta(x, x')$, where γ_η and γ_δ are correlation kernels, as mentioned in Section 3.

To optimize kernel selection for Gaussian process regression, we explore different kernels as candidates and perform cross-validation to select the best one. The process involved dividing the data into training and validation sets, fitting the model with each candidate kernel on the training set, and evaluating the performance of the mean squared error score on the validation set. After selecting the appropriate kernel, hyperparameter tuning is performed to train the final model (as discussed in Section 3.2). Then, we identify the most appropriate kernel as the squared exponential kernel:

$$\gamma_\delta(x, x') = \exp \left[-\sum_{i=1}^D \frac{(x_i - x'_i)^2}{l_{i,\delta}} \right], \quad \gamma_{sim}(x, x') = \exp \left[-\sum_{i=1}^D \frac{(x_i - x'_i)^2}{l_{i,sim}} \right] \quad (17)$$

where $l_{i,sim}$, $l_{i,\delta}$ are the correlation length or length scale for the two kernels.

To further optimize the performance of the selected kernel, we perform hyperparameter tuning using maximum likelihood estimation to find the optimal values for the kernel's hyperparameters. This helps achieve the best performance of the final model.

We opted to utilize a standard approach for splitting the LHS sample data into training, validation, and testing sets. Specifically, 70% of the LHS samples were used to train the Gaussian process, 20% for validation, and 10% for testing. To evaluate the performance of the proposed model, we measured the loss and accuracy of the training process. The final training data had a loss of 0.343, while the test data had a loss of

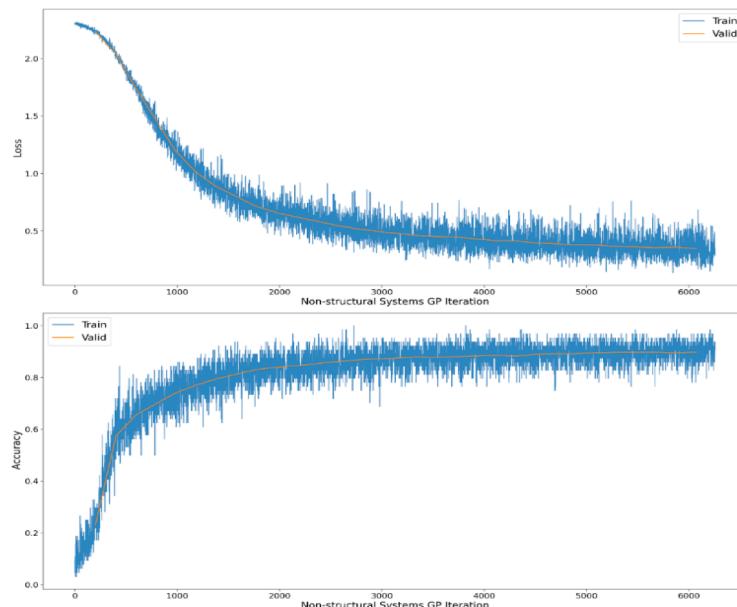


Fig. 11. GP surrogate model training result.

Algorithm 1

Blackbox variational inference with O'Hagan Bayesian calibration & geometric complexity MDL for inverse problem Uncertainty Quantification of suspended nonstructural system.

Input:

Training data d , mean and covariance functions for GPs in O'Hagan Bayesian calibration framework, variational family $q(z|\lambda)$, numbers of model type m

Output:

$q^*(z|\lambda)$ and MDL ranking

For $m = 1$ **to** M **do** // Given that M is a small constant, the time complexity $O(\cdot)$ is acceptable in this scenario

Initialize: $\lambda_{1:n}$ randomly, $t = 1$

while the training accuracy has not converged **do**

For $s = 1$ **to** S **do**

 // Draw sample from q

$z[s] \sim q(z|\lambda)$

end

$\rho := t^{\text{th}}$ value of a Robbins Monte sequence // Eq. (11) in Section 3.5.3

$\lambda := \lambda + \frac{1}{S} \sum_{s=1}^S \nabla_\lambda \log q(z[s]|\lambda) (\log p(x, z[s]) - \log q(z[s]|\lambda))$

$t = t + 1$

end

Return $q^*(z|\lambda)$

MDL := Minimum description length complexity computation // Equation in table 1 in Section 3.7.2

end

0.328, indicating the model's good performance and generalization ability. Furthermore, the training accuracy was 90.189%, and the testing accuracy was 90.527%, as shown in Fig. 11. These results demonstrate the effectiveness of our methodology in predicting the behavior of the system under study. Compared to traditional physical models, our surrogate model approach is significantly faster while achieving good extrapolation results, whereas our approach achieved several hundred-fold improvements in time. The surrogate models also enable us to consider a more extensive range of input parameters in the future, which is essential for SNS applications.

5.5. Blackbox variational inference

In order to avoid any computational inefficiency, we substitute the MCMC with the variational inference for posterior distribution estimation. Since mean variational field inference has conjugate requirements for the variational distribution $q(x)$, we propose to use the blackbox variational inference for analysis as mentioned above.

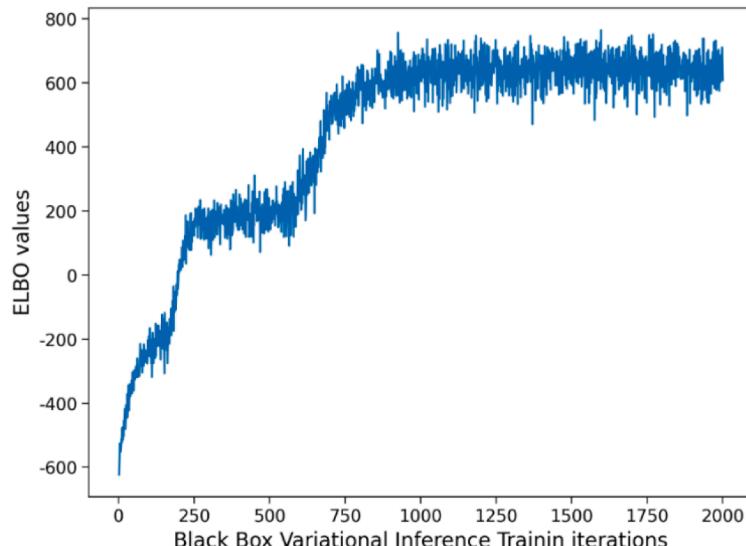


Fig. 12. Blackbox variational inference training iterations.

