Machine Learning and Model Driven Bayesian Uncertainty Quantification in Suspended Nonstructural Systems

Zhiyuan Qin, M.Z. Naser
School of Civil & Environmental Engineering and Earth Sciences (SCEEES), Clemson University, USA
Artificial Intelligence Research Institute for Science and Engineering (AIRISE), Clemson University, USA
E-mail: mznaser@clemson.edu, Website: www.mznaser.com

Abstract
This paper presents a novel framework for uncertainty quantification of inverse problems often faced in suspended nonstructural systems. This framework adopts a machine learning- and model-driven stochastic Gaussian process model calibration to quantify the uncertainty via a new black box variational inference that accounts for geometric complexity through Bayesian inference. The validity of the proposed framework is validated through one of the largest full-scale shaking table tests of suspended nonstructural systems to date, as well as accompanying simulated (numerical) data. Our findings indicate that the proposed framework is not only computationally sound and scalable but also yields optimal generalizability.

Keywords: Inverse problems; Machine learning; Gaussian process; Black box variational inference; Geometric Complexity; Suspended nonstructural system.

1.0 Introduction

In order to obtain precious predictions, it is necessary to comprehensively consider various degrees of uncertainties [1], such as those arising from measurement uncertainty, model solution error model parameters/selection, etc. [2]. Generally, there are two kinds of uncertainty: *aleatoric* and *epistemic* [3]. Aleatoric uncertainty, also known as 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 data sets or biased models, etc.

The complexity of uncertainty quantification models tends to increase with the availability of computational resources. While this increased complexity can lead to the creation of more accurate models, their associated costs may not be easily justifiable over the more affordable approximations (or surrogates) [4, 5].

With the rise of machine learning (ML) [6], surrogates have become attractive methods [7]. For example, the Gaussian Processes (GPs) have gained popularity in the supervised ML community. The use of GPs for calibrating computer models was first introduced by Kennedy and O'Hagan [8] and continues to serve as the foundation for contemporary methods. GPs are especially well suited for this purpose since they offer reliable estimates of uncertainty in nonlinear systems, even with little training data. Recent developments have enabled deeper and more expressive GPs [9, 10], despite initial challenges with their scalability as compared to, for example, neural networks.

Derivative-free Bayesian calibration or inversion commonly starts with observing the error 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]. Given that each forward run can be expensive, conducting a series of runs is computationally unaffordable, rendering MCMC impractical for real-world calibrations.

To overcome such a challenge, Peterson and Hinton [12] studied Variational Inference (VI) methods, mainly to approximate 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 processes rather than sampling processes. The variational distribution is a collection of hidden variable distributions that adopts KL-divergence (Kullback-Leibler) as a measure of inter-distribution similarity and finds the variational distribution most close to the true posterior distribution. Compared with the Monte Carlo sampling method, the variational inference method is faster, simpler, and easier to parallelize and is more suitable for big data and complex models, as noted in recent works [13].

Variational inference (VI) is a popular technique in machine learning, but it is not as widely used in statistics compared to MCMC-based sampling techniques. In civil engineering, the slow uptake of VI can be attributed to its additional modeling complexities and limited theoretical exploration. Although traditional mean field VI is commonly used [14, 15], it requires complex mathematical derivations and conjugate assumptions, which limits its practical applications. On the other hand, Black box variational inference (BBVI) [16] is a promising and advanced VI technique that remains largely unexplored in civil engineering. BBVI does not require specific model derivations and can scale well to large datasets and high-dimensional parameter spaces, which are common in many fields. In contrast, MCMC related methods become quickly impractical with the increasing size of datasets and number of parameters and do not scale well. Therefore, we propose a combination of BBVI with O'Hagan's Bayesian calibration framework [8], which can be easily derived without the need for conjugate assumptions, enabling the use of BBVI and achieving superior results for SNS systems in civil engineering, which are critical for ensuring reliability and safety and making it accessible to a wider audience of engineers and scientists.

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

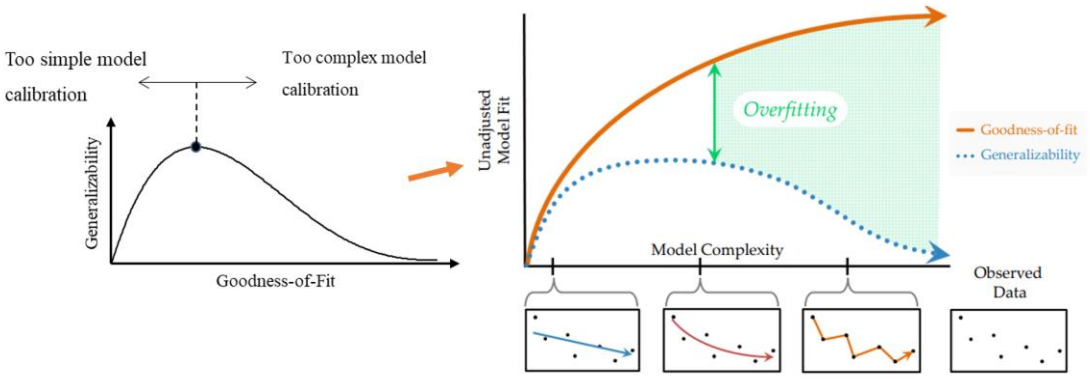
Model calibration involves estimating the best-fit values for a few identifiable calibration parameters from experiments conducted at various control parameter settings. However, models are inherently incomplete (i.e., systematically biased) and approximate, and cannot provide a complete representation of the true system behavior. Such incompleteness may originate from a

variety of fronts, such as the omission of input parameters, interactions between the model input parameters and/or control variables or assigning incorrect values to model input parameters considered to be known [8].

Thus, inevitably exhibit systematic discrepancy in predicting the true behavior of a system. However, this inherent discrepancy can be identified during model calibration by inferring an independent error model from the experimental data [4, 17] or by blending emulators with physics-based models to explain the omitted relationships between model input parameters.

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

The complexity of a model calibration campaign results in an Ockham's hill relationship between good fitness to a finite number of noisy measurements in the tested settings and the generalizability of the model predictions in the untested settings [20] (See Figure 1(a)). A model calibration campaign that lacks complexity would lose valuable information that could have otherwise been inferred from the data. On the other hand, a highly complex model is likely to fit noisy measurements, seemingly improving the goodness of fit while degrading the generalizability.



(a) Ockham's hill relationship (b) Detailed example

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

In an extreme scenario, a calibration campaign that produces a model capable of perfectly matching all possible outcomes (i.e., infinite flexibility) will yield an uninformative tool that is

impossible to falsify and one with low 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 this hypothetical infinite flexibility. This differential ability of a model to fit data is referred to as "selectivity" by Cutting et al. [21].

In Figure 1(b), we present a detailed illustration of the interplay among the model goodness-of-fit, generalizability, and model complexity. Cutting et al. [21] recognized that the number of parameters alone is an insufficient indicator of model complexity and advocated for evaluating the fitting power or scope of a model to random data. advocate for evaluating a 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 is compressed, the more information about the underlying regularities governing the process of interest would be learned [24]. This metric considers the experimental data as a code or description to be compressed by the model. MDL also evaluates models according to their ability to compress a data set by extracting the necessary information from the data without random noise. Therefore, MDL enables the selection of a model with the shortest description code (length) of the data [25]. In Bayesian inference problems, most studies overlook generalization and model complexity or use simple criteria like AIC, BIC, or 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 data space. Our comparison with other criteria demonstrates MDL's superior generalization performance and it is also well-embedded in the black box variational inference framework, and our approach achieves good Bayesian inference and Uncertainty Quantification results and improves validation accuracy of suspended nonstructural systems (SNS) systems.

In recent years, moderate or strong earthquakes have caused significant property loss, interruption of building function, and even threatened life safety due to damage to suspended nonstructural systems (SNS) [26]. Despite minor damage to the main building structures, the impact on SNS underscores their crucial role in ensuring the resilience of buildings against seismic events. While some experiments have been conducted in recent years, the effects of ultra-large areas and long duration and long periods under the conditions of super-tall buildings are still unknown. This recently completed full-scale suspended nonstructural systems (SNS) experiments, including suspended ceilings and cable trays, are the largest in the world, in which we carefully designed earthquake wave inputs in line with long duration and long periods in super-tall buildings, based on national standards and random vibration theory. In the field of civil engineering, Uncertainty Quantification and Inverse problem inference for the suspended nonstructural systems (SNS) is still unknown, and this study will fill that critical gap. Our use of a surrogate machine learning model reduces computational cost and running time, resulting in a great speed increase.

From this view, this paper presents a novel framework for uncertainty quantification of inverse problems with the application on the suspended nonstructural systems. The validity of the proposed framework is using the full-scale shaking table tests of suspended nonstructural systems (SNS) and accompanying simulated data.

2.0 **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 a range of fields, including probability and statistics, functional analysis, and more, resulting in the emergence of various mathematical, statistical, Bayesian, optimization, and approximation techniques [22]. To better understand the challenges of UQ, Stark et al. distinguished between the propagation of uncertainty (PoU) and its inverse, as shown in Figure 2 [27].

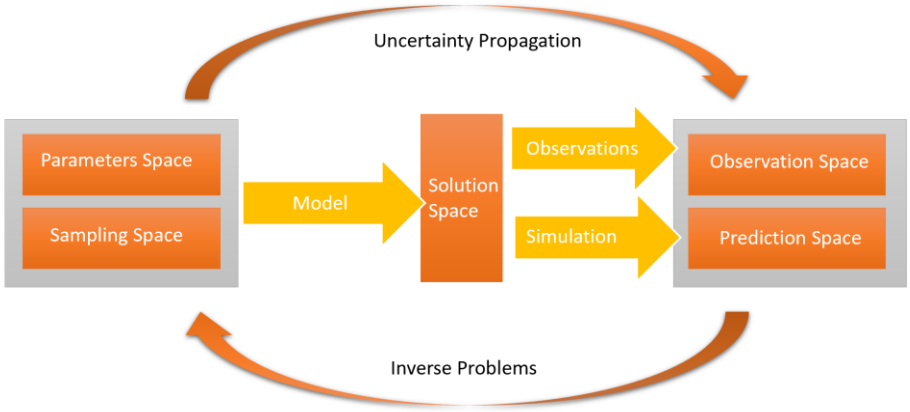


Figure 2 Classification of the main problems of uncertainty quantification

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]. It 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 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 unsolved problems. Tarantola [30] offered a Bayesian explanation of inverse problem theory and general methods for model parameter estimation, while 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 the wavelet function in L^2 space to address infinite-dimensional nonparametric regression problems.

