

Full length article

A surrogate-assisted stochastic optimization inversion algorithm: Parameter identification of dams

Li YiFei^a, M. Amin Hariri-Ardebili^{b,c,*}, Deng TongFa^d, Wei QingYang^a, Cao MaoSen^{d,e,**}

^a Department of Engineering Mechanics, Hohai University, Nanjing, China

^b Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO, USA

^c College of Computer, Mathematical and Natural Sciences, University of Maryland, College Park, MD, USA

^d Jiangxi Province Key Laboratory of Environmental Geotechnical Engineering and Hazards Control, Jiangxi University of Science and Technology, Ganzhou, China

^e Institute of Structural Dynamics and Control, Hohai University, Nanjing, China



ARTICLE INFO

Keywords:

Concrete dams
Surrogate model
Stochastic optimization inversion
Parameters identification

ABSTRACT

Dynamic monitoring data plays an essential role in the structural health monitoring of dams. This study presents a surrogate-assisted stochastic optimization inversion (SASOI) algorithm, a novel technique for static and dynamic parameter identification. This algorithm is based on probabilistic finite element simulations and Bayesian inference theory. It combines the advantages of low computational cost in surrogate models and fast convergence in the Bayesian algorithm. Taking four cases of different complexity, this paper verifies the effectiveness of the SASOI algorithm and validates its practicality for large dams. Surrogate models consider several alternatives, including polynomial chaos expansion (PCE), Kriging, polynomial chaos Kriging, and support vector regression. Implementation of the SASOI algorithm on dams shows that PCE outperforms other techniques. This algorithm improves the accuracy and efficiency of the static parameter identification methods by nearly 27 times compared to the classical inversion methods. Furthermore, the accuracy of dynamic parameter identification is higher than that of static one. The SASOI algorithm is applicable to other large-scale infrastructures.

1. Introduction

Dams are the main element of water conservancy infrastructure that provides drinking water and is used for irrigation and power generation. Dam failure may cause loss of life for those who live in the dam downstream and property loss. Therefore, the safe operation of dams is an essential task in risk management [1]. One of the main elements in safety monitoring and health diagnosis of dams is to determine the physical parameters of the dam itself and the bedrock materials [2,3].

Parameter identification is an important task for structural health monitoring and damage detection in concrete dams [4]. Different material and structural properties change during the lifetime of the dams which significantly affects the structural performance. Proper estimation of these parameters is of great importance to ensure the safe operation of dams. Parameter identification in dams is even more complex because a large volume of concrete is typically used to build a dam over the years with aggregates sought from different sources. Moreover, the deterioration (or aging) rate varies at different dam locations as a function of temperature, relative humidity, stress state, etc. All these factors cause a non-homogeneous condition in which a

limited number of concrete core samples cannot properly quantify the true material properties. Parameter identification is, indeed, a low-cost technique to estimate the unknown parameters in the numerical model of a dam (such as concrete modulus of elasticity) using the known data (i.e., measurements). The calibrated numerical model is then used for health monitoring purposes.

The classical methods for parameter identification of dams are primarily based on analytical or numerical displacement back analysis [5, 6]. This inversion method has an irreconcilable contradiction between accuracy and efficiency and has a limited scope of application. The classical methods are not applicable to dynamic dam response identification. These classical methods mainly rely on an iterative approach to change the material parameters in the finite element model to minimize the error between numerical simulation and the field measurement [7]. This is, indeed, the extreme value optimization problem discussed in Neuman et al. [8]. Generally, parameter identification is a multi-peak, non-convex, and high-dimensional optimization problem. Finding the global minima is typically challenging using the classical gradient optimization algorithm. Therefore, many intelligent algorithms such as

* Corresponding author at: College of Computer, Mathematical and Natural Sciences, University of Maryland, College Park, MD, USA.

** Corresponding author at: Institute of Structural Dynamics and Control, Hohai University, Nanjing, China.

E-mail addresses: hariri@umd.edu (M.A. Hariri-Ardebili), cmszhy@hhu.edu.cn (M.S. Cao).

genetic, particle swarm, and Jaya algorithms are continuously applied to parameter optimization inversion.

Most of these optimization algorithms are essentially an iterative process to find the extreme points in which the numerical model needs to be called many times [9]. Such a crude operation is problematic for complex large-scale structures with a nonlinear response, such as concrete dams. The calculation cost of a single process for a refined finite element model of a dam-reservoir-foundation system takes several days, making the classical optimization algorithms inefficient to use. In addition, the classical algorithms are mainly based on the deterministic back analysis with engineering judgment or a few field experimental data, which ignores the aleatory and/or epistemic uncertainty caused by empirical judgment and measurement errors [10]. Therefore, it is not easy to quantify the impact caused by the inherent random characteristics of materials [11].

To avoid the shortcomings of the classical optimization algorithms, a Bayesian optimization method is adopted in this paper [12]. The Bayesian optimization method is an advanced and state-of-the-art technique in the field of hyperparameter optimization. Nearly all modern hyper-parameter optimization techniques that yield high-quality results in terms of efficiency and effectiveness are rooted in the basic idea of Bayesian optimization. The Bayesian inversion method is based on Bayesian optimization, which only needs a limited number of samples to infer the maximum value of the likelihood function, to achieve the purpose of parameter identification [13].

Surrogate modeling (a.k.a meta-modeling) attempts to offset the increased costs of stochastic modeling by substituting the expensive computational models (e.g., analytical formula or finite element models) with inexpensive-to-use surrogates models [14]. In recent years, surrogate modeling has gained a lot of attention in parameter stochastic optimization inversion. This method is typically coupled with evolutionary optimization algorithms to establish a surrogate model (sometimes a black box) that can characterize the mapping relationship between input parameters (with random characteristics) and the system's output (i.e., the quantity of interest — QoI) [15]. Surrogate models may appear in different forms such as polynomial chaos expansion (PCE) [16], Kriging (i.e., Gaussian process regression), and support vector regression (SVR).

1.1. Applied PCE and Kriging for dams

PCE, Kriging, and polynomial chaos Kriging (PCK) have been widely used in dam engineering. To the best of the author's knowledge, the first application was presented by Ghanem et al. [17] for static analysis of embankment dams. Guo et al. [18,19] evaluated the sliding stability of embankment dams using the sparse and adaptive PCE. Sevieri et al. [10] presented a new probabilistic procedure to identify the model parameters of the gravity dams based on a generalized PCE model in a Bayesian framework. Hariri-Ardebili and Sudret [20] presented the application of PCE in both analytical and numerical analyses of concrete dams. They evaluated the stability of the 2D gravity dam, seismic analysis of the 2D gravity dam, and also 3D frequency analysis of arch dams. Amini et al. [21] extended this research for sensitivity and reliability analysis of aging dams using the copula dependency among the random variables (RVs).

Guo and Dias [22] presented a Kriging-based probabilistic analysis of an earth dam which combined the Kriging model with the Monte Carlo simulation (MCS), the global sensitivity analysis, and the first-order reliability method. Kalinina et al. [23] applied the PCE to a dam-break model, and uncertainty in the inputs was propagated to the flow quantities downstream of the dam. Shahzadi and Soulaïmani [24] combined the PCE with a deep neural network for sensitivity and uncertainty propagation in rockfill dams. Hariri-Ardebili et al. [25] combined the PCE method with random forests for sensitivity analysis of symmetry and asymmetry arch dams and to identify the most critical locations. Sevieri et al. [26] developed a hierarchical

Bayesian framework for uncertainty reduction in the seismic fragility analysis of concrete gravity dams using general PCE and Markov Chain Monte Carlo.

Apart from the analysis of existing dams, the PCE, Kriging, and PCK have been used to design a new dam. Fengjie and Lahmer [27] combined the genetic algorithm with adaptive Kriging MCS for arch dam shape optimization. Wang et al. [28,29] proposed an optimization algorithm combination of the genetic algorithm and an updated Kriging surrogate model for gravity and arch dams. Abdollahi et al. [30] proposed an uncertainty-aware dynamic shape optimization based on a two-stage adaptive Kriging-assisted quantile-based design algorithm.