Initially, we utilized the O'Hagan Bayesian calibration framework [63] to derive and determine the analytical form of the joint posterior distribution. Furthermore, five sets of model types were constructed for inference during the posterior distribution inference, providing a basis for subsequent MDL model selection.

Since we consider the frequency response as the output, then the t in the Equation can be omitted from the below content. For example, assume that a set of simulations and data observations are available, namely, $D_{\text{sim}} = \{x_i, \theta_i, \eta_i\}_{(i=1)}^{N_{\text{sim}}}$, $D_{\text{obs}} = \{x_i^*, \theta_i, y_i\}_{(i=1)}^{N_{\text{obs}}}$, respectively.

Table 5
Minimum Description Length complexity (MDL) values.

| Model type | MDL |
|--|--------|
| $A_{\text{cross}}, E, G, I_y, \text{cross}, I_z, \text{cross}, K_{\text{rod}}, K_{\text{friction}}, K_{\text{cc1}}, K_{\text{cc2}}, K_{\text{cc3}}, A_{\text{main}}, I_y, \text{cross}, K_{\text{mc}}$ | -593.7 |
| $A_{\text{cross}}, E, G, I_y, \text{cross}, K_{\text{rod}}, K_{\text{friction}}, K_{\text{cc1}}, K_{\text{cc2}}, K_{\text{cc3}}, A_{\text{main}}, I_y, \text{cross}, K_{\text{mc}}$ | -587.3 |
| $A_{\text{cross}}, E, G, K_{\text{rod}}, K_{\text{friction}}, K_{\text{cc1}}, K_{\text{cc2}}, K_{\text{cc3}}, A_{\text{main}}, I_y, \text{cross}, K_{\text{mc}}$ | -597.9 |
| $A_{\text{cross}}, E, G, K_{\text{rod}}, K_{\text{friction}}, K_{\text{cc1}}, K_{\text{cc2}}, A_{\text{main}}, I_y, \text{cross}, K_{\text{mc}}$ | -584.5 |
| $A_{\text{cross}}, E, G, K_{\text{rod}}, K_{\text{friction}}, K_{\text{cc1}}, K_{\text{cc2}}, A_{\text{main}}, K_{\text{mc}}$ | -552.1 |
| $A_{\text{cross}}, E, G, K_{\text{friction}}, K_{\text{cc1}}, K_{\text{cc2}}, A_{\text{main}}, K_{\text{mc}}$ | -522.4 |

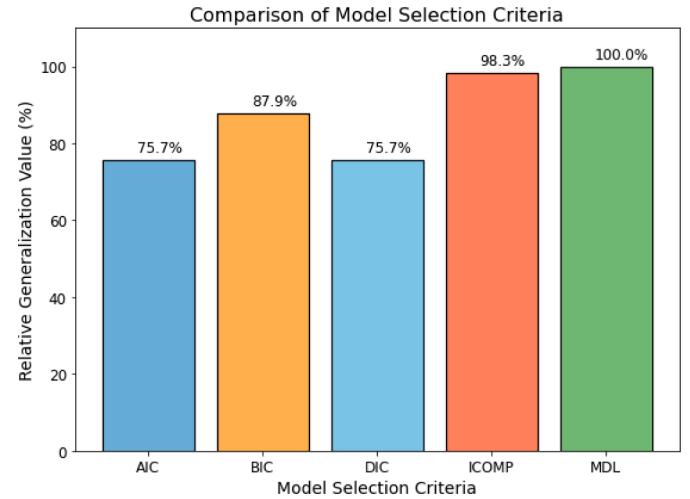


Fig. 13. Model selection criteria relative generalization values comparison.

Table 6

Comparison of model selection criteria with selected model type.

| Model selection criteria | Selected Model type | Relative Generalization Value (MDL as the benchmark) |
|---|--|---|
| Akaike information criterion (AIC) | <i>A</i> _cross, <i>E</i> , <i>G</i> , <i>I_y</i> _cross, <i>I_z</i> _cross, <i>K</i> _rod, <i>K</i> _friction, <i>K</i> _cc1, <i>K</i> _cc2, <i>K</i> _cc3, <i>A</i> _main, <i>I_y</i> _cross, <i>K</i> _mc | 75.7% |
| Bayesian information criterion (BIC) | <i>A</i> _cross, <i>E</i> , <i>G</i> , <i>I_y</i> _cross, <i>K</i> _rod, <i>K</i> _friction, <i>K</i> _cc1, <i>K</i> _cc2, <i>K</i> _cc3, <i>A</i> _main, <i>I_y</i> _cross, <i>K</i> _mc | 87.9% |
| Deviance information criterion (DIC) | <i>A</i> _cross, <i>E</i> , <i>G</i> , <i>I_y</i> _cross, <i>I_z</i> _cross, <i>K</i> _rod, <i>K</i> _friction, <i>K</i> _cc1, <i>K</i> _cc2, <i>K</i> _cc3, <i>A</i> _main, <i>I_y</i> _cross, <i>K</i> _mc | 75.7% |
| Information-theoretic measure of complexity (ICOMP) | <i>A</i> _cross, <i>E</i> , <i>G</i> , <i>K</i> _rod, <i>K</i> _friction, <i>K</i> _cc1, <i>K</i> _cc2, <i>A</i> _main, <i>I_y</i> _cross, <i>K</i> _mc | 98.3% |
| Minimum description length (MDL) | <i>A</i> _cross, <i>E</i> , <i>G</i> , <i>K</i> _rod, <i>K</i> _friction, <i>K</i> _cc1, <i>K</i> _cc2, <i>K</i> _cc3, <i>A</i> _main, <i>I_y</i> _cross, <i>K</i> _mc | 100% (Benchmark) |