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 are discussed in [35]. Maximum posterior estimation is widely used for extracting information from the posterior distribution, as 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. Various methods are available, with thorough analyses of Bayesian computation techniques for parametric models [36] and infinite-dimensional spaces [37] in the literature.

2.3 Surrogate models

A surrogate model provides an approximation of 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 [38].

Gaussian processes (GPs) are nonparametric Bayesian models of interest to this work [39]. According to Rasmussen and Williams [40], any finite sample of the function follows a multidimensional Gaussian distribution, with the QoI function viewed as a specific sample of a Gaussian process. The Gaussian process is a probability distribution of the response function, which makes 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 the Bayesian information criterion (BIC) criteria, with posterior distributions of hyperparameters often sampled using MCMC methods in applications sensitive to parameter uncertainty.

Gaussian Processes (GPs) have been widely used in many fields, providing 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 a single-layer GP model. Recent advances in GP research have applied them to sparse variational inference, enabling efficient computation for large datasets. The induction point approach combined with expectation propagation has 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 further improve model accuracy [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 the implicit function variables. These advancements have made deep GPs a powerful tool for modeling complex and high-dimensional systems.

3.0 Framework for ML & model driven uncertainty quantification of inverse problems

3.1 Gaussian process surrogate model

3.1.1 Gaussian process surrogate model

We define the expectation and covariance of a real-valued GP *f(x), m(x)* and *k (x, x')*:

$$\mathrm{m(x)} = \mathbb{E}[f(x)] \tag{1}$$

$$\mathrm{k}(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x'))] \tag{2}$$

The GP can be expressed as:

$$f(x) \sim GP(m(x) mk(x, x')) \tag{3}$$

The size of the covariance between any two random variables in a GP can be calculated using the covariance function *k(r)*, where *r = d(x, x')* is the distance between the two inputs. The covariance (kernel) function quantifies the correlation between random variables and can be specified using various options, including the squared exponential covariance function:

$$K(r) = \alpha \exp\left(-\frac{\gamma^2}{2l^2}\right) \tag{4}$$

where α, γ, and l are hyperparameters of kernel functions, especially l is a scale parameter that describes how complex the kernel function is in the parameter space, and γ is defined as a measure of distance in Euclidean space[1]. Suppose that the input points$\mathrm{X}_* \in \mathbb{R}^{\mathrm{n*d}}$ of *n* dimensions are known, and the corresponding random variable $f_* \in \mathbb{R}^{\mathrm{n}}$ can be represented as an n-dimensional Gaussian distribution with an expectation of 0 and a covariance matrix of *K*:

$$f_* \sim N(0, K(X_*, X_*)) \tag{5}$$

where, each element in *K* can be calculated using equation (4). Assuming that each observation is mapped using an unknown function *f(x)* with independent Gaussian noise in equation (6) and the correlation of the observations of the inputs x_p, x_q is in equation (7):

$$\mathrm{y_i} = f(x_i) + \epsilon_i \tag{6}$$

$$cov(y_p, y_q) = k(x_p, x_q) + \sigma^2 \delta_{pq} \quad cov(y) = K(X, X) + \sigma_n^2 I \tag{7}$$

where δ_{pq} is the Dirac function when $p = q$, $\delta_{\mathrm{pq}} = 1$, otherwise δ_{pq} is 0. The joint distribution of y and f_* is shown in equation (8), and the conditional distribution is shown in (9):

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right) \tag{8}$$

$$f_* | X, y, X_* \sim n(\overline{f_*}, cov(\overline{f_*})) \tag{9}$$

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

where,

$$\overline{f_*} \triangleq \mathbb{E}[f_*|X, y, X_*] = k_*^T(K + \sigma_n^2 I)^{-1} y$$

$$cov(f_*) = \mathbb{V}[f_*] = k(x, x_*) - k_*^T(K + \sigma_n^2 I)^{-1} k_*$$

Then, we can get marginal likelihood $P(y|X)$directly from the prior y∼N$(0, K + \sigma_n^2 I)$:

$$logp(y|X) = -\frac{1}{2} y^T (K + \sigma_n^2 I)^{-1} y - \frac{1}{2}\log|K + \sigma_n^2 I| - \frac{n}{2} log 2\pi \quad (10)$$

3.1.2 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 function:

$$k_y(x_p, x_q) = \sigma_f^2 \exp\left(-\frac{1}{2l^2}(x_p - x_q)^2\right) + \sigma_n^2 \delta_{pq} \quad (11)$$

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 equation (10), we can attain the logarithmic form of $p\ (y|X, \theta)$ with hyperparameters represented by the vector θ explicitly:

$$logp(y|X, \theta) = -\frac{1}{2} y^T K_y^{-1} y - \frac{1}{2}\log|K_y| - \frac{n}{2}\log(2\pi) \quad (12)$$

where, $K_y = K_f + \sigma^2 I$ is the covariance matrix containing the covariance function and the observed error, $-\frac{1}{2} y^T K_y^{-1} y$ is the only data fitting term that contains observations; $\frac{1}{2}\log|K_y|$ is a calculated penalty that only considers covariance information; $\frac{n}{2}\log(2\pi)$ is a normalized constant. To maximize the marginal likelihood, its partial derivative form for each hyperparameter is:

$$\frac{\partial}{\partial\theta_j} logp(y|X, \theta) = \frac{1}{2} y^T K^{-1} \frac{\partial K}{\partial\theta_j} K^{-1} y - \frac{1}{2} tr\left(K^{-1}\frac{\partial K}{\partial\theta_j}\right) = \frac{1}{2} tr((\alpha\alpha^T - K^{-1})\frac{\partial K}{\partial\theta_j}) \quad (13)$$

where, $tr(\cdot)$ is the trace of the matrix, $\alpha = K^{-1}y$, K^{-1} determines the computational complexity of (13), Once K^{-1} is obtained, it has a gradient complexity of $O(n^2)$ for hyperparameter θ_j.

3.2 Bayesian inference

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

$$\Pr(\boldsymbol{\Theta}|\boldsymbol{D}, H) = \frac{\Pr(\boldsymbol{D}|\boldsymbol{\Theta}, \mathrm{H})\Pr(\boldsymbol{\Theta}|H)}{\Pr(\boldsymbol{D}|\mathrm{H})} \quad (14)$$

$$\text{and } \Pr(\boldsymbol{D}|H) = \int \Pr(\boldsymbol{D}|\boldsymbol{\Theta}, \mathrm{H})\Pr(\boldsymbol{\Theta}|H)\, d\boldsymbol{\Theta} \quad (15)$$

where $\boldsymbol{\Theta}$ 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 $\boldsymbol{\Theta}$. H represents the model or hypothesis which is believed to best represent the available $\boldsymbol{D}$. The terms expressed in equation (14) are as such: $Pr(\boldsymbol{\Theta}|H) \equiv \pi(\Theta)$ is the prior distribution which describes our prior knowledge of θ before any available D; $Pr(\boldsymbol{D}|\boldsymbol{\Theta}, H) \equiv L(\Theta)$ is the likelihood function which represents the degree of similarities between D and H; $Pr(\boldsymbol{\Theta}|\boldsymbol{D}, H) \equiv P(\Theta)$ is the posterior probability distribution of the parameters, which describes our updated information of θ after the information gained by D; $Pr(\boldsymbol{D}|H) \equiv Z$ is the Bayesian evidence which serves as the normalizing constant of the posterior. In the estimation process, the Bayesian evidence factor can be ignored since it is independent of the parameters Θ, and inferences are obtained by computing or sampling the unnormalized posterior (refer to [43] for full details):

$$P(\theta|D,H) \propto Pr(D|\theta,H) \cdot Pr(\theta|H) \tag{16}$$

3.3 Bayesian calibration

A statistically descriptive model calibration procedure [43] uses Bayesian inference to model the relationship between computer simulations and observed data y, while considering parameter uncertainties, model discrepancy, and observation error.

$$y(x) = \eta(x, t^*) + \delta(x) + \epsilon \tag{17}$$

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.

3.4 Stochastic sampling methods

In highly 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 $Pr(\Theta|\boldsymbol{D}, H)$ 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 KL divergence monotonically decreases during Markov transition, according to standard theory [44]:

$$D_{KL}[q_t||p] \leq D_{KL}[q_{t-1}||p] \tag{18}$$

where, q_t converges to the stationary distribution p as $t \rightarrow \infty$ under proper conditions..

3.5 Variational inference

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 Kullback-Leibler 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 stochastic sampling methods, 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.

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

$$\mathrm{D_{KL}}(q(z)||p(z)) = -\int q(z) \log \frac{p(z)}{q(z)} dz \tag{19}$$

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):

$$\begin{aligned} log\, p(x) &= log\, \mathbb{E}_{q(z;\lambda)} \left[\frac{p(x,\lambda)}{q(z;\lambda|x)} \right] \\ &\geq \mathbb{E}_{q(z;\lambda)}\, log \left[\frac{p(x,z)}{q(z;\lambda|x)} \right] = ELBO(q) \text{ by Jensen's inequality} \end{aligned} \tag{20}$$

The distance measured between its logarithmic marginal distribution and the ELBO is the KL divergence distance between variational distributions:

$$\begin{aligned} log\, p(x) &= \int q(z;\lambda)\, log \left(\frac{p(z,\ x)}{q(z;\lambda)} \right) dz - \int q(z;\lambda)\, log \left(\frac{p(z|x)}{q(z;\lambda)} \right) dz \\ &= ELBO(q) + D_{kl}(q(z;\lambda)||p(x,z)) \end{aligned} \tag{21}$$

Maximizing ELBO is equivalent to minimizing $D_{KL}(q(z|x)||p(z))$, 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 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 [44].

3.5.3 Non-conjugate (Black box) 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 Black Box Variational Inference (BBVI) [16] is proposed, allowing unbiased estimation of the variational parameter gradient by sampling from the variational distribution. This method does not constrain the probability distribution of intermediate variables and is a more general inference method. ELBO is maximized through updates based on the gradient or stochastic gradient of variational parameters, expressed as the expectation of the variational distribution for general models:

$$\nabla_\lambda L = \mathbb{E}_q[\nabla_\lambda log q(z|\lambda)(\log p(x,z) - log q(z|\lambda))] \quad (22)$$

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

$$\nabla_\lambda \hat{L}_{stoch} = \frac{1}{K} \sum_{k=1}^{K} \nabla_\lambda log q(z_k|\lambda)(\log p(x,z_k) - log q(z_k|\lambda)) \quad (23)$$

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 can be divided into model unknown, and model known designs [45]. Monte Carlo (MC) method is widely used for unknown models due to its *O (1/N$^{(1/2)}$)* mean convergence speed, which is independent of the model and dimension of input parameter variables, and its ability to provide stable solutions for unconventional distributions and extremely nonlinear and discontinuous models [46]. However, the MC method suffers from inefficient random sampling from the experimental space, which can be improved by targeted division of the test space and selecting representative sample points within each division. The random configuration method, based on the tensor product of high-precision interpolation points, has been introduced to address this issue, but it is limited by the "curse of dimensionality". Sparse nodes, such as the sparse Gauss-Hermite criterion, have been introduced to address this issue in high-dimensional numerical integrals in UQ experimental design.