1.2. Application of SVR on dams

Support vector regression has been adopted for different response prediction models in dam engineering. Su et al. [31,32] implemented SVR for early-warning dam safety assessment. Ranković et al. [33] developed an SVR-based identification model for the prediction of dam structural behavior. Tabari and Sanayei [34] used the SVR model to predict the intermediate block displacement of the dam crest. Chen et al. [35] proposed a time-varying identification model for crack monitoring data from concrete dams based on SVR and the Bayesian framework. Hariri-Ardebili and Pourkamali-Anaraki [36] applied the SVR for reliability analysis of concrete dams using both the simplified method and the nonlinear numerical simulations. Lin et al. [37] compared SVR with other soft computing methods to develop a forecasting model for dam deformation. Zhou et al. [38] applied the SVR for seismic fragility analysis of high concrete-faced rockfill dams. Ren et al. [39] developed a multiple-point monitoring model for dam displacements based on correlated multiple-output SVR.

1.3. Parameter identification in dams

Su et al. [40] used the genetic simulated annealing algorithm to identify the mechanical parameters of the dam and its foundation. The results proved the efficiency and robustness of the algorithm. Karimi et al. [41] used an artificial neural network coupled with finite element, and boundary element approaches for system identification of concrete dams. Chen et al. [42] adapted the improved cuckoo search algorithm and the improved particle swarm optimization to adjust the mechanical parameters and identify the dam-zoning elasticity modulus for the heightened concrete dam. Yang et al. [43] established the statistical model of dam deformation monitoring data based on the backward elimination partial least squares and then adapted the improved particle swarm optimization to identify the elastic modulus of the dam body and foundation rock. Fedele et al. [44] applied ten separate artificial neural networks to inverse ten elastic moduli of the dam body and bedrock. To enhance noise robustness, artificial noise was added to the training samples.

Static monitoring data (e.g., static displacement and temperature) is typically used for parameter identification in concrete dams. Kang et al. [45] proposed a kernel extreme learning machine-based accelerated Jaya algorithm to minimize the objective function of dam material parameter identification. Liu et al. [46] proposed a parameter optimization inversion method based on unconstrained Lagrangian SVR and cultural genetic algorithm. They applied this technique to optimize the zoned elasticity module of a high arch dam in its initial impound period. Bao et al. [47] adopted a multi-output least-squares SVR machine combined with an improved differential evolution algorithm to estimate multiple mechanical parameters of the Jinping-I arch dam. The proposed algorithm outperformed four other methods.

With rapid development in structural dynamic response identification technology, information regarding the vibration mode has been used to identify the overall mechanical characteristics of the structures [48]. Li et al. [49] proposed a stochastic optimization inversion

method for dynamic parameters of high arch dams based on the response surface model and genetic algorithm. The vibration tests were performed on a prototype arch dam during flooding, and the dynamic elastic modulus of the dam and foundation were estimated. Kang et al. [50] proposed a method based on the Kriging and Jaya algorithm to identify the dynamic parameters of concrete dams rapidly and compared it with traditional particle swarm optimization and genetic algorithms. Liu et al. [51] computed the first three major frequencies and the basic modal shapes of the dam from the acceleration response induced by the underwater explosion and used them to construct an objective function. The dynamic elastic modulus of the dam was identified based on the SVR.

1.4. Contributions and objectives

As discussed before, the conventional parameter identification method has a limited scope of application and is not applicable to dynamic response identification. While using Bayesian optimization for parameter identification yields high-quality results, it is still a very demanding procedure, especially for complex infrastructures like dams. This shortcoming necessitates the application of surrogate models. Most of the literature mentioned above only uses one surrogate model within the optimization algorithm to identify the unknown parameters of the structure. There is no previous comprehensive research on such an integrated model's importance, accuracy, and efficiency in parameter identification.

In this paper, we propose a novel technique that combines (a) an inversion algorithm, (b) Bayesian optimization, and (c) surrogate models for the parameter identification of structural systems. The proposed hybrid framework is applicable to both static and dynamic systems. Another novelty of this work relies on integrating four surrogate models within the proposed framework that increases the accuracy and reliability of the results. To the best of the author's knowledge, and following the extensive literature review in the previous sections, this is the first study that addresses the rapid identification of static and dynamic mechanical parameters of concrete dams in the context of multiple surrogate models combined with Bayesian inversion theory.

The main objective of this study is to answer the following questions: (1) How to develop an accurate and efficient surrogate model to replace the fine-grid numerical model? (2) How to determine the most optimal surrogate model? (3) How to combine a surrogate model with Bayesian inference to accelerate the unknown parameter identification in the concrete dam? and (4) What are the main advantages of the SASOI algorithm compared to the classical iterative-based optimization inversion method?

To adequately address these questions, four different case studies are discussed, which cover the following aspects: (1) comparing the problems with explicit function vs. black-box-type models, (2) comparing the static vs. dynamic simulations, (3) comparing the simplified analytical models vs. advanced stochastic simulations, and (4) comparing the numerical simulation with experimental results.

A brief review of the selected surrogate models and the Bayesian inversion theory is provided in Section 2 for those readers who are less familiar with these concepts. A detailed description of the general procedures and characteristics of the SASOI algorithm is presented in Section 3 and the indicators and benchmarks for algorithm evaluation are in Section 4. Then, the static and dynamic parameters identification of dams using the SASOI algorithm are presented in Section 5, followed by conclusions in Section 6.

2. Underpinning theories of SASOI

2.1. Surrogate models

2.1.1. Polynomial chaos expansion

The classic PCE was first introduced by Wiener [52]. The key concept of PCE is to expand the model response onto a basis consisting

of multi-variate polynomials, which are orthogonal with respect to the joint distribution of the input variables. In this setting, characterizing the response probability density function (PDF) is equivalent to evaluating the PCE coefficients [53]. There are multiple strategies to calculate the PCE coefficients, which might be divided into intrusive or non-intrusive approaches. Consider a M dimensional random vector with independent components $\mathbf{X} = \{X_1, X_2, \dots, X_M\}$ described by the joint PDF $f_{\mathbf{X}}, i = 1, 2, \dots, M$. Thus, the scalar QoIs resulting from this system is also a random variable, denoted $Y = \mathcal{M}(\mathbf{X})$. Knowing that Y has a finite variance, it can be represented as a PCE [54]:

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^M} \zeta_{\alpha} \psi_{\alpha}(\mathbf{X}) \quad (1)$$

where $\alpha = \{\alpha_1, \dots, \alpha_n\} (\alpha_i \geq 0) \in \mathbb{N}^M$ is the multidimensional index notation vector that identifies the components of the multivariate polynomials ψ_{α} and the $\zeta_{\alpha} \in \mathbb{R}$ are the expansion coefficients to be determined. $\psi_{\alpha}(\mathbf{X}) = \prod_{i=1}^M \phi_{\alpha_i}^{(i)}(x_i)$ are multivariate polynomials orthonormal with respect to $f_{\mathbf{X}}$, among them, $\phi_{\alpha_i}^{(i)}$ is the univariate orthogonal polynomial in the i th variable of corresponding polynomial degree α_i .

In real-world problems, the truncated form of Eq. (1) is used as follows:

$$\mathcal{M}(\mathbf{X}) \approx \mathcal{M}^{PCE}(\mathbf{X}) = \sum_{\alpha \in \mathcal{A}} \zeta_{\alpha} \psi_{\alpha}(\mathbf{X}) \quad (2)$$

where $\mathcal{A} \in \mathbb{N}^M$ is the truncation set of multi-indices of cardinality P . There are two main truncation schemes, i.e., standard and hyperbolic [55]. The former one corresponds to all polynomials in the M input variables of total degree less than or equal to p :

$$\mathcal{A}^{M,p} = \{\alpha \in \mathbb{N}^M : |\alpha| \leq p\} \quad \text{card } \mathcal{A}^{M,p} \equiv P = \binom{M+p}{p} = \frac{(M+p)!}{p!M!} \quad (3)$$

A modification of the standard scheme, the hyperbolic truncation scheme makes use of the parametric q to define the truncation:

$$\mathcal{A}^{M,p,q} = \{\alpha \in \mathcal{A}^{M,p} : \|\alpha\|_q \leq p\}, \quad \|\alpha\| = \left(\sum_{i=1}^M \alpha_i^q \right)^{1/q} \quad (4)$$

where using $q = 1$, the hyperbolic truncation yields the standard truncation scheme in Eq. (3). For $q < 1$, hyperbolic truncation includes all the high-degree terms in every single variable, but high-order interaction terms should be avoided to the extent possible. An illustration presents a set of 2D hyperbolic truncation with varying p and q can be found in Hariri-Ardebili and Sudret [20].