Let $k = sim$, δ , and $R_k(D_k)$ is the correlation matrix with $\gamma_k(x, x') \in D_k$, $D_{obs}(\theta) := \{(x_i, \theta)\}_{i=1}^{N_{obs}}$ for $x_i \in D_{obs}$. Considering the prior distribution and the independence between $\eta(x, \theta)$ and $\delta(x)$, the posterior

density of the gaussian process with mean and variance can be written as:

$$\mu_{y_{obs}}(x^*, \theta) = t_{obs}(x^*, \theta)V_{obs}(\theta)^{-1}y \quad (18)$$

$$\sigma_{y_{obs}}^2(x^*, \theta) = \sigma_{obs}^2(x^*) - t_{obs}(x^*, \theta)V_{obs}^{-1}t(x^*, \theta) \quad (19)$$

$$\text{where, } t_{obs}(x^*, \theta) = \begin{bmatrix} \alpha\sigma_{sim}^2 R_{sim}((x^*, \theta), D_{sim}) \\ \alpha^2\sigma_{sim}^2 R_{sim}((x^*, \theta), D_{sim}) + \sigma_\delta^2 R_{sim}(x^*, D_{obs}) \end{bmatrix} \quad (20)$$

$$V(\theta) = \begin{bmatrix} V^{(sim, sim)} & V^{(sim, obs)}(\theta) \\ V^{(obs, sim)}(\theta) & V^{(obs, obs)}(\theta) \end{bmatrix}$$

and the above diagonal block matrices are given by

$$V^{(sim, sim)} = \sigma_{sim}^2(R_{sim}(D_{sim}) + \sigma_{e_{sim}}^2 I) \quad (21)$$

$$V^{(obs, obs)} = \sigma_{obs}^2(R_{obs}(D_{obs}) + \sigma_{e_{obs}}^2 I) + \sigma_{sim}^2\alpha^2(R_{sim}(D_{obs}(\theta)) + \sigma_{e_{sim}}^2 I) \quad (22)$$

The off-diagonal blocks are given by

$$V^{(sim, obs)}(\theta) = \alpha V^{(sim, sim)}(D_{sim}, D_{obs}(\theta)) \quad (23)$$

We utilized BBVI to make Bayesian inference of the inverse problem, as described in [Algorithm 1](#). The convergence of the ELBO iterations is

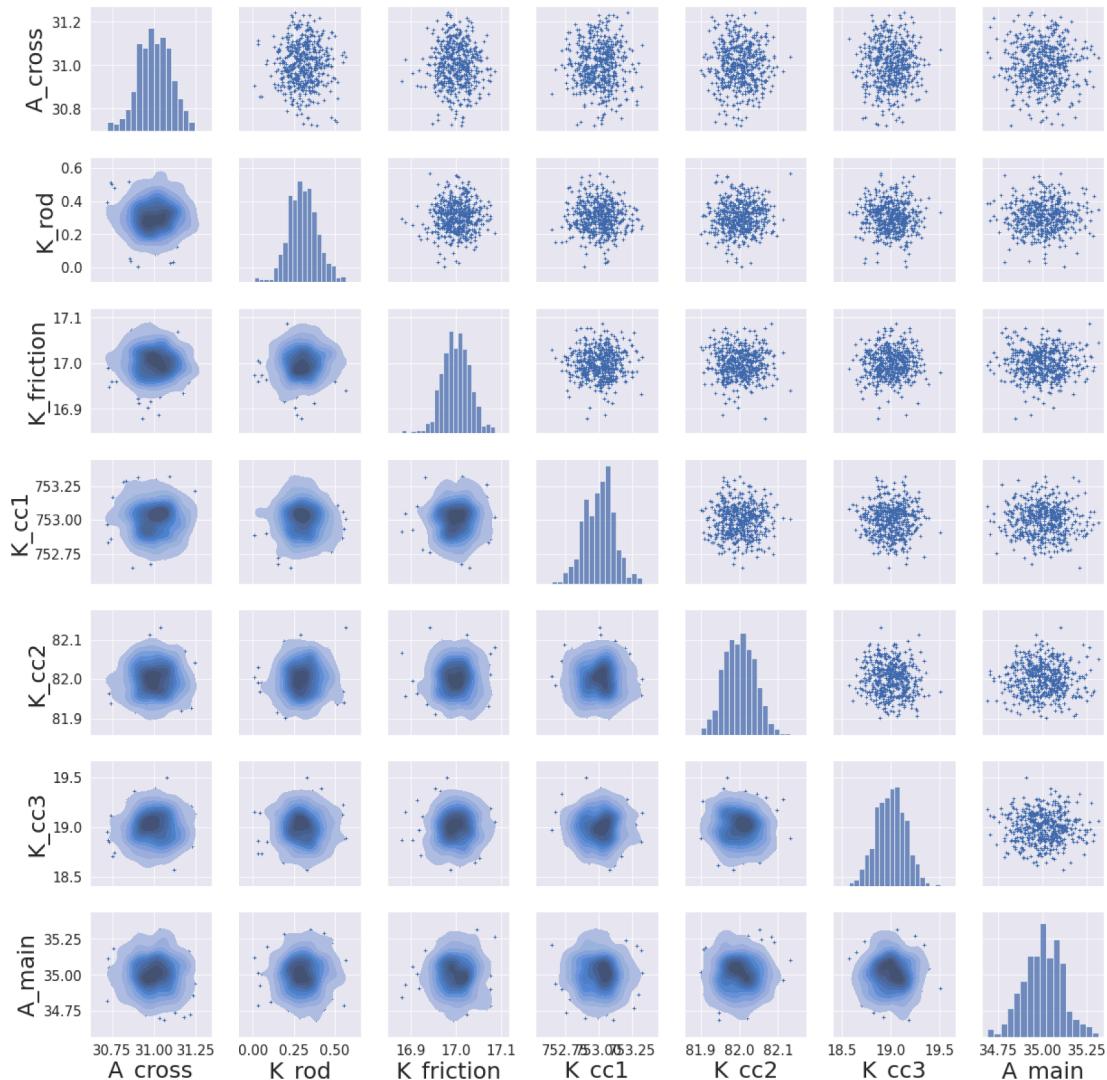


Fig. 14. Posterior distribution of seven of inverse problems parameters.