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 $\{\mathrm{c_i}\}_{\|i\|_1=1}^{N^d}$, where $\mathrm{i} = (\mathrm{i}_1, \ldots, \mathrm{i}_\mathrm{d})$

is a D-dimensional indicator and $\|i\|_1 = \sum_{j=1}^{d} i_j$; *Step 2:* Select N c_i so that the indicators i^1, i^2 of any two small hypercubes satisfies: $i_j^1 \neq i_j^2, j = 1, \ldots, d$; *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 Uncertainty Quantification (UQ) that measures the impact of perturbations on the QoI function *y(x, θ)* 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 [47] to analyze subsequent SNS.

3.7.2 Minimum description length model selection

According to Sober [48], informative models are less complex. Kuhn [49] also noted that simpler models are preferable to complex ones. Turney [50] 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.

Table 1 Widely used model selection metrics

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

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

MDL [22] is a 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 [49,50], allowing for greater learning of underlying regularities governing the process of interest [24], This metric has a simple mathematical structure:

$$MDL(x^n) = -\log f_\theta(x^n) + L(\theta) \tag{24}$$

where, $-log\, f_\theta(x^n)$ is the average shortest encoding length that can be achieved by losslessly encoding n-dimensional data x^n using model M, that is, the lower bound of entropy. $L(\theta)$ represents the complexity of the model itself. Since the absolute value of $L(\theta)$ cannot be calculated directly, MDL uses the K–L divergence between the model $f_\theta(x^n)$ and the real generative data $q(x^n)$ as a measure of model complexity. Using the concept of Normalized Maximum Likelihood, Rissanen further demonstrated the redundancy of the observation model has the following asymptotic optimal lower bound of Minimax [51]:

$$\inf_{f_\theta} \sup_{\theta\in\Theta} \mathrm{K\text{–}L}\,(q, f_\theta) = \frac{k}{2}\log\frac{n}{2\pi e} + \log\int_K \sqrt{\det I(\theta)}\, d\theta + O(1) \tag{25}$$

where, θ is a compact subset of the parameter space Θ, k represents the number of parameters entering the model, I(θ) is the Fisher information matrix of the parameter distribution, and it is defined by an expectation $I_{ij} = -<\left(\frac{\partial^2 \log f(D|\theta)}{\partial\theta_i \partial\theta_j}\right)>_\theta$ over the parametric model $f_\theta(x^n)$. Our conclusion rigorously shows that model complexity estimates must approach the lower bound progressively. It has good generalizability [49] because it doesn't assume knowledge of the true data distribution or make specific assumptions about the observed data distribution.

3.8 Framework of machine learning & model driven UQ of inverse problems

Following the above methodology, Figure 3 sums up the proposed framework which will be used in section 5 as below. The model-driven approach [52] uses finite element numerical modeling to simulate data via design of experiments, providing a large dataset to train the data-driven, machine learning based surrogate model. This approach is used to solve forward problems. In contrast, the data-driven (machine learning-driven) approach [52, 53] 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. Our approach addresses these inverse problems, and the forward and inverse problems are complementary. In the new approach, we propose a new black box variational inference method combined with O'Hagan's Bayesian calibration framework, and embed MDL model selection to enhance the accuracy, efficiency, and robustness of both forward and inverse problems with generalization.

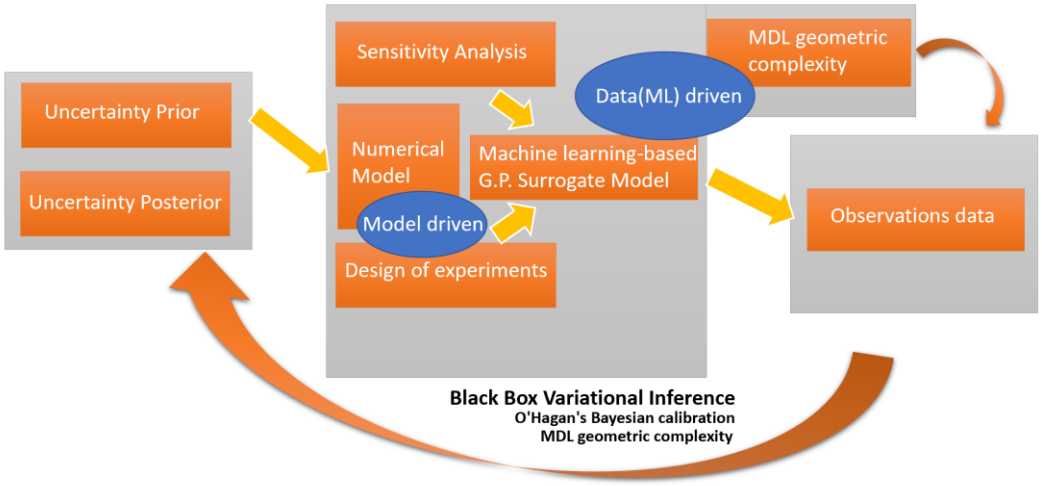


Figure 3 Framework of machine learning-based data & model driven UQ of inverse problems

4.0 SNS shaking table experiment

4.1 SNS shaking table experiment configurations

To validate our above framework, we utilized a recent large-scale shaking table experiment that we conducted in [26], and a brief summary is presented herein for completion. The experiment aimed to investigate the seismic effects of nonstructural systems under various seismic input wave conditions [54]. The 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 stories high, measuring 5.40 m in height, 12.84 m in length, and 11.64 m in width. The systems were suspended from the platform during the test. (see Figure 5).

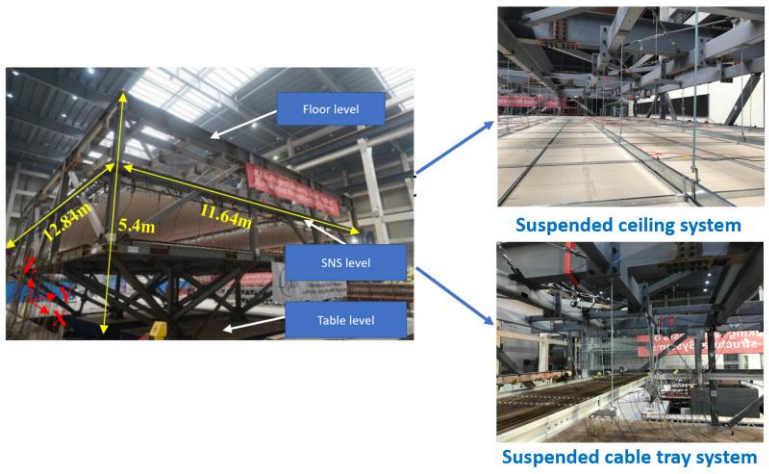


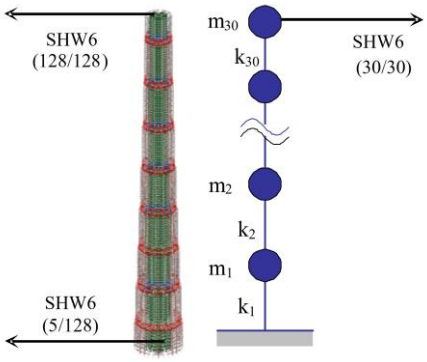
Figure 4 View of steel platform [97] and suspended ceiling and cable tray systems.

4.2 Loading protocol

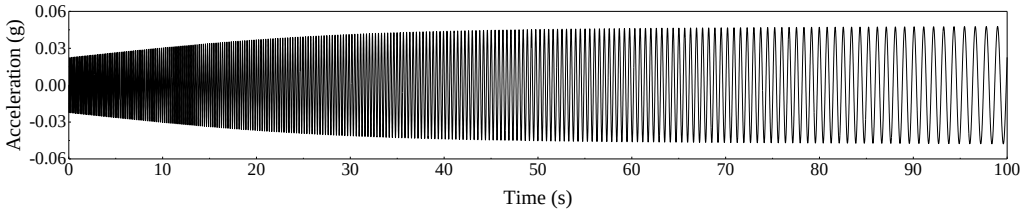
To accurately simulate the long duration and long period seismic input experienced by supertall

buildings, the test specimen was loaded with various frequencies after analysis, Table 2 provides a comprehensive list of all the motions tested of one type of SNS systems in our study [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 Figure 5(b) is a sine wave that oscillates 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) in figure 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 3.01s, 1.18 s, and 0.72 s, respectively. The characteristic period of the 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 the 30-story model is closer to the characteristic period of the ground motion SHW6 than that of the 128-story model, so the PFA of SHW6 (30/30) is larger than that of SHW6 (128/128). Details about the input motions are in [26].



(a) The 128-layer Benchmark (left) and 30-layer Stick (right) model of floor wave



(b) Acceleration time history curve of the Sweep wave

Figure 5 Schematic of partial motions of input

457 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	-

458 Notes:

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

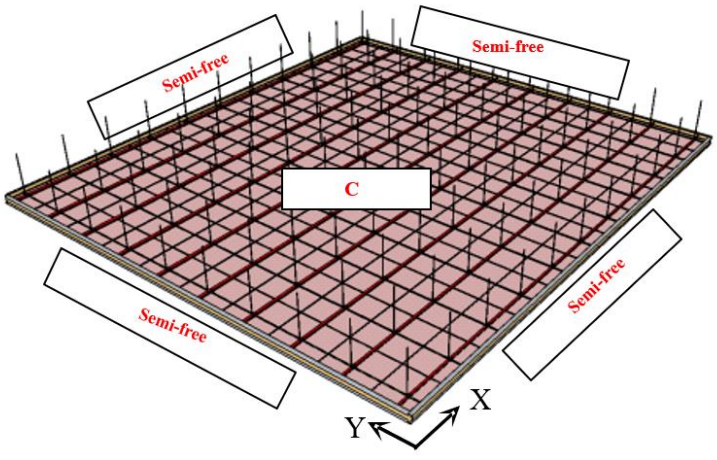
460 [2)] During Runs 12 and 14, the floor acceleration responses at the 5th and 128th floors of the 128-story building subjected
461 to the ground motion SHW6 are input, and during Run 16, the floor acceleration response at the 30th floor of the 30-
462 story building is input.