The infinite series in Eq. (1) can be written as a sum of its truncated version Eq. (2) and a residual:

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{j=0}^{P-1} \zeta_j \psi_j(\mathbf{X}) + \varepsilon_p \equiv \zeta^T \psi(\mathbf{X}) + \varepsilon_p \quad (5)$$

where P is defined based on Eq. (3), ε_p is the truncation error, and superscript T means transpose, $\zeta = \{\zeta_0, \dots, \zeta_{P-1}\}^T$ is a vector containing the coefficients and $\psi(\mathbf{X}) = \{\psi_0(\mathbf{X}), \dots, \psi_{P-1}(\mathbf{X})\}^T$ is the vector that assembles the values of all the orthonormal polynomials in \mathbf{X} .

There are multiple techniques to find the expansion coefficients in Eq. (1). The least angle regression (LAR) algorithm uses low-rank truncation schemes. It aims to find coefficient vectors with only a few non-zero entries (i.e., sparse solutions), while the other coefficients are set to zero [56]. The LAR algorithm can be formulated by expanding the least square minimization and adding a penalty term $\lambda \|\zeta\|_1$ as:

$$\hat{\zeta} = \arg \min_{\zeta \in \mathbb{R}^P} \mathbb{E} [(\zeta^T \psi(\mathbf{X}) - Y)^2] + \lambda \|\zeta\|_1 \quad (6)$$

where $\|\zeta\|_1 = \sum_{\alpha \in \mathcal{A}} |\zeta_{\alpha}|$ is the regularization term that forces the minimization to favor low rank solutions.

2.1.2. Kriging

Sacks et al. [57] introduced the concept of Kriging, which was used to represent an input/output mapping of an expensive computational model in the context of a surrogate model. Santner et al. [58] presented an in-depth introduction to Kriging as a meta-modeling tool. Dubourg et al. [59] summarized the common type of Kriging as simple Kriging, ordinary Kriging, and Universal Kriging. The first two are special cases of universal Kriging. The universal Kriging aims to find the best linear unbiased predictor while minimizing the prediction's mean square error. This is composed of a polynomial term used for global trend prediction and a Gaussian process term used for local deviation regression. In addition, the correlation function (i.e., kernel or covariance function) is a crucial ingredient for a Kriging model since it contains the assumptions about the approximation function and controls the smoothness of the Kriging model. Lataniotis et al. [60] introduced some typical one-dimensional correlation functions like linear, exponential, Gaussian, and Matérn.

Again, suppose that the model output $Y = \mathcal{M}(x)$ is a realization of a Gaussian process indexed by $x \in \mathcal{D}_X \subset \mathbb{R}^M$. A Kriging meta-model can be expressed as Lataniotis et al. [61]:

$$Y \approx \mathcal{M}^K(x) = \beta^T f(x) + Z(x, \omega) \quad (7)$$

where $\beta^T f(x)$ is the mean value (i.e., trend) of the Kriging model, β represents the regression coefficient vector, $f(x) = [f_1(x), \dots, f_M(x)]$ is the polynomial basis function. In addition, $Z(x, \omega)$ is a zero mean, unit variance, stationary Gaussian process, and its covariance function is defined as:

$$\text{cov}(Z(x_i), Z(x_j)) = \sigma^2 R(x_i, x_j; \theta) \quad (8)$$

where σ^2 is the (constant) variance of Gaussian process $Z(x, \omega)$, and $R(x_i, x_j; \theta)$ is the correlation function which describes the "similarity" between two observations with hyper-parameters $\theta = [\theta_1, \dots, \theta_n]^T$. The Matérn-5/2 correlation function is selected in this study [61].

For $\mathcal{Y} = \{\mathcal{M}(x^{(1)}), \dots, \mathcal{M}(x^{(N)})\}^T$ which is assumed to follow a multivariate Gaussian distribution, the unknown Kriging parameters $\gamma = (\beta, \sigma^2, \theta)$ can be estimated by maximizing the likelihood function as:

$$\mathcal{L}(\gamma; \mathcal{Y}) = \frac{(\det C)^{-1/2}}{(2\pi)^{N/2}} \exp\left[-\frac{1}{2}(\mathcal{Y} - F\beta)^T C^{-1}(\mathcal{Y} - F\beta)\right] \quad (9)$$

where the covariance matrix $C = \sigma^2 R + \Sigma_n$ sums up the covariance matrix of the Gaussian processes and noisy response; $F = [f(x_1), \dots, f(x_N)]^T$ is the observation (design) matrix of the Kriging metamodel trend.

Taking the partial derivative of the log-likelihood function with respect to β and σ^2 to zeros, the hyper-parameters θ can be obtained from solving the optimization problem in Eq. (10). This optimization problem can be solved using the covariance matrix adaptation-evolution strategy (CMA-ES). This is a de-randomized stochastic search algorithm introduced by Hansen and Ostermeier [62].

$$\hat{\theta} = \arg \min_{\theta \in \mathcal{D}_\theta} [-\log \mathcal{L}(\theta; \mathcal{Y})] \quad (10)$$

2.1.3. Polynomial chaos Kriging

PCK is a state-of-the-art non-intrusive surrogate modeling approach that combines the advantages of PCE and Kriging: the regression-type PCE captures the global behavior of the computational model, whereas the interpolation-type Kriging approximates local variations [63]. A detailed derivation of the PCK surrogate model can be found in Schobi et al. [64].

PCK can be interpreted as a universal Kriging model with a specific trend that composes a set of orthonormal polynomials [63]:

$$Y \approx \mathcal{M}^{PCK}(x) = \sum_{\alpha \in \mathcal{A}} \beta_\alpha \psi_\alpha(X) + \sigma^2 Z(x, \omega) \quad (11)$$

where $\sum_{\alpha \in \mathcal{A}} \beta_\alpha \psi_\alpha(X)$ is a weighted sum of orthonormal polynomials describing the trend of the PCK model, σ^2 and $Z(x, \omega)$ denote the variance and the zero mean, unit variance, stationary Gaussian process.

The construction of the PCK model is mainly divided into the following two parts: (1) determination of the truncated set of polynomials contained in the trend, and (2) calibration of the unknown Kriging parameters $(\beta, \sigma^2, \theta)$. These two parts can be combined in various ways, of which the most commonly used methods are sequential PCK (SPCK) and optimal PCK (OPCK) [63]. The former is adopted in this study. In SPCK, the set of polynomials and the Kriging surrogate model are determined sequentially. In the first step, the truncated set of polynomials \mathcal{A} in Eq. (2) is determined by sparse PCE based on LAR selection. Then, the truncated set \mathcal{A} is embedded into Eq. (11) as a trend that composed of $P = |\mathcal{A}|$ regressors. Finally, the PCK surrogate model is calibrated as a universal Kriging model.

2.1.4. Support vector regression

SVR was developed based on statistical learning theory by Vapnik [65]. The core idea is to use kernel functions to project the original data into high-dimensional feature space and search for the best prediction function in linear feature space. A large group of kernel functions can be used including non-stationary linear, polynomial, sigmoid, Gaussian, exponential, Matérn-3/2 and Matérn-5/2 [57,66]. This allows the SVR model to have significant generalization ability and nonlinear problem-handling capability, largely circumventing the over-fitting of the data.