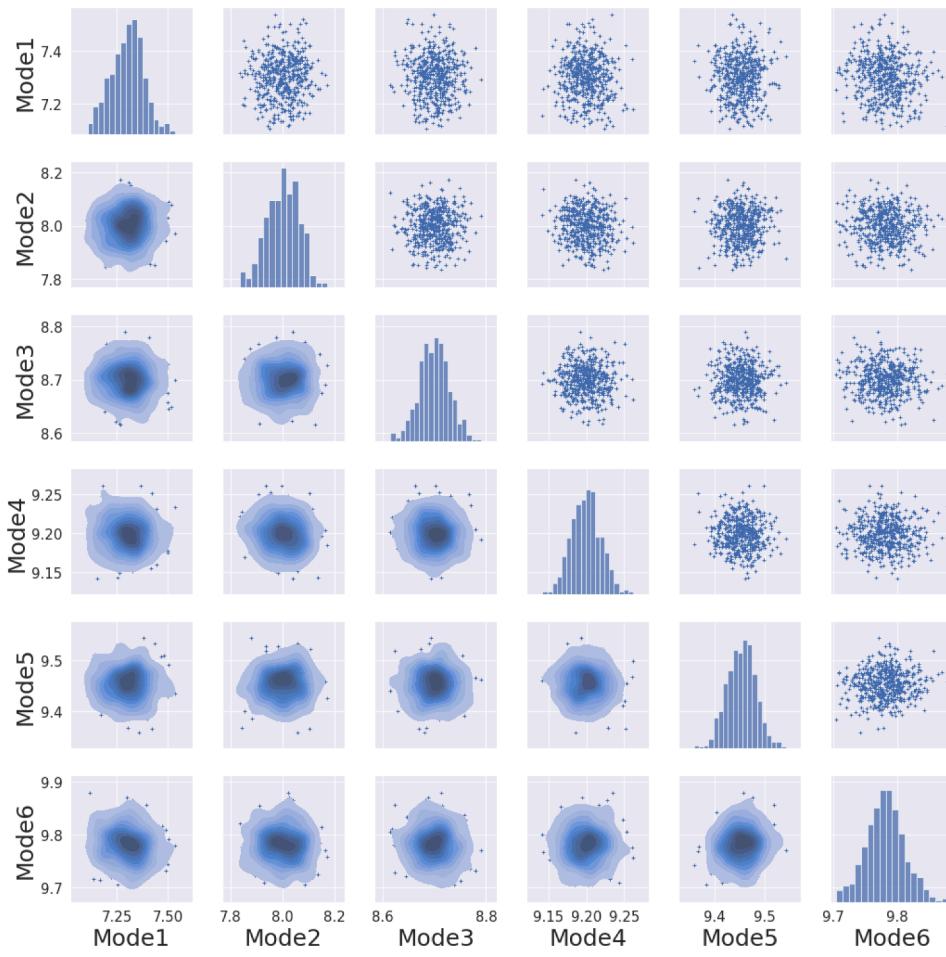


Fig. 15. Posterior distribution of the first six frequency modes of the suspended ceiling system.

shown in Fig. 12.

5.6. Minimum description length (MDL) model selection and comparison with other methods

As outlined in Section 3, we utilize MDL for model selection. Based on the blackbox variational inference results of the posterior and MDL equations, we calculate the MDL complexity values of seven groups of models using the following method (see Table 5).

After computing the MDL complexity values for each group of models, we select the model with the minimum MDL value. This model, with 11 parameters, provides the optimal result for the inverse problem, as indicated by the posterior distribution results in Section 5.7. Furthermore, we compare the MDL results with the other model selections below (see Table 6, Fig. 13).

The generalizability value is used to express the ability of the model calibration to predict the true model. This value is defined by the following Equation [23]:

$$Gr(\%) = \left\{ 1 - \frac{\sum_{i=1}^m |y(x_i, \theta) - y_{true}(x_i, \theta_n)|}{\sum_{i=1}^m |y_{true}(x_i, \theta_n)|} \right\} \times 100 \quad (24)$$

where m is the number of discretized points that are used to evaluate the generalizability, and $y_{true}(x_i, \theta_n)$ is the response of the actual model. This study uses MDL as the benchmark to compare other model selection criteria.

The comparison between the geometric complexity measure minimum description length (MDL) and other model selection criteria with the result in the next section shows its intuitive advantage in terms of

generalization performance applied in our SNS systems. AIC only considers the calibration parameters and will choose the more complex model, which leads to over-fitting, and DIC has a similar penalty for complexity to AIC. BIC has a more substantial penalty for complexity than AIC because it includes the factor of the natural logarithm of the sample size. ICOMP has almost the same generalizability but becomes less fitting compared with MDL because it considers calibration parameters, sample size, sensitivity, and interdependence.

In summary, the presented analysis demonstrates that MDL strikes a better balance between model complexity and generalization performance, resulting in superior predictive performance in SNS systems. The MDL approach considers calibration parameters, sample size, and geometric complexity information, resulting in a more comprehensive and well-rounded evaluation of model complexity, leading to optimal generalization performance.