463 ## 5.0 Applications for suspended nonstructural systems (SNS)

464 ### *5.1 Data*

465 #### 5.1.1 Type C general information

466 We selected one of the SNS experiments, the type C Suspended Ceiling System (SCS) experiment,
467 as a case study to test our proposed uncertainty quantification methodology. This SCS consists of
468 a single suspended ceiling with a full semi-free boundary constraint condition, as depicted in figure
469 6. The suspended ceiling covers an area of 150 m^2, which is currently the largest worldwide.



470

Figure 6 SCS type C [26]

5.1.2 Instruments

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 Figure 7). The accelerometers measured the absolute acceleration, while the displacement transducer measured the displacement of the ceiling relative to the platform.

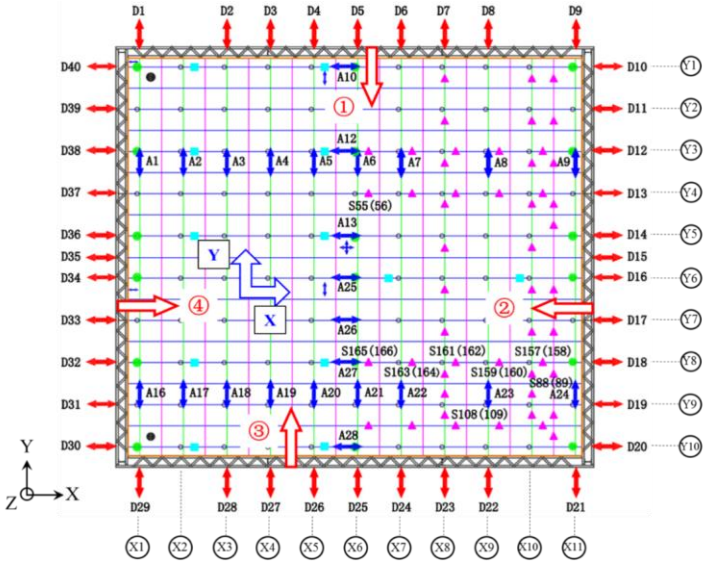


Figure 7 Instrumentations on the ceiling [26].

Notes: displacement transducer; accelerometer; strain gauge on the threaded rod; strain gauge on lay-in panel; strain gauge on cross tee and sub cross tee; strain gauge on carrying channel.

5.1.3 Boundary conditions

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

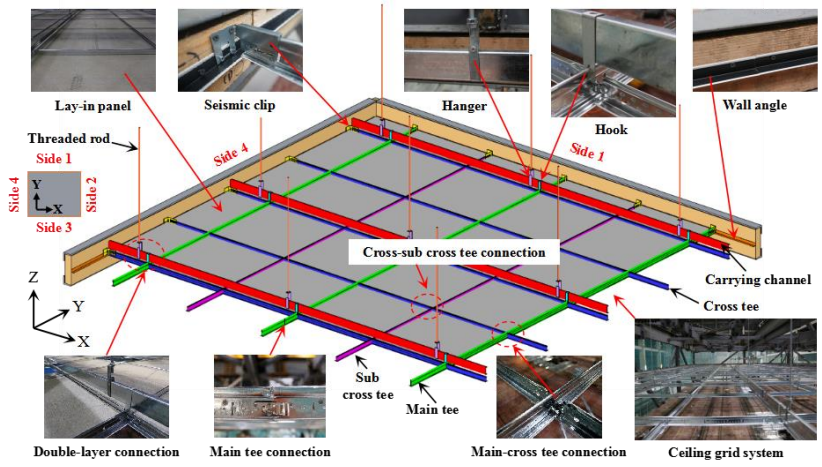
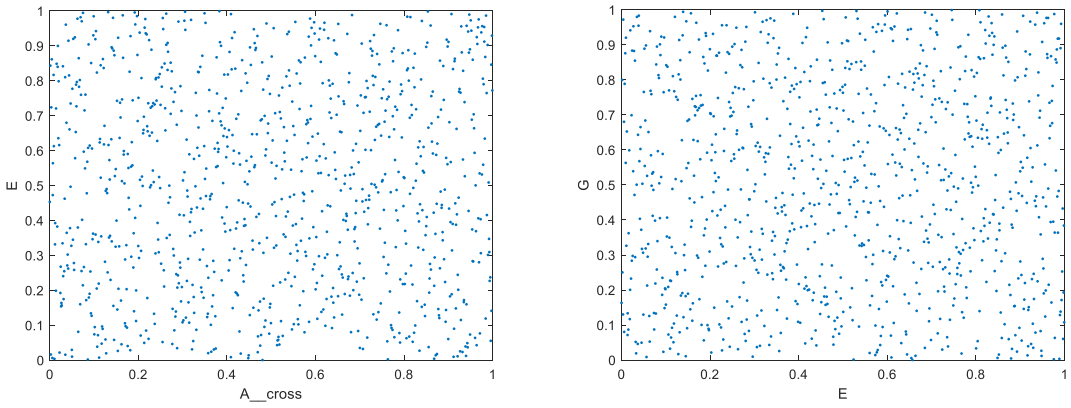


Figure 8 Composition of the double-layer SCS [97].

5.2 Training points

As previously mentioned, the LHS method is used for sampling combinations of training points. For each pair of the two sets of training data, we generated 1000 Latin Hypercube samples and selected two pairs of the training point data to be illustrated (also see Figure 9). And this will be used for forward running (model-driven) the computational numerical model of SNS systems.

It is based on Latin Hypercube Sampling (LHS) method. It uses 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 a result of maximizing the minimum distance between points. A detailed account of the method process can be found in section 3.6.



(a) A_cross and E LHS sampling (b) E and G LHS sampling

Figure 9 Latin Hypercube samples results

5.3 Priors

The parameters[2] selected in this work are based on linear model parameters (see Table 3). The fifteen selected inverse problem parameter names, symbols, prior, and observation prior data are as follows:

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)

[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.

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)

5.4 Sensitivity analysis

To conserve computing resources, a sensitivity analysis was conducted to identify the key uncertainty parameters (refer to Figure 10). The findings indicate that the 4^{th} and 13^{th} 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, which are listed in Table 4. A detailed account of the analysis process can be found in section 3.7.1.

Our sensitivity analysis revealed that certain parameters, such as those associated with main tee connections, cross tee connections, and boundary connections, significantly impact the accuracy of the model's predictions. During the validation process, for instance, we observed that the acceleration prediction results were not precise between 60-61 seconds. 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.

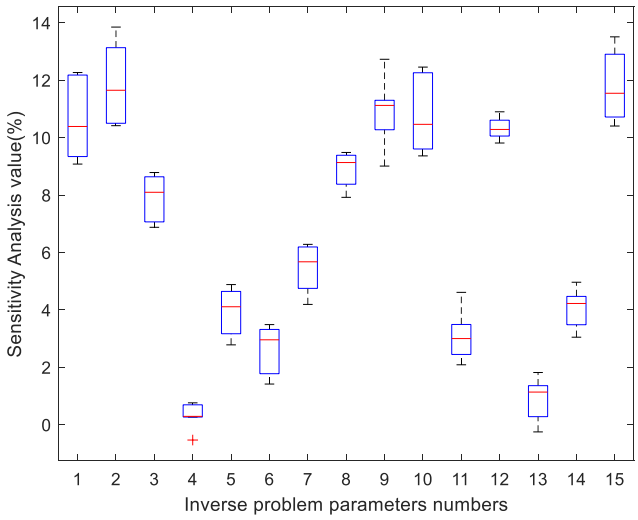


Figure 10 Sensitivity analysis result

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~ (30, 36)
E	Elastic modulus	U~ (1.9, 2.3) *100000
G	Shear modulus	U~ (7.4,8.2) *10000
I_y_*cross*	Y-dir moment of inertia	U~ (4000,4700)
I_z_*cross*	Z-dir moment of inertia	U~ (650,750)
K_rod	Threaded rods connection stiffness	U~ (0.5,3)
K_friction	Friction elastic stiffness	U~ (15,25)
K_cc1	X-dir Cross tee connections stiffness	U~ (750,850)
K_cc2	Y-dir Cross tee connections stiffness	U~ (75,125)
K_cc3	Z-dir Cross tee connections stiffness	U~ (10,35)
A_main	Main tee cross-section area	U~ (35,42)
I_y_*cross*	Y-dir moment of inertia	U~ (7250,7800)
K_mc	X-dir Main tee connections stiffness	U~ (1400,1600)

5.5 Data driven Machine learning-based Gaussian process surrogate model training

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

$$M\ddot{x}(t) + C\dot{x}(t) + Kx(t) = \mathrm{F(t)} \tag{26}$$

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. *F(t)* is the applied

excitation force on the system.

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

$$y_{obs}(x,t) = d\ddot{x}(t) + \epsilon_i \tag{27}$$

where, y_{obs} is the observed or measured response at time instant t. $\ddot{x}(t)$ and ϵ_i are the values of the acceleration response of the system and observation noise. The matrix d is a sensor placement matrix associated with accelerometers' locations.

We now consider Kennedy and O'Hagan's formulation [4], where data observation (frequency response) is 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 $X \subset \mathbb{R}^{\mathrm{D}}$, and θ is a set of parameters to be calibrated or inferred.

The relationship between data observation and simulation is assumed to be

$$y_{obs}(x,t) = \alpha y_{sim}(x,\theta,t) + \delta(x,t) + \epsilon(x) \tag{28}$$

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

$$\alpha = \frac{cov[y_{obs}(x,\ t),\ \eta_{sim}(x,\theta,t)]}{var[y_{sim}(x,\theta,t)]} \tag{29}$$

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 involves 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. Refer to section 3.1 for further details. By following this process, we can select the most appropriate kernel as squared exponential kernel:

$$\gamma_\delta(x,x') = \exp\left[-\sum_{i=1}^{D}\frac{(x_i-x_i')^2}{l_{i,\delta}}\right],\ \gamma_{sim}(x,x') = \exp\left[-\sum_{i=1}^{D}\frac{(x_i-x_i')^2}{l_{i,sim}}\right] \tag{30}$$

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

To optimize the performance of the selected kernel, we can perform hyperparameter tuning using maximum likelihood estimation to find the optimal values for the kernel's hyperparameters. This will help achieve the best performance for the final model. Section 3.1 provides more details on the process of hyperparameter tuning.

It is worth noting that in this study, we utilized a standard approach for splitting the LHS sample data into training, validation, and testing sets. Specifically, 70% of the LHS samples were used for training the Gaussian process, 20% for validation, and the remaining 10% for testing. To evaluate the performance of our 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 0.328, indicating good performance and generalization ability of the model. Furthermore, the training accuracy was 90.189%, and the testing accuracy was 90.527%, as shown in Figure 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, where our approach achieved a several hundred-fold improvements in time. The surrogate models also enable us to consider a larger range of input parameters in the future, which is important for SNS applications.