Consider some pairs of inputs $x \in \mathcal{D}_X \subset \mathbb{R}^M$ and outputs $y = \mathcal{M}(x) \in \mathbb{R}$ given an experimental design $\mathcal{X} = \{x_1, \dots, x_n\}$ and the corresponding model responses $\mathcal{Y} = \{\mathcal{M}(x^{(1)}), \dots, \mathcal{M}(x^{(N)})\}^T$, a linear SVR model is formulated as:

$$\mathcal{M}^{SVR}(x) = \omega^T x + b = \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i^T x + b \quad (12)$$

where $\omega = \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i \in \mathbb{R}^n$ is a vector of weight coefficient, $\alpha_i \geq 0$ and $\alpha_i^* \geq 0$ are Lagrange multipliers, and $b \in \mathbb{R}$ is an offset parameter to be estimated.

For nonlinear problems, the input variables are mapped into a high or infinite dimensional feature space by the nonlinear transform $x \rightarrow \varphi(x)$. In this space, the linear prediction function in Eq. (12) can be extended as follows:

$$\mathcal{M}^{SVR}(x) = \omega^T \varphi(x) + b = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \varphi(x_i)^T \varphi(x) + b \quad (13)$$

where the inner product $\varphi(x_i)^T \varphi(x)$ is the so-called kernel function $k(x_i, x)$.

Lagrange multipliers α_i and α_i^* are computed by maximizing the Lagrangian function:

$$\begin{aligned} L(\alpha, \alpha^*) = & -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(x_i, x_j) \\ & - \sum_{i=1}^N (\alpha_i + \alpha_i^*) \varepsilon + \sum_{i=1}^N (\alpha_i - \alpha_i^*) y_i \\ \text{s.t. } & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \quad \text{and} \quad 0 \leq \alpha_i, \alpha_i^* \leq C, \quad i = \{1, \dots, N\} \end{aligned} \quad (14)$$

where $C \in \mathbb{R}^+$ is a regularization parameter for the regression problem, ε is the insensitive tube width which provides SVR with its sparsity property.

The Gaussian kernel is selected in this study which is the most popular form:

$$k(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right) \quad (15)$$

where $\sigma > 0$ is the hyper-parameter of the kernel function.

Once a kernel is chosen, the most important step is to properly fit the hyper-parameters $\gamma = \{C, \varepsilon, \sigma\}^T$ that lead to the most accurate model in terms of generalization. CMA-ES is adopted to calibrate the SVR model to find the hyper-parameters' optimal values similar to the Kriging models.

2.2. Bayesian inversion

In some cases, the unknown parameters in the system cannot be directly measured, or they do not have a clear physical meaning. In such a case, the inverse problem helps estimate those parameters indirectly from the experimental data. In the inverse problems, the information about the observations is propagated back to the computational model to obtain insight into the input parameters [67].

Most of the Bayesian inverse problems share the same ingredients [68,69]: a computational forward model \mathcal{M} , a set of input parameters $\mathbf{x} \in D_X$ that need to be inferred, and a set of QoIs \mathbf{y} . Assuming that the random input parameters vector $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ cannot be measured directly, we can only resort to measuring the QoIs. Let us consider N independent measurements \mathbf{y}_i are available and gathered in a data set $\mathcal{Y} \stackrel{\text{def}}{=} \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$. Since the forward model $\mathbf{x} \mapsto \mathcal{M}$ is based on several simplifications, a discrepancy term needs to be added to the computational model as $\mathbf{y} = \mathcal{M}(\mathbf{x}) + \varepsilon$. In practice, $\varepsilon \in \mathbb{R}^{N_{\text{out}}}$ represents the effects of measurement error and model inaccuracy and can be assumed to be an additive Gaussian discrepancy with zero mean value and unknown variance σ^2 .

In this setting, the unknown parameter vector \mathbf{x} consists of two parts, i.e., computational forward model parameters \mathbf{x}_M and discrepancy parameters \mathbf{x}_ε . Assuming that the two parts as being priorly independent, and a priori distribution $\pi(\mathbf{x}_\varepsilon)$ for the unknown variance σ^2 can be obtained, then their joint prior distribution can be formulated as:

$$\pi(\mathbf{x}) = \pi(\mathbf{x}_M) \pi(\sigma^2) \quad (16)$$

The likelihood function can thus be written as:

$$\begin{aligned} \mathcal{L}(\mathbf{x}_M, \sigma^2; \mathcal{Y}) &= \prod_{i=1}^N \frac{1}{\sqrt{(2\pi\sigma^2)^{N_{\text{out}}}}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y}_i - \mathcal{M}(\mathbf{x}_M))^T (\mathbf{y}_i - \mathcal{M}(\mathbf{x}_M))\right) \end{aligned} \quad (17)$$

Based on the prior distribution in Eq. (16) and likelihood function in Eq. (17), the corresponding posterior distribution can be formulated as:

$$\begin{aligned} \pi(\mathbf{x}_M, \sigma^2 | \mathcal{Y}) &= \frac{1}{Z} \pi(\mathbf{x}_M) \pi(\sigma^2) \mathcal{L}(\mathbf{x}_M, \sigma^2; \mathcal{Y}), \\ Z &= \int_{D_X} \mathcal{L}(\mathbf{x}_M, \sigma^2; \mathcal{Y}) \pi(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (18)$$

where the normalizing factor Z , known as the evidence or marginal likelihood, shall ensure that this distribution integrates to 1.

The posterior distribution summarizes the updated information about the unknown parameters ($\mathbf{x}_M, \mathbf{x}_\varepsilon$) based on the prior knowledge and the observed data, which is often characterized through its first statistical moments. The posterior mean vector, as well as the posterior covariance matrix (i.e., quantified uncertainty), are:

$$\mathbb{E}[\mathbf{X} | \mathcal{Y}] = \int_{D_X} \mathbf{x} \pi(\mathbf{x} | \mathcal{Y}) d\mathbf{x} \quad (19)$$

$$\text{cov}[\mathbf{X} | \mathcal{Y}] = \int_{D_X} (\mathbf{x} - \mathbb{E}[\mathbf{X} | \mathcal{Y}])(\mathbf{x} - \mathbb{E}[\mathbf{X} | \mathcal{Y}])^T \pi(\mathbf{x} | \mathcal{Y}) d\mathbf{x} \quad (20)$$

3. Proposed parameter identification algorithm: SASOI

The procedure of the proposed SASOI algorithm for concrete dams involves the following main steps:

Step A: Set probabilistic input *priori* model: Considering that the input parameters to be identified are difficult to measure, the stochastic theory is introduced to reduce the epistemic uncertainties related to the mechanical parameters of the materials. According to the properties of the input parameters, a suitable distributional model is selected for each one. Then, taking the empirical judgment value, or a few

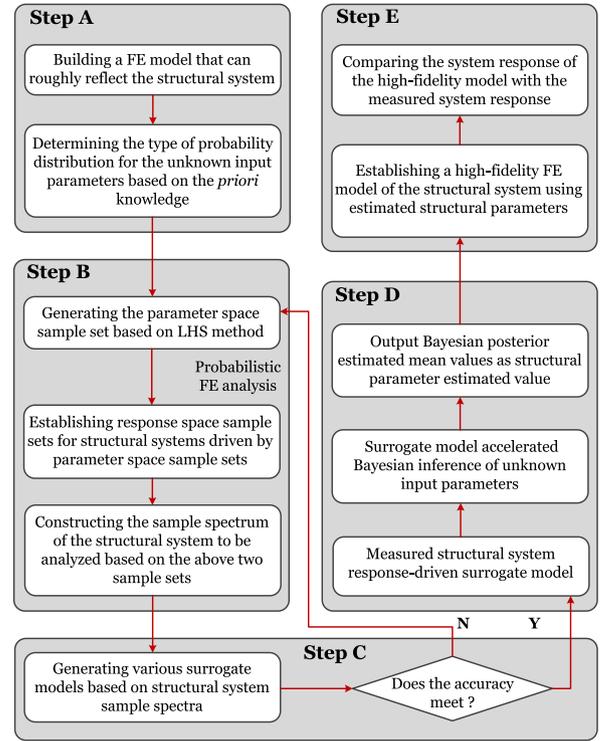


Fig. 1. Proposed SASOI flowchart for concrete dams; Step C is based on Section 2.1; Step D is based on Section 2.2; Step E uses formulation in Section 3.1.

measured values as the mean value of the distribution function and given the variance, the probabilistic input prior model of the unknown parameters can be constructed.