5.7. Further analysis and discussion

While it is true that variational inference is an approximation of MCMC, it is still a powerful tool for improving the accuracy of predictions. While MCMC can take several hours to generate posterior estimates, variational inference can accomplish the same task in a few minutes, as noted in our analysis. The posterior distribution resembles a normal distribution due to using an approximate normal distribution as the variational family in blackbox variational inference combined with O'Hagan's calibration framework. This approach allows us to find a simple variational family to approximate the high-dimensional posterior distribution while maintaining good accuracy, making fewer derivations, and saving computational time. It has advantages over mean-field

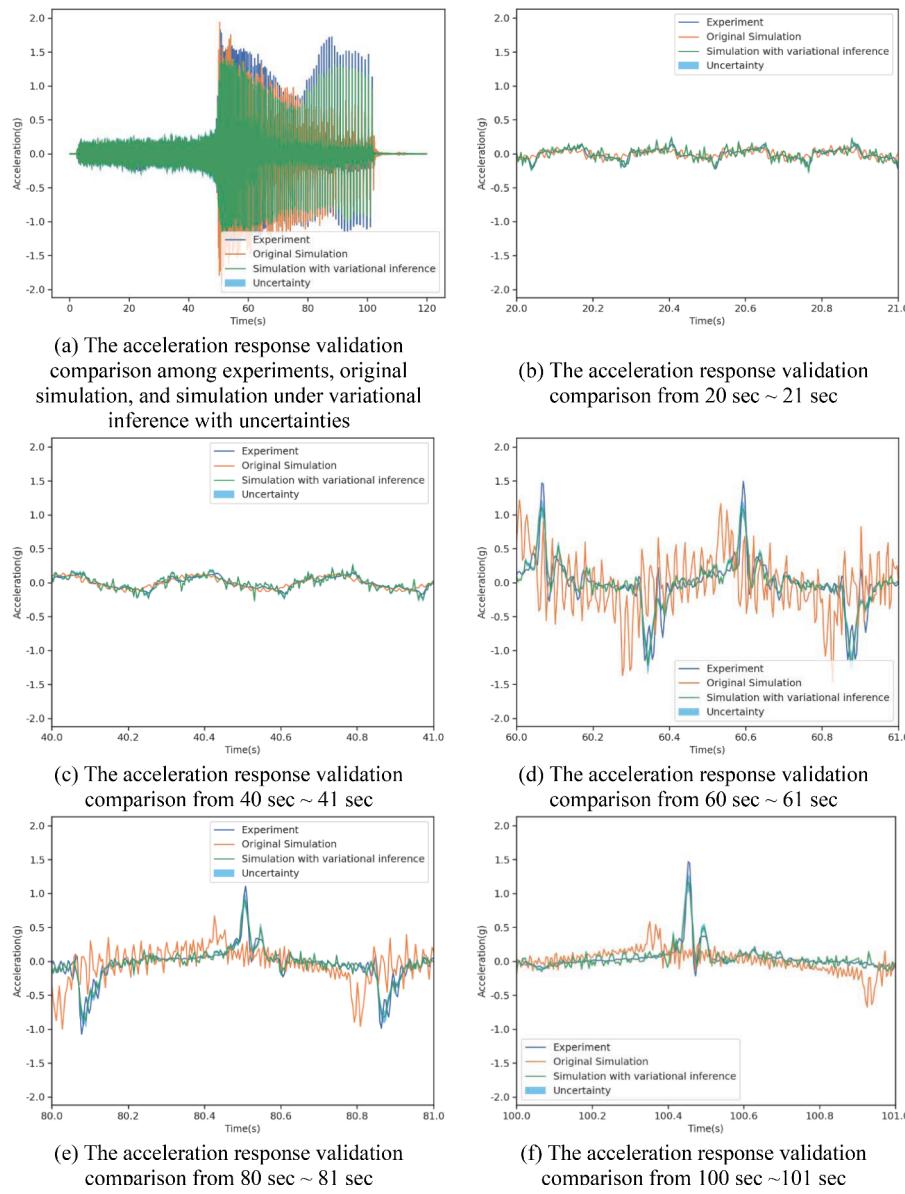


Fig. 16. Comparison between the measured and predicted responses.

variational inference, which involves lengthy derivations and conjugate distributions such as Gaussian mixture models. In addition, it avoids the complex burn-in process of the MCMC iteration and ensures accuracy. Our new approach, embedded with a geometric complexity measure, produces posterior estimates with low bias and variance, indicating that it can accurately estimate the parameters of the system with good generalization.

The above can be clearly seen in Figs. 14 and 15. For example, Fig. 14 shows the convergence of the posterior distribution for the simulated dataset, indicating that our approach is able to find good posteriors for the parameters. Additionally, Fig. 15 shows the posterior distribution of the first six frequency modes of the system, which are of particular interest in the context of SNS. We can see that the posterior mean is very close to the observation data mentioned above, indicating that our approach can produce accurate and robust estimates for SNS systems with uncertainty quantification.

Based on the results of the posterior distribution of the inverse problems under our new approach, we now move to verify and evaluate the seismic acceleration and displacement response of the tested SNS. Fig. 16 shows that the numerical simulation with variational inference

has a much better validation result than the original simulation in amplitude and trend. A validation improvement rate of approximately 50%–70% was achieved. As discussed in the [26], part of the original acceleration response was validated well, and some were not; for a clearer view, we uniformly selected five short time ranges to double check: 20 s ~ 21 s, 40 s ~ 41 s, 60 s ~ 61 s, 80 s ~ 81 s and 100 s ~ 101 s.