It is essential to note that the chosen splitting of the LHS sample data into training, validation, and testing sets is a widely accepted and well-established approach in the field of machine learning. This approach ensures that the model can generalize well to unseen data, which is crucial for the practical application of the model.

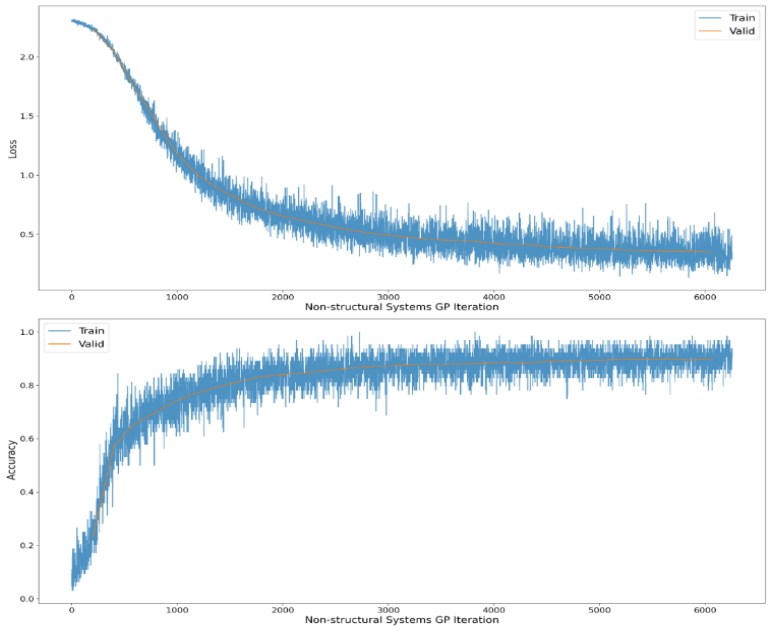


Figure 11 GP surrogate model training iteration result

5.6 Blackbox variational inference

In order to avoid any computational inefficiency, we substitute the MCMC by the variational inference for posterior distribution estimation, and 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.

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 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, $D_{sim} = \{x_i, \theta_i, \eta_i\}_{(i=1)}^{N_{sim}}$, $D_{obs} = \{x_i^*, \theta_i, y_i\}_{(i=1)}^{N_{obs}}$, respectively.

Now name $k = sim, \delta$, and $R_k(D_k)$ is the correlation matrix with $\gamma_k(x, x') \in D_k$, $D_{obs}(\theta) \coloneqq \{(x_i, \theta)\}_{i=1}^{N_{obs}}$ 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:

$$\mu_{y_{obs}}(x^*, \theta) = t_{obs}(x^*, \theta) V_{obs}(\theta)^{-1} y \tag{31}$$

$$\sigma^2_{y_{obs}}(x^*, \theta) = \sigma^2_{obs}(x^*) - t_{obs}(x^*, \theta) V_{obs}^{-1} t(x^*, \theta) \tag{32}$$

where, $$t_{obs}(x^*, \theta) = \begin{bmatrix} \alpha\sigma^2_{sim} R_{sim}\big((x^*, \theta), D_{sim}\big) \\ \alpha^2\sigma^2_{sim} R_{sim}\big((x^*, \theta), D_{sim}\big) + \sigma^2_{\delta} R_{sim}(x^*, D_{obs}) \end{bmatrix}$$

$$V(\theta) = \begin{bmatrix} V^{(sim,sim)} & V^{(sim,obs)}(\theta) \\ V^{(obs,sim)}(\theta) & V^{(obs,obs)}(\theta) \end{bmatrix} \tag{33}$$

and the above diagonal block matrices are given by

$$V^{sim,sim)} = \sigma^2_{sim}\big(R_{sim}(D_{sim}) + \sigma^2_{\epsilon_{sim}} I\big) \tag{34}$$

$$V^{(obs,obs)} = \sigma^2_{obs}\big(R_{obs}(D_{obs}) + \sigma^2_{\epsilon_{obs}} I\big) + \sigma^2_{sim}\alpha^2 (R_{sim}(D_{obs}(\theta)) + \sigma^2_{\epsilon_{sim}} I) \tag{35}$$

The off-diagonal blocks are given by

$$V^{(sim,\ obs)}(\theta) = \alpha\, V^{(sim,sim)}(D_{\_sim}, D_{\_obs}(\theta)) \tag{36}$$

We utilized Black Box Variational Inference to make Bayesian inference of the inverse problem, as described in algorithm 1 below:

Algorithm 1 Black box variational inference with O'Hagan Bayesian calibration & geometric complexity MDL for inverse problem Uncertainty Quantification of suspended nonstructural system

Input:

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

Output:

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

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

2 **Initialize:** $\boldsymbol{\lambda}_{1:n}$ randomly, $t = 1$

while *the* training accuracy *has not converged* **do**
  For s = 1 to S **do**
    // Draw sample from q
    $\mathbf{z[s]} \sim \mathbf{q(z|\lambda)}$
  end
  $\rho := t^{th}\ value\ of\ a\ Robbins\ Monte\ sequence$ // Equation (23) 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 $q^*(\boldsymbol{z}|\boldsymbol{\lambda})$
$MDL :=$ Minimum description length $complexity\ computation$ // Equation in table 1 in section 3.7.2
end

And the convergence of the ELBO iteration is presented in Figure 12.

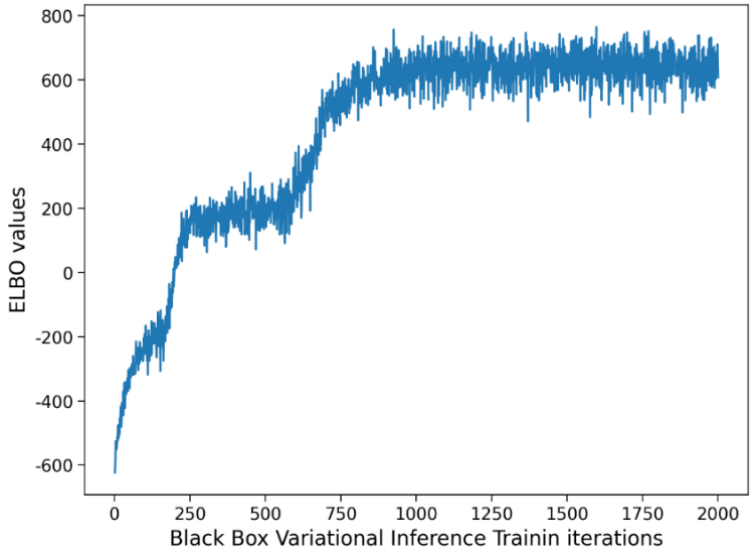


Figure 12 Blackbox variational inference training iterations

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

As outlined in section 3, we will utilize MDL for model selection. Based on the black box variational inference results of the posterior and MDL equations, we calculate the MDL complexity values of seven groups of models using the following method.

Table 5 Minimum Description Length complexity (MDL) values

Model type	MDL
A_cross, E, G, I_y_cross, I_z_cross, K_rod, K_friction, K_cc1, K_cc2, K_cc3, A_main, I_y_cross, K_mc	-593.7
A_cross, E, G, I_y_cross, K_rod, K_friction, K_cc1, K_cc2, K_cc3, A_main, I_y_cross, K_mc	-587.3
A_cross, E, G, K_rod, K_friction, K_cc1, K_cc2, K_cc3, A_main, I_y_cross, K_mc	**-597.9**
A_cross, E, G, K_rod, K_friction, K_cc1, K_cc2, A_main, I_y_cross, K_mc	-584.5
A_cross, E, G, K_rod, K_friction, K_cc1, K_cc2, A_main, K_mc	-552.1
A_cross, E, G, K_friction, K_cc1, K_cc2, A_main, K_mc	-522.4

After computing the MDL complexity values for each group of models, we select the model with the minimum MDL value. This model, with eleven parameters, is the optimal result for the inverse problem, as indicated by the posterior distribution results in Section 5.8. Furthermore, we compare the MDL results with those of other model selection below.

Table 5 Comparison of model selection criteria with selected model type

Model selection criteria	Selected Model type	Relative Generalization Value (MDL as benchmark)
Akaike information criterion (AIC)	*A_cross, E, G, I_y_cross, I_z_cross, K_rod, K_friction, K_cc1, K_cc2, K_cc3, A_main, I_y_cross, K_mc*	75.7%
Bayesian information criterion (BIC)	*A_cross, E, G, I_y_cross, K_rod, K_friction, K_cc1, K_cc2, K_cc3, A_main, I_y_cross, K_mc*	87.9%
Deviance information criterion (DIC)	*A_cross, E, G, I_y_cross, I_z_cross, K_rod, K_friction, K_cc1, K_cc2, K_cc3, A_main, I_y_cross, K_mc*	75.7%
Information-theoretic measure of complexity (ICOMP)	*A_cross, E, G, K_rod, K_friction, K_cc1, K_cc2, A_main, I_y_cross, K_mc*	98.3%
Minimum description length (MDL)	*A_cross, E, G, K_rod, K_friction, K_cc1, K_cc2, K_cc3, A_main, I_y_cross, K_mc*	**100% (Benchmark)**

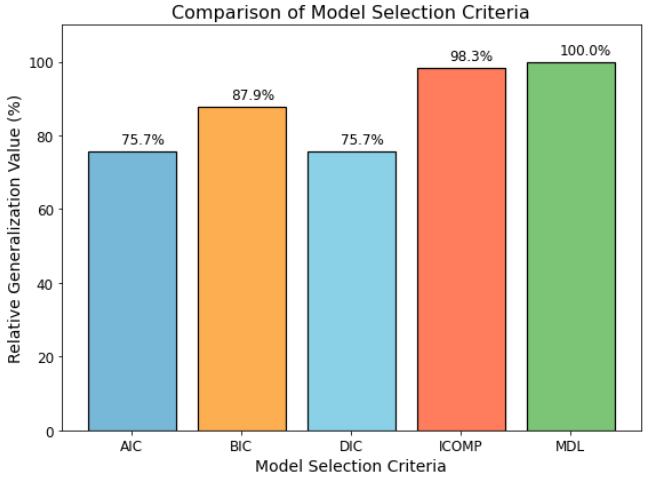


Figure 13 Model selection criteria relative generalization values comparision

Generalizability value is used to express the ability of the model calibration to predict the true model. Generalizability value is defined by the following equation [23]:

$$Gr(\%) = \left\{1 - \frac{\sum_{i=1}^{m} \left| y(x_i, \theta) - y_{truth}(x_i, \theta_n) \right|}{\sum_{i=1}^{m} \left| y_{truth}(x_i, \theta_n) \right|}\right\} \times 100 \tag{38}$$

where m is the number of discretized points that are used to evaluate Generalizability, and $y_{truth}(x_i, \theta_n)$ is the response of the true model. In this study, we use MDL as the benchmark to compare other model selection criteria.