Step B: Choose a design of experiment (DOE): To construct the surrogate model, we first need to conduct an experimental design. There are many types of sampling techniques available for experimental design. In this study, the Latin hypercube sampling (LHS) method is selected, which has the advantages of uniformity and efficiency and is widely used in constructing a surrogate model. After generating a certain number of parameter space sample sets based on Step A and the LHS method, the probabilistic finite element analyses are performed to create the response space sample set. Finally, the sample spectrum of the structural system is constructed based on the above two sample sets.

Step C: Construct and evaluate the surrogate model: The surrogate model is constructed based on the parameter space sample sets and the response space sample sets generated in Step B. The accuracy of the surrogate model is evaluated by leave-one-out (LOO) cross-validation error, Err_{LOO} . This paper compares four surrogate modeling technologies (i.e., PCE, Kriging, PCK, and SVR).

Step D: Surrogate-assisted accelerated Bayesian inversion: The core of this step is to substitute the original high computational cost calculation (i.e., numerical) model with a cost-effective yet accurate surrogate model. The predicted QoIs of the surrogate model should be consistent with those from original finite element analyses, verified in the previous step. Next, the predicted QoIs by surrogate models are used to accelerate the Bayesian inference and estimate the Bayesian posterior mean of the input parameters.

Step E: Compare and evaluate the accuracy of the inversion algorithm: The original calculation model is updated based on the structural parameters estimated in step D. These values are then compared with the updated QoIs from measured or experimental observations.

This paper aims to establish a general procedure that rapidly identifies concrete dam parameters based on the SASOI algorithm through

the five steps above. This study will demonstrate the efficiency and accuracy of this algorithm and its feasibility through several simple-to-complex case studies. In the context of computational modeling and uncertainty quantification, inverse problems are the main classes of applications that benefit from Bayesian inference. Following the generic framework introduced by Nagel and Sudret [70], Wagner et al. [71], Fig. 1 illustrates a flowchart for the SASOI algorithm with application in dams.

The overall goal of this section is to answer two important questions: (1) Can the surrogate model be used as an alternative to the calculation model, that is, the accurate evaluation of the surrogate model; if yes, (2) How accurate is the SASOI algorithm? and how much computational time/cost can be saved using this algorithm? These questions are addressed based on the metric in Section 3.1.

3.1. Metrics to evaluate SASOI

For the first problem, the Err_{LOO} is selected to evaluate the accuracy of the surrogate model in this study, which is profitable to compare the performance of different surrogate models when evaluated on the same validation set [16]. First, only a single DOE point is retained at a time, and then, a meta-model $\mathcal{M}^{meta\setminus i}$ is constructed based on a reduced experimental design $\mathcal{X}\setminus\mathbf{x}^{(i)} = \{\mathbf{x}^{(j)}, j = 1, \dots, N, j \neq i\}$. Finally, comparing its prediction on the excluded point, $\mathbf{x}^{(i)}$ with the real value $y^{(i)}$, the error is calculated as:

$$Err_{LOO} = \frac{\sum_{i=1}^N (\mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{meta\setminus i}(\mathbf{x}^{(i)}))^2}{\sum_{i=1}^N (\mathcal{M}(\mathbf{x}^{(i)}) - \hat{\mu}_Y)^2} \quad (21)$$

where $\hat{\mu}_Y = \frac{1}{N} \sum_{i=1}^N \mathcal{M}(\mathbf{x}^{(i)})$ is the sample mean of the experimental design response.

The answer to the second question will be achieved by comparing the error between updated prediction QoIs and experimental or measured observation. Two error validation metrics are used in this paper including root mean square error (RSME) and mean absolute percentage error (MAPE):

$$RMSE = \left(\frac{1}{N} \sum_{k=1}^N (y_k - \tilde{y}_k)^2 \right)^{1/2} \quad (22)$$

$$MAPE = \frac{100}{N} \sum_{k=1}^N \frac{|y_k - \tilde{y}_k|}{y_k} \quad (23)$$

4. Validation of SASOI by simple benchmark problems

This section presents two basic benchmark problems to validate the SASOI algorithm for static and dynamic parameter identification situations: (1) a simply-supported beam and (2) a cantilever aluminum plate. The first example has an accurate analytical solution, and the second one benefits from experimentally measured values.

4.1. Static parameter identification of a beam

4.1.1. Construction of static parameter prior distribution space

This example evaluates the mid-span deflection of a simply-supported beam subjected to uniformly distributed load $p = 12,000$ N/m. The beam has a rectangular cross-section of width $b = 0.15$ m, height $h = 0.30$ m, and length $L = 5.0$ m. The beam consists of uniform material with a modulus of elasticity, E . The input parameters (b, h, L, p) are considered known parameters that obey a constant distribution. However, the E is assumed to be unknown input parameters with a lognormal distribution, and its prior distribution is $LN(30, 4.5)$ GPa. According to the beam theory, the analytical expression for the mid-span deflection of a simply-supported beam is $V_{mid} = \frac{5 p L^4}{32 E b h^3}$. This simple formula serves as the original calculation model and connects the unknown modulus of elasticity to the measurable mid-span deflection.

The best source to guess the initial distribution of an unknown parameter is to use the data in the literature that already has been calibrated to the experimental tests [72]. Two types of distributional models are widely used in structural engineering (in general) and dam engineering (in particular): truncated normal distribution, and the lognormal distribution. The former one has the highest entropy and is the simplest choice in case there is no evidence so far on experimental results [73,74]. The latter model is more reasonable for mechanical properties of the material (e.g., mass density) as the lower tail is bounded to zero which indicates a physical property cannot take a negative value [75].

A set of $N = 5$ independent experiments are carried out on this beam, and the measurements are [71]: 12.84, 13.12, 12.13, 12.19, and 12.67 mm. Due to measurement error, the measured deflections vary across experiments. We can calculate the mean of mid-span deflection as $\bar{V}_{mid} = 12.59$ mm. The corresponding modulus of elasticity from the analytical solution is $\bar{E} = 22.98$ GPa.

4.1.2. Results and discussions

This section presents the results of the surrogate model and parameter identification based on the SASOI algorithm in the simply-supported beam model. The unknown input parameter is sampled N_{sim} times by the LHS method, and the corresponding QoI(s) are extracted using the original calculation model. The data set for constructing the surrogate model consists of two parts: the data set for the design of the experiment, N_{DOE} , and the data set for validation, N_{val} , both belong to a subset of N_{sim} .

All surrogate model construction and Bayesian inference have been carried out with the open-source software UQLab [76], which allows for easy reproducibility. The numerical computing environment in this study is based on a high-performance UNIX workstation, which has two nodes; each node has 36-core CPU and 192 GB of memory. The calculation software adopts MATLAB [77] and ANSYS APDL [78] and calls 12-core CPU to perform the probabilistic FE analysis.

Intuitively, for larger N_{DOE} , a more accurate surrogate model is expected. As a rule of thumb, the surrogate model can be a perfect alternative for the original computational model when its accuracy reaches $1e^{-3}$. While higher accuracy (i.e., smaller error term) is desirable, it may not be beneficial/efficient from a computational cost point of view. Therefore, keeping the surrogate model balanced in terms of accuracy and cost is essential.