In addition, in Fig. 16, we compare the original simulations with our new approach. We observed that, except for simulation (f), all the original simulations had a similar trend to that of the experimental data. Additionally, except for simulations (d), (e), and (f), which is likely due to inaccurate parameters for the component connections and boundary friction properties resulting in larger discrepancies, all the original simulations had an amplitude that fit well with the experimental data. However, after applying our new approach, we observed that the trend of simulation (f) better fits with the experimental data, and the amplitude of simulations (d), (e), and (f) almost fitted well with the experimental data under the confidential uncertainty interval.

Also, the original displacement response can be seen to be validated reasonably well (with a minor time delay difference). Fig. 17 shows that the numerical simulation with variational inference has a much better

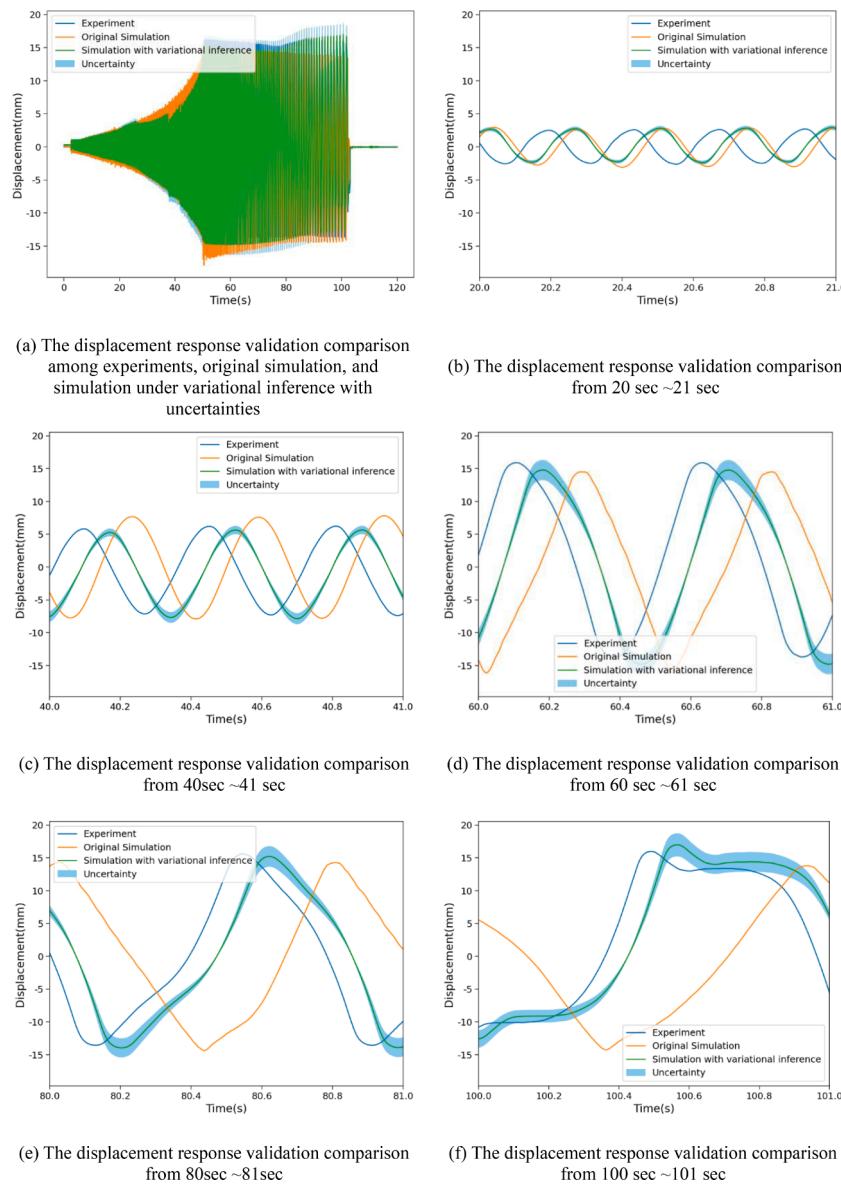


Fig. 17. Comparison between the measured and predicted responses.

displacement response validation result than the original simulation in terms of amplitude and trend with a confidential uncertainty interval. For a clearer view, a pick of uniformly select five time-range with a 1 s interval at 20 s ~ 21 s, 40 s ~ 41 s, 60 s ~ 61 s, 80 s ~ 81 s, and 100 s ~ 101 s. We can find that the validation improved especially the time delay difference result, and the overall improvement rate is around 50%.

In Fig. 17, we compared the original simulations with our new approach. We observed that almost all the original simulations fit well in amplitude with the experimental data, but they all exhibited some delay effect. After applying our new approach, we observed that the amplitude of simulations (c) and (d) were enhanced, and all the delay effects were reduced. The best performance improvement was observed in simulation (e), which exhibited an accumulation effect with the time duration that was enhanced by our new approach. Overall, our results demonstrate the effectiveness of our proposed approach for enhancing the accuracy of uncertainty quantification in SNS, and we were able to produce accurate and precise posterior estimates that provide valuable insights into the behavior of these systems.

6. Conclusions

This study presents a novel uncertainty quantification methodology of inverse problem based on variational inference with an efficient machine learning-based surrogate model for predicting the response of suspended nonstructural systems (SNS) in supertall buildings during long duration and long period seismic events. Our approach embeds geometric complexity measure Minimum Description Length (MDL) as a model selection criterion to strike a better balance between model complexity and generalization performance, resulting in superior predictive performance in SNS systems. The proposed optimization-based variational inference method significantly improves the low efficiency of traditional Markov Chain Monte Carlo (MCMC) and ensures high precision. When combined with ML-based Gaussian process surrogate models, the same method dramatically reduced the inference time of forward and inverse problems, enabling us to perform the largest full-scale SNS experiments in the world.

Our study also presents a unique combination of blackbox variational inference (BBVI) with O'Hagan's Bayesian calibration framework, achieving superior results for SNS systems in civil engineering. BBVI does not require specific model derivations and can be scaled well to

large datasets and high-dimensional parameter spaces, which are common in many fields. Moreover, our approach achieved good and robust Bayesian inference and uncertainty quantification results and improved the validation accuracy of SNS systems. Our comparison with other criteria demonstrates MDL's intuitive advantage of MDL in terms of optimal generalization performance.