Based on the comparison between the geometric complexity measure minimum description length

(MDL) and other model selection criteria with the result in the next section, it shows the intuitive advantage of it in terms of generalization performance applied in our SNS systems. AIC only consider the calibration parameters and will choose more complex model and leads to over-fitting, and DIC has a similar penalty for complexity to AIC. BIC has a stronger penalty for complexity than AIC because it includes the factor of the natural logarithm of the sample size. ICOMP almost has the same generalization, but become less-fitting, compared with MDL because it also not only consider calibration parameters, and sample size but also their sensitivity and interdependence. In summary, our study demonstrates that MDL strikes a better balance between model complexity and generalization performance, resulting in superior predictive performance in SNS systems. The MDL approach takes into account both calibration parameters and sample size, and geometric complexity information resulting in a more comprehensive and well-rounded evaluation of model complexity, leading to optimal generalization performance.

5.8 Results of new approach and discussions

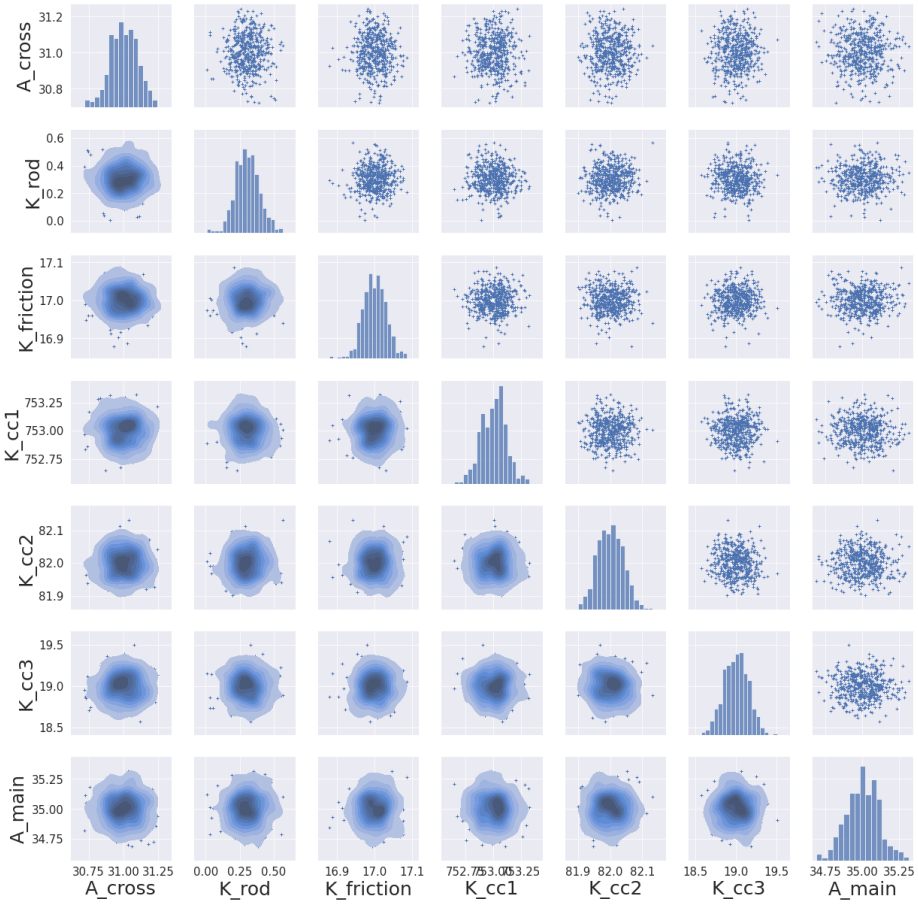


Figure 14 Posterior distribution of seven of inverse problems parameters

While it is true that variational inference is an approximation of MCMC, it is still a powerful tool for improving accuracy levels. Variational inference has been shown in the literature to

significantly improve inference time compared to Markov Chain Monte Carlo (MCMC) by several orders of magnitude. In details, MCMC can take several hours to generate posterior estimates, variational inference can accomplish the same task in just a few minutes. The posterior distribution resembles a normal distribution due to the use of an approximate normal distribution as the variational family in black box 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 less derivations and saving computational time. It has advantages over mean-field variational inference, which involves lengthy derivations and the use of conjugate distributions, such as Gaussian mixture models. Additionally, it avoids the complex burn-in process of MCMC iteration and ensures accuracy.

Our new approach embedded with geometric complexity measure produces posterior estimates with low bias and variance, indicating that it is able to accurately estimate the parameters of the system with good generalization. Figure 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. Due to space limitations in the figure, we randomly selected seven of the inverse problem parameters to display. Additionally, Figure 15 shows the posterior distribution of the first six frequency modes of the system, which are of particular interest in the context of suspended nonstructural systems. We can see that the posterior mean is very close to the observation data prior we mentioned above, indicating that our approach is able to produce accurate and robust estimates for SNS systems with uncertainty quantification.

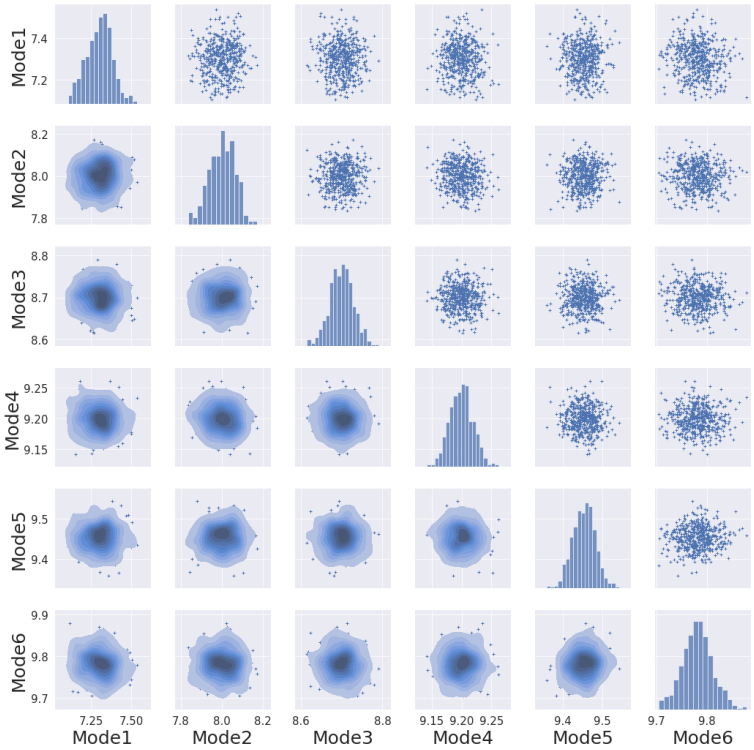
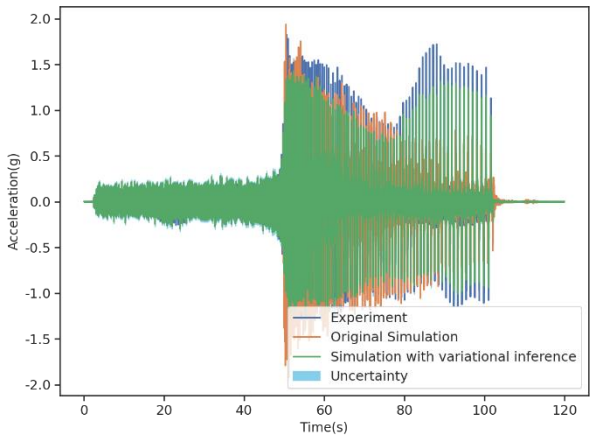
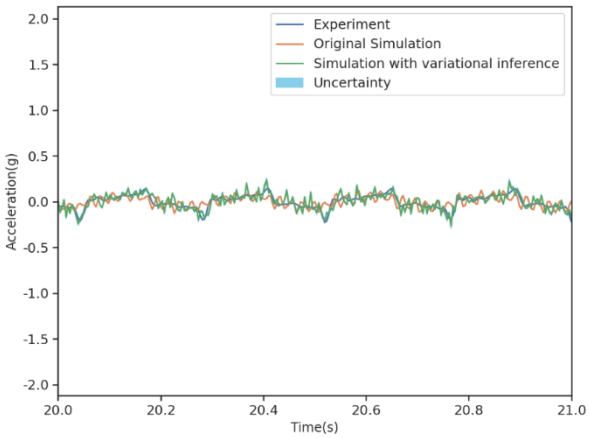


Figure 15 Posterior distribution of the first six frequency modes of the suspended ceiling system

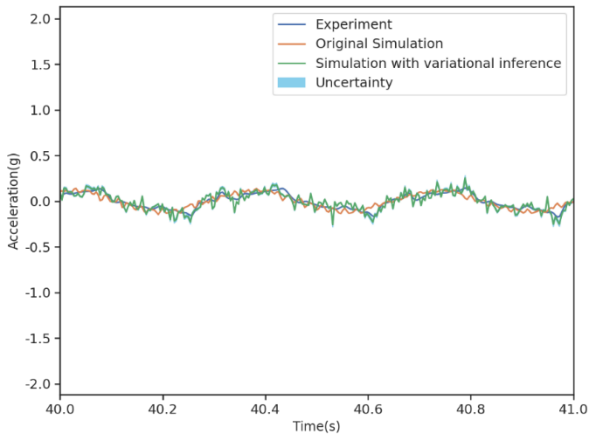
Based on the results of the posterior distribution of the inverse problems under our new approach above, we will verify and evaluate the seismic acceleration and displacement response, and the results are as follows. From Figure 16, we can find that the numerical simulation with variational inference has a much better validation result than the original simulation in terms of amplitude and trend. The validation improvement rate is achieved by around 50% ~70%. As discussed in the [26], part of the original acceleration response were validated well, and some were not; for a clearer view, we uniformly selected five short time ranges to double check: 20 sec ~ 21 sec, 40 sec~ 41 sec, 60 sec~ 61 sec, 80 sec~ 81 sec and 100 sec ~101 sec.