Different surrogate models are constructed with varying N_{DOE} and fixed N_{val} in this study to explore the optimal value needed to be used for surrogate models. A total of $N_{sim} = 100$ initial samples are drawn for the unknown parameter E using the LHS method, and the corresponding mid-span deflections V_{mid} are then calculated. For each of the surrogate algorithms, six meta-models are constructed with different sizes of N_{DOE} (i.e., 5, 10, 20, 30, 40, and 50) and the same size of N_{val} . This results in 24 surrogate models (i.e., four techniques and six sample sizes). In each case, the Err_{LOO} is shown in Fig. 2(a), and the following observations can be drawn:

- While the error term Err_{LOO} for the SVR model decreases with the increase of N_{DOE} , the required accuracy (i.e., less than 5%) is not achieved until $N_{DOE} = 30$.
- Aside from the SVR meta-model, the other three surrogate models have very high accuracy for the entire range of N_{DOE} . The accuracy of PCE and PCK is better than the Kriging-only model.
- For SVR and Kriging surrogate models, increasing the size of N_{DOE} increases the accuracy continuously. For the PCE and PCK surrogate models, there is a clear drop from $N_{DOE} = 5$ to 10; however, Err_{LOO} is small enough that the meta-model reaches saturation and does not change for N_{DOE} greater than 10. One may note that this is an elementary analytical study with only one unknown variable.

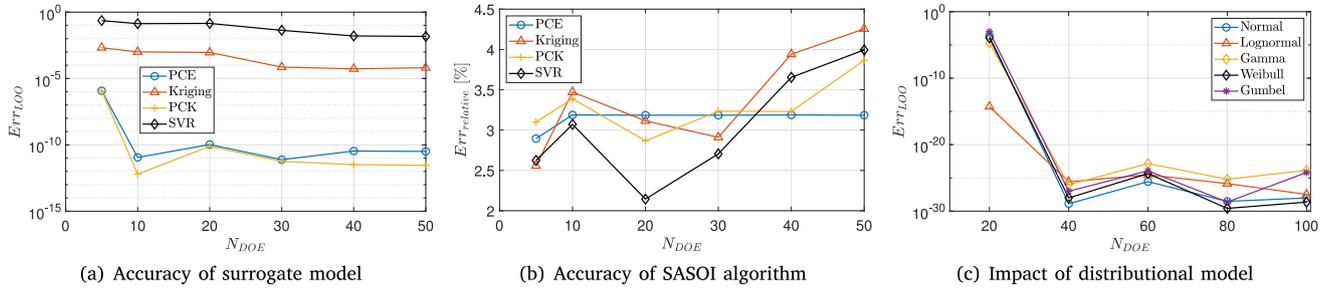


Fig. 2. Simply-support beam example with variable N_{DOE} , four surrogate algorithms, and multiple assumed distributional models.

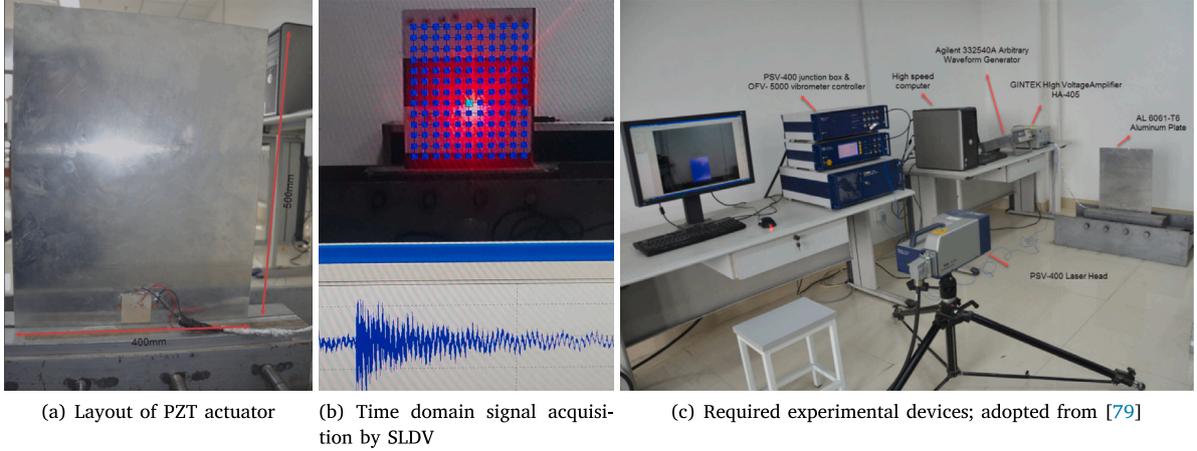


Fig. 3. Experimental setup for modal analysis of a cantilever aluminum plate.

Next, the developed surrogate models are integrated into the SASOI algorithm for parameter identification of simply-support beam with accelerated Bayesian inference of the unknown parameter E . Using the obtained \bar{V}_{mid} , the relative error is computed, and the accuracy of the SASOI algorithm is evaluated. Once the SASOI algorithm is combined with each of the four surrogate models, the following notations specify each combination: PCE_SOI, Kriging_SOI, PCK_SOI, and SVR_SOI. The results are shown in Fig. 2(b) and Table 1 with the following major observations:

- There is no positive correlation between the accuracy of the SASOI algorithm and the accuracy of its corresponding surrogate model. According to Fig. 2(a), the SVR surrogate model had unreliable accuracy for $N_{DOE} < 30$; however, the accuracy of SVR_SOI is still comparable to the other three.
- The robustness of the PCE_SOI algorithm is better than the other three since the posterior value of parameter E and mid-span deflection V_{mid} tend to be constant for $N_{DOE} \geq 10$.
- Although there are some differences in the accuracy of the SASOI algorithms, the accuracy varies 2%–5% in all cases, which meets the acceptance criteria for this simple structure.

While the unknown parameter in this example, E , was assumed to follow a Lognormal distribution, it is interesting to see if the choice of initial distributional model is important in the SASOI algorithm. For this purpose, four other distributional models are also tested, i.e., Normal, Gamma, Weibull, and Gumbel. All five distributional models are used in conjunction with the PCE model only because it is the best surrogate model. Fig. 2(c) shows the variation of Err_{LOO} as a function of N_{DOE} . As seen, there is no meaningful difference among the results which indicates the SASOI is nearly independent of the choice of initial distributional model. Indeed, the original Lognormal model shows a slightly better error metric at $N_{DOE} = 20$ compared to the other four.

Table 1

Comparison of the relative error between the posterior value of V_{mid} and its measured mean \bar{V}_{mid} in [mm].

N_{DOE}	5	10	20	30	40	50
PCE_SOI	12.22	12.19	12.19	12.19	12.19	12.19
	2.90%	3.18%	3.18%	3.18%	3.18%	3.18%
Kriging_SOI	12.27	12.15	12.2	12.22	12.09	12.05
	2.57%	3.49%	3.12%	2.91%	3.95%	4.27%
PCK_SOI	12.2	12.16	12.23	12.18	12.18	12.1
	3.11%	3.39%	2.87%	3.23%	3.23%	3.87%
SVR_SOI	12.26	12.2	12.32	12.25	12.13	12.1
	2.62%	3.07%	2.14%	2.71%	3.65%	4.00%

4.2. Dynamic parameter identification of a plate

4.2.1. Model and experiment setup

The cantilever aluminum plate shown in Fig. 3 is the second verification example for dynamic parameter identification. The AL6061-T6 aluminum alloy flat plate has an elastic modulus of 710 MPa, a mass density of 2700 kg/m³, and the dimensions of 600 × 400 × 3 mm³. The plate is clamped at one end with a clamped depth of 100 mm by a steel anchor beam (See Fig. 3(a)), leading to a cantilever plate of 500 × 400 × 3 mm³.

As shown in Fig. 3(a), a piezoelectric lead zirconate-titanate (PZT) actuator with the dimension of 50 × 50 × 1 mm³ is surface-bonded at the plate width center near the clamped plate base. It is used to excite the plate using a sweep sine excitation ranging from 100 to 2000 Hz by Agilent® 33250 A wave signal generator. To obtain a detectable excitation and response, the Pintek® HA-405 high voltage amplifier is connected with the excitation circuit before the swept-frequency signal

Table 2
Material properties for the cantilever aluminum plate.