In summary, our study significantly contributes to the SNS community by offering a comprehensive and well-rounded evaluation of model complexity and generalization performance and ensuring building resilience and safety against seismic risk events. Our findings suggest that our proposed methodology can present significant advantages over traditional methods in terms of computational efficiency and accuracy and can be extended to many other fields dealing with large-scale, high-dimensional datasets. We plan, as well as invite our readers, to investigate additional inverse problems for uncertainty quantification in the nonlinearities of the SNS system in future works – possibly by replacing the current machine learning-based Gaussian process model with a physics-informed deep learning model that incorporates prior physical knowledge to improve accuracy and explanatory power, particularly in situations with limited data [53,55,56].

CRediT authorship contribution statement

Zhiyuan Qin: Conceptualization, Methodology, Writing – original draft, Software. **M.Z. Naser:** Conceptualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgments

First, the authors would like to thank the reviewers for their careful reading, questions, and comments. The authors acknowledge financial support from the National Natural Science Foundation of China (Grant No. U2239253) and International Joint Research Laboratory of Earthquake Engineering of Tongji University (Grant No. 0200121005/058).

References

- [1] Tabandeh Armin, Sharma Neetesh, Gardoni Paolo. Uncertainty propagation in risk and resilience analysis of hierarchical systems. *Reliab Eng Syst Saf* 2022;219: 108208. <https://doi.org/10.1016/j.ress.2021.108208>. ISSN 0951-8320.
- [2] Oberkampf WL, Roy CJ. Verification and validation in scientific computing. Cambridge: Cambridge University Press; 2010.
- [3] Kiureghian AD, Ditlevsen O. Aleatory or epistemic? Does it matter? *Struct Saf* 2009;31:105–12.
- [4] Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Stat Sci* 1989;4:409–23. <https://doi.org/10.1214/ss/1177012413>.
- [5] Guo Hongyuan, You Dong, Gardoni Paolo. Adaptive subset simulation for time-dependent small failure probability incorporating first failure time and single-loop surrogate model. *Struct Saf* 2023;102:102327. <https://doi.org/10.1016/j.strusafe.2023.102327>. VolumeISSN 0167-4730.
- [6] Pan Yongjun, Sun Yu, Li Zhixiong, Gardoni Paolo. Machine learning approaches to estimate suspension parameters for performance degradation assessment using accurate dynamic simulations. *Reliab Eng Syst Saf* 2023;230:108950. <https://doi.org/10.1016/j.ress.2022.108950>. VolumeISSN 0951-8320.
- [7] Guo Hongyuan, Dong You, Gardoni Paolo. Efficient subset simulation for rare-event integrating point-evolution kernel density and adaptive polynomial chaos kriging. *Mech Syst Signal Process* 2022;169:108762. <https://doi.org/10.1016/j.ymssp.2021.108762>. VolumeISSN 0888-3270.
- [8] Kennedy MC, O'Hagan A. Bayesian calibration of computer models. *J Roy Stat Soc Ser B* 2001;63:425–64. <https://doi.org/10.1111/1467-9868.00294>.
- [9] Burt, D.R., Rasmussen, C.E., and van der Wilk, M.: Rates of convergence for sparse variational gaussian process regression, arXiv [preprint], arXiv:1903.03571, 2019.
- [10] Shang Xiaobing, Su Li, Fang Hai, Zeng Bowen, Zhang Zhi. An efficient multi-fidelity Kriging surrogate model-based method for global sensitivity analysis. *Reliab Eng Syst Saf* 2023;229:108858. <https://doi.org/10.1016/j.ress.2022.108858>. VolumeISSN 0951-8320.
- [11] Hastings WK. Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* 1970;57(1):97–109.
- [12] Hinton G E, van Camp D. keeping neural networks simple by minimising the description length of weights. In: Proceedings of COLT-93; 1993. p. 5–13.
- [13] Nathoo F, Babul A, Moiseev A, Virji-Babul N, Beg M. A variational Bayes spatiotemporal model for electromagnetic brain mapping. *Biometrics* 2014;70(1): 132–43.
- [14] Ni Y, Li J, Chatzi E. Probabilistic model updating via variational Bayesian inference and adaptive Gaussian process modeling. *Comput Methods Appl Mech Eng* 2021; 383:114915.
- [15] Damm MR, Kirkland M. Structural deterioration modeling using variational inference. *J Comput Civil Eng* 2019;33(1):04018051.
- [16] Ranganath R, Gerrish S, Blei D. Blackbox variational inference. *Artif Intell Stat* 2014:814–22.
- [17] Higdon D, Gattiker J, Williams B, Rightley M. Computer model calibration using high-dimensional output. *J Am Stat Assoc* 2008;103:570–83.
- [18] Bard Y. Nonlinear parameter estimation. New York/London: Academic Press; 1974.
- [19] Nguyen Khanh TP, Fouladridar Mitra, Grall Antoine. Model selection for degradation modeling and prognosis with health monitoring data. *Reliab Eng Syst Saf* 2018;169:105–16. <https://doi.org/10.1016/j.ress.2017.08.004>. VolumePagesISSN 0951-8320.
- [20] MacKay DJC. Bayesian Interpolation. *Neural Comput* 1992;4:415–47.
- [21] Cutting JE, Bruno N, Brady NP, Moore C. Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. *J Exp Psychol: General* 1992;121(3):364–81.
- [22] Rissanen J. Modeling by shortest data description. *Automatica* 1978;14(5):465–71.
- [23] Pitt M, Myung I, Zhang S. Toward a method of selecting among computational models of cognition. *Psychol Rev* 2002;109(3):472–91.
- [24] Myung I, Balasubramanian V, Pitt M. Counting probability distributions: differential geometry and model selection. *Proc Natl Acad Sci* 2000;97(21): 11170–5.
- [25] Myung IJ, Pitt MA, Kim W. Model evaluation, testing and selection. In: Lamberts K, Goldstone R, editors. *Handbook of cognition*. London, UK: Sage Publications; 2005.
- [26] Wang Yong, Jiang Huanjun, Motoyui Shojiro, Kasai Kazuhiko, Qin Zhiyuan, Huang Youlu. Study on seismic performance of suspended ceiling system with semi-free boundary condition. *Eng Struct* 2023;275(Part A):115208. ISSN 0141-0296.
- [27] Sullivan TJ. Introduction to uncertainty quantification [M]. Springer; 2015.
- [28] Chantrasmí T, Iaccarino G. Forward and backward uncertainty propagation for discontinuous system response using the Pade-Legendre method. *Int J Uncertain Quantif* 2012;(2):2.
- [29] Tikhonov A N. On the stability of inv. Prob. Dokl. Akad. Nauk SSSR. 1943:195–8.
- [30] Tarantola A. Inverse problem theory and methods for model para. estimation. siam 2005.
- [31] Engl H W, Hanke M, Neubauer A. Regularization of inverse problems. Springer Science & Business Media; 1996.
- [32] Cotter S L, Dashti M, Stuart A M. Approximation of Bayesian inverse problems for PDEs. *SIAM J Numer Anal* 2010;48(1):322–45.
- [33] Dashti M, Harris S, Stuart A. Besov priors for Bayesian inverse problems. arXiv preprint arXiv:1105.0889. 2011.
- [34] Lasanen S. Non-Gaussian statistical inverse problems, Part I: posterior distributions [J/OL]. *Inverse Probl Imag* 2012;6(2):215–66.
- [35] Jia J, Peng J, Gao J. Bayesian approach to inverse problems for functions with avariable-index Besov prior. *Inverse Probl* 2016;32(8):085006 (32pp).
- [36] Kypraios T, Neal P, Prangle D. A tutorial introduction to Bayesian inference for stochastic epidemic models using approximate Bayesian computation. *Math Biosci* 2017;287:42–53.
- [37] Chen P, Villa U, Ghattas O. Hessian-based adaptive sparse quadrature for infinite dimensional Bayesian inverse problems. *Comput Methods Appl Mech Eng* 2017.
- [38] Sun Z, Wang J, Li R, et al. LIF: a new Kriging based learning function and its application to structural reliability analysis. *Reliab Eng Syst Saf* 2017;157:152–65.
- [39] Zimmer C, Meister M, Nguyen-Tuong D. Safe active learning for time-series modeling with gaussian processes. In: *Advances in Neural Infor. Processing Systems*. 31; 2018. p. 2730–9.
- [40] Snoek J, Larochelle H, Adams R P. Practical bayesian optimization of machine learning algorithms [C]. *Advances in neural information processing systems*. 2012. p. 2951–9.
- [41] Cutajar K, Bonilla E V, Michiardi P, et al. Practical learning of deep gaussian processes via random fourier features. 2016.
- [42] Salimbeni H, Deisenroth M. Doubly stochastic variational infer. for Deep GPs. 2017.
- [43] Willcox KE, Ghattas O, Heimbach P. The imperative of physics-based modeling and inverse theory in computational science. *Nat Comput Sci* 2021;1:166–8. <https://doi.org/10.1038/s43588-021-00040-z>.
- [44] Xu Yanwen, Kohtz Sara, Boakye Jessica, Gardoni Paolo, Wan Pingfeng. Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges. *Reliab Eng Syst Saf* 2023;230:108900. <https://doi.org/10.1016/j.ress.2022.108900>. VolumeISSN 0951-8320.
- [45] Kennedy MC, O'Hagan A. Bayesian calibration of computer models. *J R Stat Soc Ser B (Stat Methodol)* 2001;63:425–64. <https://doi.org/10.1111/1467-9868.00294>.