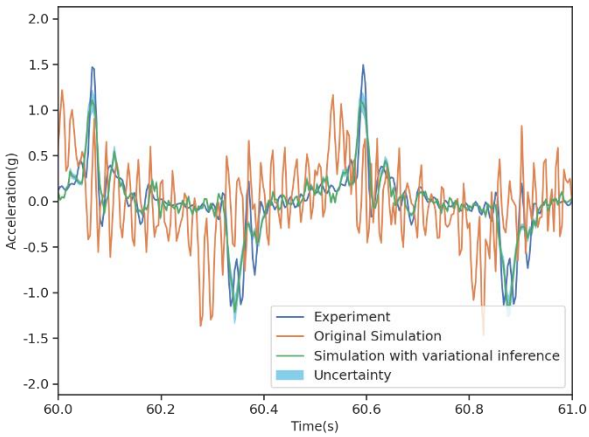
(a) The acceleration response validation comparison among experiments, original simulation, and simulation under variational inference with uncertainties



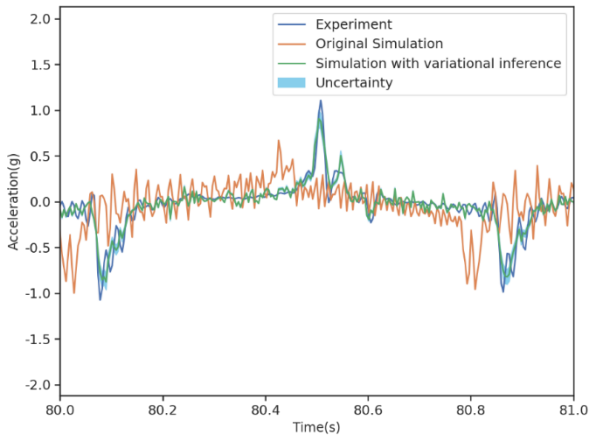
(b) The acceleration response validation comparison from 20 sec ~ 21 sec



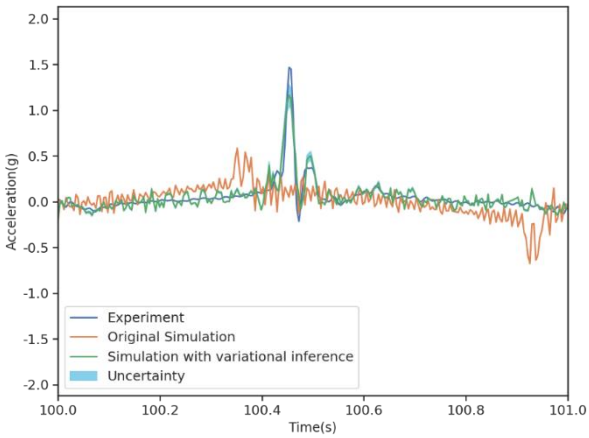
(c) The acceleration response validation comparison from 40 sec ~ 41 sec



(d) The acceleration response validation comparison from 60 sec ~ 61 sec



(e) The acceleration response validation comparison from 80 sec ~ 81 sec

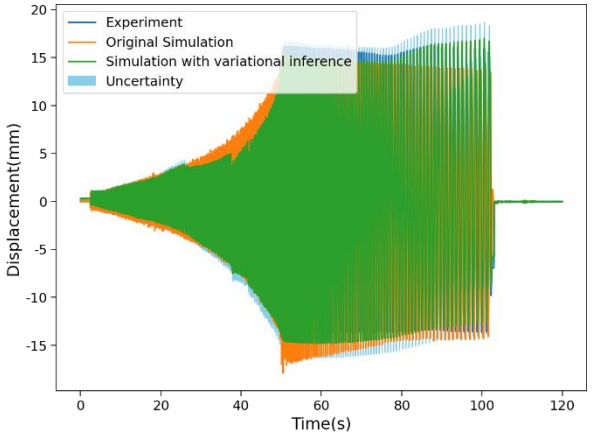


(f) The acceleration response validation comparison from 100 sec ~101 sec

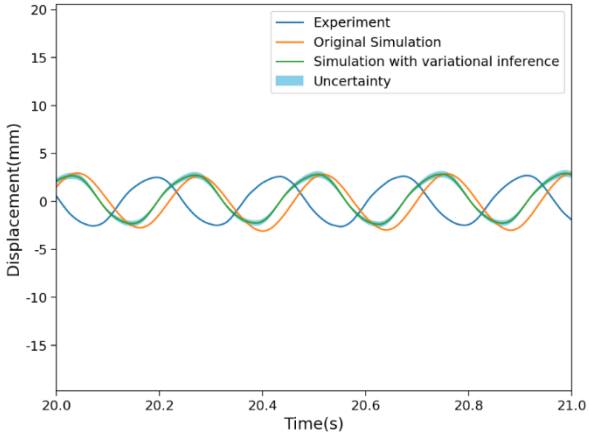
Figure 16 Comparison between the measured and predicted responses

In Figure 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 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 amplitude that fit well with the experimental data. However, after applying our new approach, we observed that the trend of simulation (f) was now fitted with the experimental data, and the amplitude of simulations (d), (e), and (f) were almost fitted well with the experimental data under the confidential uncertainty interval.

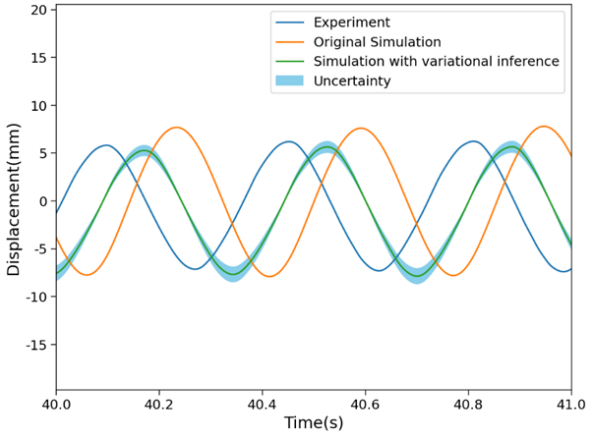
Also, the original displacement response can be seen as majorly validated well, and some were not, especially with some time delay difference. Figure 17 shows that the numerical simulation with variational inference has a much better displacement response validation result than the original simulation in terms of amplitude and trend with a confidential uncertainty interval. For a clearer view, a pick of uniformly select five time-range to double check: 20 sec ~ 21 sec, 40 sec~ 41 sec, 60 sec~ 61 sec, 80 sec~ 81 sec, and 100 sec ~101 sec. We can find that the validation improved especially the time delay difference result, and the overall improved rate is around 50%.



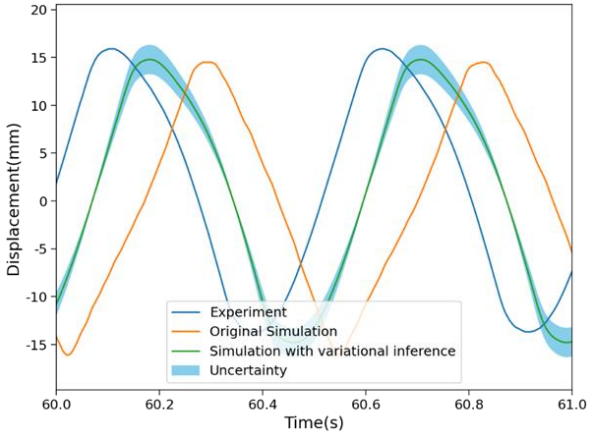
(a) The displacement response validation comparison among experiments, original simulation, and simulation under variational inference with uncertainties



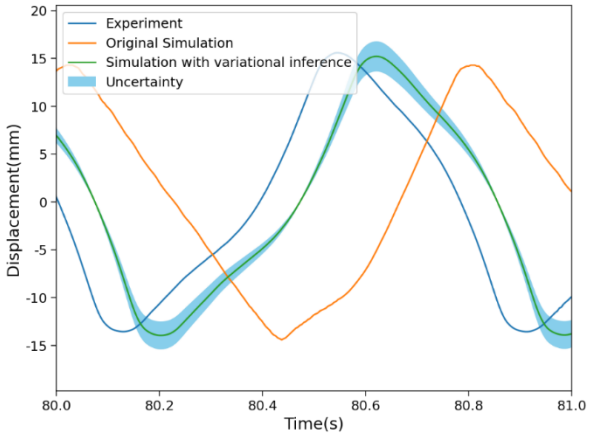
(b) The displacement response validation comparison from 20 sec ~21 sec



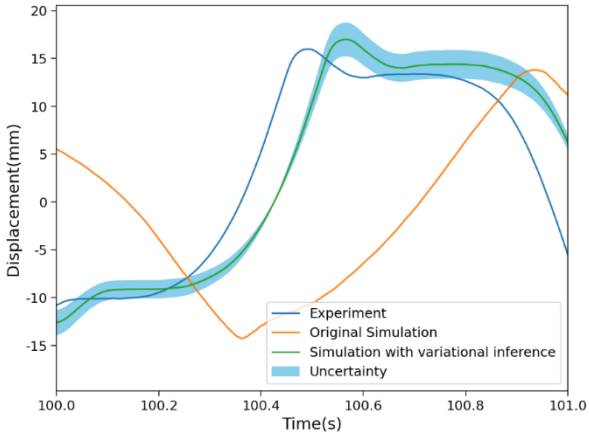
(c) The displacement response validation comparison from 40sec ~41 sec



(d) The displacement response validation comparison from 60 sec ~61 sec



(e) The displacement response validation comparison



(f) The displacement response validation comparison

Figure 17 Comparison between the measured and predicted responses

In Figure 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 under confidential uncertainty interval, and all the delay effects were reduced. The best performance improvement was observed in simulation (e), which exhibited an accumulation effect with 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 suspended nonstructural systems, and we were able to produce accurate and precise posterior estimates that provide valuable insights into the behavior of these systems.

Our proposed black-box variational inference, combined with O'Hagan's Bayesian calibration framework, provides a promising approach for inverse problems in Suspended Nonstructural Systems (SNS). Compared to traditional variational inference, which relies on conjugate assumptions and complex mathematical derivations, our method offers a simpler and more accessible solution that is easy for engineers and scientists to use. Furthermore, our approach incorporates a geometric complexity measure minimum description length into the framework, which results in accurate and robust Bayesian inference and Uncertainty Quantification outcomes. By improving the robustness and validation accuracy of SNS systems with excellent generalization capabilities, our method offers a valuable tool for predicting and understanding the behavior of these systems.

6.0 Conclusions

This article presents a novel uncertainty quantification methodology of inverse problem based on variational inference with an efficient machine learning (ML)-based surrogate model for predicting the response of suspended nonstructural systems (SNS) in super-tall 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 is seen to significantly improve the low efficiency of traditional Markov Chain Monte Carlo (MCMC) and ensures a high level of precision. When combined with ML-based Gaussian process surrogate models, the same method dramatically reduces 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 novel combination of Black box 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 scale 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 in terms of optimal generalization performance.