Parameters	Symbol	Unit	Model	Quantity
Elastic modulus	E	MPa	Lognormal	$LN(710, 71)$
Mass density	ρ	kg/m ³	Lognormal	$LN(2700, 270)$
Poisson's ratio	ν	–	–	0.3

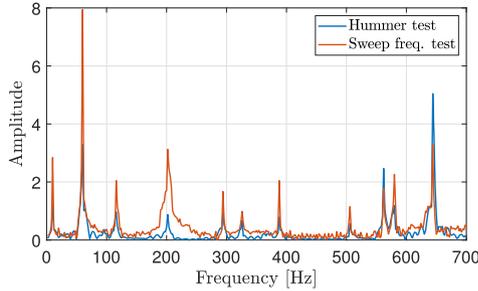


Fig. 4. Frequency response function of cantilever aluminum plate based on hammer test and sweep frequency test.

is transmitted to the PZT actuator, and the voltage amplitude is 200 Vpp [79]. Due to the incomparable advantage of the Scanning Laser Doppler Vibrometer (SLDV) in collecting the dynamic response of the structure, the plate vibration response is excited by hammering and frequency sweeping and then measured using the Polytec® PSV-400 SLDV. The natural frequency time-domain signal acquisition process of the cantilever plate is shown in Fig. 3(b), and the experimental setup for modal analysis of a cantilever plate is illustrated in Fig. 3(c).

4.2.2. Construction of prior distribution space

In this benchmark case, it is assumed that the dynamic elastic modulus and mass density of the cantilever aluminum plate are unknown input parameters, and its material properties are shown in Table 2. LHS-based random sampling is used to construct the dynamic parameter prior distribution space. The finite element model of the cantilever plate is developed using shell elements, and a subset of the computed natural frequencies of the structure is extracted by modal analysis as QoIs. The numerical calculation environment in this study is to that explained in Section 4.1.

4.2.3. Comparison and remarks

The vibration time-domain signal of the cantilever plate is acquired by two experimental modal analysis (EMA) methods: the hammering method and sweep frequency method and the frequency response function (FRF) is obtained based on the FFT method. According to Fig. 4, there is an acceptable consistency between the two methods, and both can identify the frequencies with peak amplitude (e.g., 10 Hz, 60 Hz, 115 Hz, 201 Hz, 293 Hz, ...). The first ten peak frequencies are extracted and used as the measured QoIs.

A total of $N_{sim} = 120$ samples are drawn using the LHS method, and ten sets of calculated natural frequencies corresponding to the measured values are computed based on probabilistic finite element analysis. Subsequently, five batches with different N_{DOE} sizes (i.e., 15, 25, 50, 75, and 100) and a fixed $N_{val} = 20$ are used to build the surrogate models. The resulted Err_{LOO} is shown in Fig. 5 with the following major observations:

Fig. 5(a) evaluates the surrogate models for varying numbers of N_{DOE} and three QoIs corresponding to the first three natural frequencies. The error term Err_{LOO} of the SVR model decreases by increasing

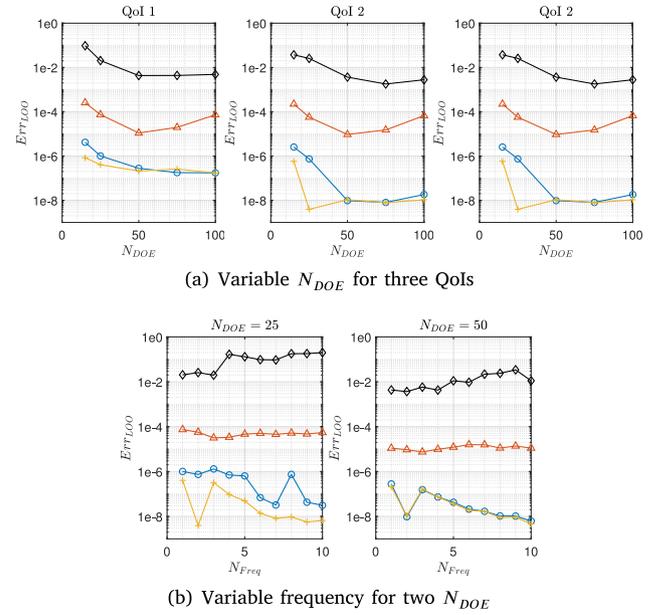


Fig. 5. Quantification of Err_{LOO} in cantilever plate based on different surrogate models.

Table 3
Parameter identification results for cantilever plate.

Parameters	Unit	PCE_SOI	Kriging_SOI	PCK_SOI	SVR_SOI
E	MPa	708.05	708.95	703.33	697.32
ρ	Kg/m ³	2725.13	2728.5	2707.59	2682.34

the size of N_{DOE} . It yields to a reliable accuracy for $N_{DOE} \geq 50$. However, the other three surrogate models always achieve high accuracy regardless of N_{DOE} size. PCK and PCE illustrate the best performance.

Fig. 5(b) compares all first ten frequencies for two N_{DOE} of 25 and 50. Comparing four surrogate models, again, SVR is the worst model, while PCK and PCE are the best ones. For all four surrogate models, increasing the N_{DOE} from 25 to 50 decreases the Err_{LOO} error considerably. Comparing the results of different frequencies reveals three different behaviors: using the Kriging model, the Err_{LOO} error is nearly constant along with the frequency number. PCK and PCE have some fluctuations, but the error is generally reduced by increasing the frequency number. However, using the SVR surrogate model, the Err_{LOO} error increases by an increase in frequency number. Indeed, the accuracy of the SVR model with $N_{DOE} = 25$ for the first three frequencies is completely different from the other seven.

Hereafter, four surrogate models with $N_{DOE} = 50$ are used to combine with Bayesian inference to identify the structural unknown input parameters. The results of dynamic parameter identification are shown in Table 3. Subsequently, by comparing the identified values of the unknown input parameters with the given calibration values, in this case, See Fig. 6(a), one can observe the inversion accuracy of the SASOI algorithm, which is very high.

In addition, Fig. 6 illustrates the results of surrogate models using two metrics. By comparing the frequency curve composed of the first 160 calculated natural frequencies of the cantilever plate, it is clear that the priori and posteriori frequency curves (i.e., before and after the parameter identification) are identical. This plot also shows the measured natural frequency values distributed on the curve. Detailed comparison of measured values and each of four surrogate models for all ten frequencies are summarized in A.9. This example verifies the accuracy of the SASOI algorithm for structural dynamic parameter identification.

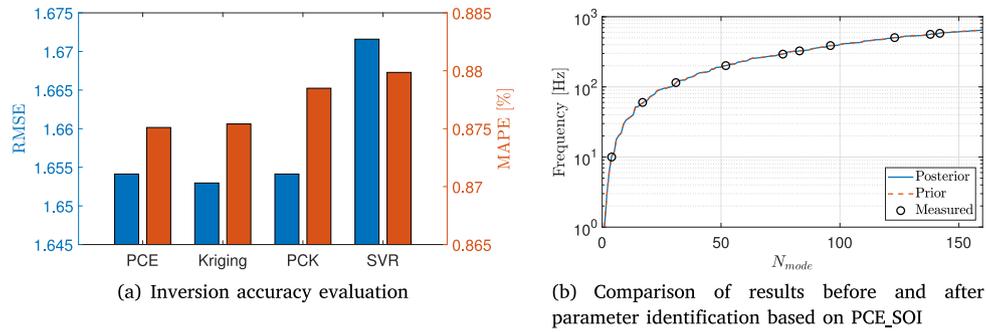


Fig. 6. Comparison of accuracy of SASOI algorithms based on various surrogate models; $N_{DOE} = 50$ and $N_{val} = 20$.

5. Application of SASOI in dam engineering

So far, the proposed SASOI algorithm has been verified for both the static and dynamic parameter identification problems using simple structural models. In this section, the same algorithm is applied to two examples of concrete arch dams. The first one is to identify the static parameters in an arch dam with measure data, while the second one explores the dynamic characteristics using the small-scale experimental program.