- [46] Blei D M, Kucukelbir A, McAuliffe J D. Variational inference: a review for statisticians. *J Am Stat Assoc* 2017;112(518):859–77.
- [47] Helton J C, Davis F J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab Eng Syst Saf* 2003;81(1):23–69.
- [48] Schöbi R, Sudret B, Marelli S. Rare event estimation using polynomial-Chaos Kriging. *ASCE-ASME J Risk Uncertain Eng Syst, Part A: Civil Eng* 2016;D4016002.
- [49] Oladyshkin S, Nowak W. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. *Reliab Eng Syst Saf* 2012;106:179–90.
- [50] SOBER E. [1975]: simplicity. Oxford: Oxford University Press; 1975.
- [51] Hansen MH, Yu B. Model selection and the principle of minimum description length. *J Am Stat Assoc* 2001;96(454):746–74.
- [52] Google and DeepMind research team. Why neural networks find simple solutions: the many regularizers of geometric complexity. In: 36th Conference on Neural Information Processing Systems. NeurIPS; 2022. 2022.
- [53] Rissanen J. Strong optimality of the normalized ML models as universal codes and information in data. *IEEE Trans Inf Theory* 2001;47(5):1712–7.
- [54] Xilin Lv. Shaking table model test methods and techniques for building structures. 2nd Edition. Beijing: Science Press; 2016 [M](In Chinese).
- [55] Tapeh A, Naser MZ. Artificial intelligence, machine learning, and deep learning in structural engineering: a scientometrics review of trends and best practices. *Arch Comput Method Eng* 2022. <https://doi.org/10.1007/s11831-022-09793-w>.
- [56] Naser MZ. Mapping functions: a physics-guided, data-driven and algorithm-agnostic machine learning approach to discover descriptive expressions of engineering phenomena. *Measurement* 2021. <https://doi.org/10.1016/j.measurement.2021.110098>.