In summary, our study provides significant contributions to the SNS community by offering a comprehensive and well-rounded evaluation of model complexity and generalization performance and contribute to ensuring building resilience and safety against seismic risk events. Our findings suggest that our proposed methodology can offer 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.

As future work, we plan to investigate additional inverse problems for uncertainty quantification in the nonlinearities of the SNS system. We aim to replace the current machine learning-based Gaussian process model with a physics-informed deep learning model that incorporates prior physical knowledge [53, 55, 56] to improve accuracy and explanatory power, particularly in situations with limited data.

7.0 Acknowledgments

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

8.0 References

[1] Armin Tabandeh, Neetesh Sharma, Paolo Gardoni, Uncertainty propagation in risk and resilience analysis of hierarchical systems, Reliability Engineering & System Safety, Volume 219, 2022, 108208, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2021.108208.

[2] Oberkampf WL, Roy CJ. Verification and validation in scientific computing[M]. Cambridge: Cambridge University Press, 2010

[3] Kiureghian AD, Ditlevsen O. Aleatory or epistemic? Does it matter? [J] Structural Safety, 2009, 31: 105-112

[4] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P.: Design and Analysis of Computer Experiments, Stat. Sci., 4, 409–423, https://doi.org/10.1214/ss/1177012413, 1989.

[5] Hongyuan Guo, Dong You, Paolo Gardoni,Adaptive subset simulation for time-dependent small failure probability incorporating first failure time and single-loop surrogate model, Structural Safety, Volume 102, 2023, 102327, ISSN 0167-4730, https:// doi.org/ 10.1016/j.strusafe.2023. 102327.

[6] Yongjun Pan, Yu Sun, Zhixiong Li, Paolo Gardoni, Machine learning approaches to estimate suspension parameters for performance degradation assessment using accurate dynamic simulations, Reliability Engineering & System Safety, Volume 230, 2023, 108950, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2022.108950.

[7] Hongyuan Guo, You Dong, Paolo Gardoni, Efficient subset simulation for rare-event integrating point-evolution kernel density and adaptive polynomial chaos kriging, Mechanical Systems and Signal Processing, Volume 169, 2022, 108762, ISSN 0888-3270, https://doi.org/10.1016/j.ymssp.2021.108762.
[8] Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, J. Roy. Stat. Soc. Ser. B, 63, 425–464, https://doi.org/10.1111/1467-9868.00294, 2001.
[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] Xiaobing Shang, Li Su, Hai Fang, Bowen Zeng, Zhi Zhang, An efficient multi-fidelity Kriging surrogate model-based method for global sensitivity analysis, Reliability Engineering & System Safety, Volume 229, 2023, 108858, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2022.108858.
[11] Hastings WK. Monte Carlo sampling methods using Markov Chains and their applications[J]. Biometrikam, 1970, 57(1):97-109
[12] Hinton G E, van Camp D. keeping neural networks simple by minimising the description length of weights. [1993] Proceedings of COLT-93. 5-13
[13] Nathoo, F., Babul, A., Moiseev, A., Virji-Babul, N., and Beg, M. (2014). A variational Bayes spatiotemporal model for electromagnetic brain mapping. Biometrics, 70(1):132–143.
[14] Ni, Y., Li, J., & Chatzi, E. (2021). Probabilistic model updating via variational Bayesian inference and adaptive Gaussian process modeling. Computer Methods in Applied Mechanics and Engineering, 383, 114915.
[15] Dann, M. R., & Birkland, M. (2019). Structural deterioration modeling using variational inference. Journal of Computing in Civil Engineering, 33(1), 04018051.
[16] Ranganath R, Gerrish S, Blei D. Black box variational inference[C]. Artificial Intelligence and Statistics. 2014: 814-822.
[17] Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008), "Computer model calibration using high-dimensional output," Journal of the American Statistical Association, 103, 570–583.
[18] Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press, New York/London.
[19] Khanh T.P. Nguyen, Mitra Fouladirad, Antoine Grall, Model selection for degradation modeling and prognosis with health monitoring data, Reliability Engineering & System Safety, Volume 169, 2018, Pages 105-116, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2017.08.004.
[20] MacKay, D.J.C. (1992), "Bayesian Interpolation," Neural Comput. 4:415–447.
[21] Cutting, J. E., Bruno, N., Brady, N. P., & Moore, C. (1992). Selectivity, scope, and simplicity of models: A lesson from fitting judgments of perceived depth. Journal of Experimental Psychology: General, 121(3), 364-381
[22] Rissanen, J. (1978). Modeling by shortest data description. Automatica, 14(5), 465-471.
[23] Pitt, M., Myung, I., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. Psychologial Review, 109(3), 472-491.
[24] Myung, I., Balasubramanian, V., & Pitt, M. (2000). Counting probability distributions: Differential geometry and model selection. Proceedings of the National Academy of Sciences, 97(21), 11170-11175.
[25] Myung, I. J., Pitt, M. A., & Kim, W. (2005). Model evaluation, testing and selection. In

Lamberts, K. & Goldstone, R., (Eds.), Handbook of Cognition. London, UK: Sage Publications
[26] Yong Wang, Huanjun Jiang, Shojiro Motoyui, Kazuhiko Kasai, Zhiyuan Qin, Youlu Huang, Study on seismic performance of suspended ceiling system with semi-free boundary condition, Engineering Structures, Volume 275, Part A, 2023, 115208, ISSN 0141-0296.
[27] Sullivan T J. Introduction to uncertainty quantification [M]. Springer, 2015.
[28] Chantrasmi T, Iaccarino G. Forward and backward uncertainty propagation fordiscontinuous system response using the Pade-Legendre method [J]. International Journal for Uncertainty Quantification. 2012, 2 (2).
[29] Tikhonov A N. On the stability of inv. Prob. [C]. In Dokl. Akad. Nauk SSSR. 1943: 195–198.
[30] Tarantola A. Inverse problem theory and methods for model para. estimation [M]. siam, 2005.
[31] Engl H W, Hanke M, Neubauer A. Regularization of inverse problems [M]. Springer Science & Business Media, 1996.
[32] Cotter S L, Dashti M, Stuart A M. Approximation of Bayesian inverse problems for PDEs [J]. SIAM Journal on Numerical Analysis. 2010, 48 (1): 322–345.
[33] Dashti M, Harris S, Stuart A. Besov priors for Bayesian inverse problems [J]. arXiv preprint arXiv:1105.0889. 2011.
[34] Lasanen S. Non-Gaussian statistical inverse problems, Part I: Posterior distributions [J/OL]. Inverse Problems and Imaging. 2012, 6 (2): 215–266.
[35] Jia J, Peng J, Gao J. Bayesian approach to inverse problems for functions with avariable-index Besov prior [J/OL]. Inverse Problems. 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 [J/OL]. Mathematical Biosciences. 2017, 287: 42–53.
[37] Chen P, Villa U, Ghattas O. Hessian-based adaptive sparse quadrature for infinite dimensional Bayesian inverse problems [J/OL]. Computer Methods in Applied Mechanics & Engineering. 2017.
[38] Sun Z, Wang J, Li R, et al. LIF: A new Kriging based learning function and its application to structural reliability analysis [J]. Reliability Engineering & System Safety. 2017, 157: 152–165.
[39] Zimmer C, Meister M, Nguyen-Tuong D. Safe Active Learning for Time-Series Modeling with Gaussian Processes[C] Advances in Neural Infor. Processing Systems. 31. 2018: 2730-2739.
[40] Snoek J, Larochelle H, Adams R P. Practical bayesian optimization of machine learning algorithms [C]. In Advances in neural information processing systems. 2012: 2951–2959.
[41] Cutajar K, Bonilla E V, Michiardi P, et al. Practical Learning of Deep Gaussian Processes via Random Fourier Features[J]. 2016.
[42] Salimbeni H, Deisenroth M. Doubly Stochastic Variational Infer. for Deep GPs [J]. 2017.
[43] M.C. Kennedy, A. O'Hagan Bayesian calibration of computer models J R Stat Soc Ser B (Statistical Methodol, 63 (2001), pp. 425-464, 10.1111/1467-9868.00294]
[44] Blei D M, Kucukelbir A, McAuliffe J D. Variational inference: A review for statisticians[J]. Journal of the American Statistical Association, 2017, 112(518): 859-877.
[45] Helton J C, Davis F J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems [J]. Reliability Engineering & System Safety. 2003, 81 (1): 23–69.
[46] Schöbi R, Sudret B, Marelli S. Rare Event Estimation Using Polynomial-Chaos Kriging [J].

ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering. 2016: D4016002.
[47] Oladyshkin S, Nowak W. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion [J]. Reliability Engineering & System Safety. 2012,106: 179–190.
[48] SOBER, E. [1975]: Simplicity. Oxford: Oxford University Press, 1975
[49] M. H. Hansen, B. Yu. Model selection and the principle of minimum description length. Journal of the American Statistical Association, 2001, 96(454): 746-774
[50] Google and DeepMind research team. (2022). Why Neural Networks Find Simple Solutions: The Many Regularizers of Geometric Complexity. In 36th Conference on Neural Information Processing Systems (NeurIPS 2022).
[51] J. Rissanen. Strong optimality of the normalized ML models as universal codes and information in data. IEEE Transactions on Information Theory, 2001, 47(5): 1712-1717
[52] Willcox, K.E., Ghattas, O. & Heimbach, P. The imperative of physics-based modeling and inverse theory in computational science. Nat Comput Sci 1, 166–168 (2021). https://doi.org/10.1038/s43588-021-00040-z
[53] Yanwen Xu, Sara Kohtz, Jessica Boakye, Paolo Gardoni, Pingfeng Wan, Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges, Reliability Engineering & System Safety,Volume 230, 2023, 108900, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2022.108900.
[54] Lv Xilin. Shaking table model test methods and techniques for building structures (Second Edition) [M]. Beijing: Science Press, 2016. (In Chinese)
[55] Tapeh, A., Naser, M.Z. (2022). Artificial Intelligence, Machine Learning, and Deep Learning in Structural Engineering: A Scientometrics Review of Trends and Best Practices. Archives of Computational Methods in Engineering. https://doi.org/10.1007/s11831-022-09793-w.
[56] Naser M.Z. (2021). “Mapping Functions: A Physics-guided, Data-driven and Algorithm-agnostic Machine Learning Approach to Discover Descriptive Expressions of Engineering Phenomena.” Measurement. https://doi.org/10.1016/j.measurement.2021.110098