5.1. Static parameter identification of Dayakou Dam

5.1.1. Engineering background

Dayakou Dam is a 95 m high double curvature arch dam. See Fig. 7(a). The width of the dam at the crest and the bottom are 5.0 m and 22.0 m, respectively. The thickness-to-height ratio of the crown cantilever is 0.232. There are five sections in the dam with four transverse joints arranged with a spacing of about 60 m. Three overflow surface holes are placed in the middle of the dam crest, with a weir crest elevation of 643.5 m (the bottom elevation is 558 m), and the net width of each hole is 10 m. The dam was completed in November 2015 and began to store water in February 2017.

A high-fidelity finite element model is developed for the dam and its surrendering foundation with as many details as possible. The finite element model is based on hexahedron and tetrahedron isoparametric elements in ANSYS, as shown in Fig. 7(b). The foundation model is extended 1.5 and 2.5 times the dam height in the upstream and downstream directions. It is also extended twice the dam height to the left, right, and bottom.

Fig. 7(b) also illustrates six material zones in the arch dam model, while their basic parameters are reported in Table 4. The position of the measurement points to monitor the arch dam deformation is shown in Fig. 7(c). The sensors on the positive vertical line are arranged to be measured at points 1 and 2, and the sensor on the inverted vertical line is arranged to be recorded at point 3. By analyzing the absolute displacement values of three recorded data on the A1 measurement line in the middle of the dam, the measured displacement along the river, U_y , can be obtained.

This example aims to identify the static parameters of the dam subjected only to water pressure after the dam construction. Therefore, the influence of seepage on the deformation is not significant. The initial reservoir water level (at the time of storage initiation) was 570 m. Then, it was increased to the normal water level of 648 m in July 2017. By processing a large amount of water and air temperature data during the impounding period, the displacement monitoring data along the river on June 15 and July 15, 2017, are selected as the measured displacement. The corresponding upstream reservoir water levels are 647.3 and 648.0 m, respectively, while the downstream water level is 571 m. Using this information, the hydrostatic pressure on the dam is calculated. One should note that the temperature difference

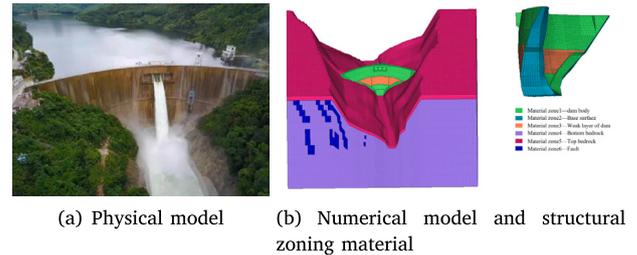


Fig. 7. Description of Dayakou arch dam example.

Table 4

Basic parameters for Dayakou arch dam high-fidelity finite element model.

Zone	Number of elements	Density [kg/m ³]	Elastic modulus [GPa]	Poisson's ratio [-]	Coefficient of thermal expansion [1/°C]
1	47,788	2400	20	0.19	7.0×10^{-6}
2	2304	2400	20	0.19	7.0×10^{-6}
3	4040	2400	16	0.21	7.0×10^{-6}
4	129,569	-	12	0.22	-
5	22,160	-	6	0.26	-
6	4983	-	5	0.3	-

between the inside and outside of the concrete dam during the construction phase will significantly impact the stress distribution. The applied loads are the dam self-weight, hydrostatic pressure (from upstream and downstream reservoirs), and thermal load.

Table 5
Material properties for Dayakou arch dam prior distribution space.

Parameter	Symbol	Unit	Model	Quantity
Zone 1; Elastic Modulus	E_1	GPa	Normal	$N(20, 2)$
Zone 1; Coeff. Thermal Exp.	α_1	$1/^\circ\text{C}$	Normal	$N(7.0 \times 10^{-6}, 0.70 \times 10^{-6})$
Zone 2; Elastic Modulus	E_2	GPa	Normal	$N(20, 2)$
Zone 2; Coeff. Thermal Exp.	α_2	$1/^\circ\text{C}$	Normal	$N(7.0 \times 10^{-6}, 0.70 \times 10^{-6})$
Zone 3; Elastic Modulus	E_3	GPa	Normal	$N(20, 2)$
Zone 3; Coeff. Thermal Exp.	α_3	$1/^\circ\text{C}$	Normal	$N(7.0 \times 10^{-6}, 0.70 \times 10^{-6})$
Zone 4; Elastic Modulus	E_4	GPa	Normal	$N(12, 1.2)$
Zone 5; Elastic Modulus	E_5	GPa	Normal	$N(6, 0.6)$
Zone 6; Elastic Modulus	E_6	GPa	Normal	$N(5, 0.5)$

Table 6
Parameter identification results for Dayakou Dam.

Parameter	Unit	PCE_SOI	Kriging_SOI	PCK_SOI	SVR_SOI	BO
E_1	GPa	26.38	24.853	22.002	25.371	22.711
α_1	$1/^\circ\text{C}$	6.062×10^{-6}	5.740×10^{-6}	7.096×10^{-6}	5.653×10^{-6}	6.379×10^{-6}
E_2	GPa	22.077	16.703	23.619	18.254	22.406
α_2	$1/^\circ\text{C}$	2.764×10^{-6}	6.988×10^{-6}	6.710×10^{-6}	6.652×10^{-6}	5.985×10^{-6}
E_3	GPa	21.153	17.926	18.102	19.301	18.219
α_3	$1/^\circ\text{C}$	7.867×10^{-6}	6.382×10^{-6}	7.258×10^{-6}	6.309×10^{-6}	6.618×10^{-6}
E_4	GPa	5.086	7.181	3.734	5.913	10.259
E_5	GPa	4.814	5.061	4.125	5.541	5.192
E_6	GPa	4.983	5.091	6.576	5.11	4.27

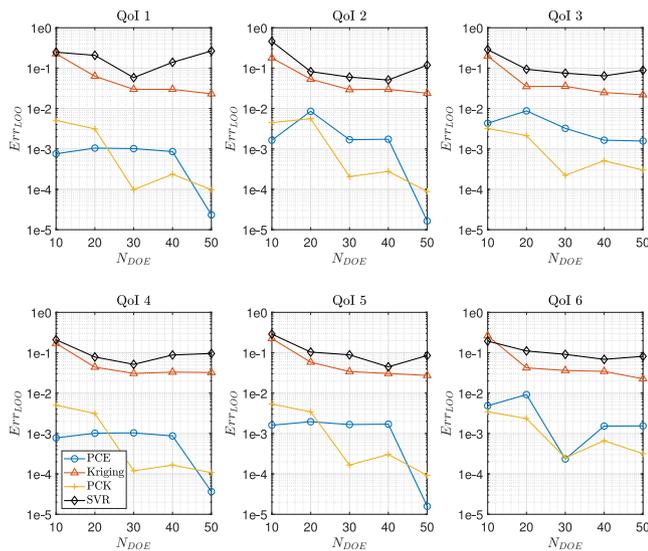


Fig. 8. Quantification of Err_{LOO} in Dayakou Dam example based on different surrogate models; Note: first three QoIs correspond to 2017/6/15 and other three correspond to 2017/7/15.

5.1.2. Construction of static parameter prior distribution space

In this example, the elastic modulus and coefficient of thermal expansion for the concrete and rock are assumed to be unknown input parameters and follow a normal distributional model. The mean value of the actual survey is used as the mean of *priori* distribution, and all other material parameters are kept constant in the finite element model. The material properties for each partition in this finite element model are shown in Table 5. The LHS-based method is used to draw the samples from unknown input parameters and perform probabilistic finite element simulations. The calculated displacement along the river at the target points is extracted as QoI to be used in the surrogate models, and the numerical calculation environment is the same as in Section 4.1.

5.1.3. Comparison and discussion

The SASOI algorithm is applied to identify the unknown input parameters in the dam. The first step towards SASOI implementation is to build a reliable surrogate model to be replaced with the original finite element model. Therefore, five groups of N_{DOE} with different sizes (i.e., 10, 20, 30, 40, and 50) are used to construct the surrogate models, and their Err_{LOO} is shown in Fig. 8. Six QoIs are discussed in this example, in which QoIs 1 to 3 are calculated displacement on 2017/6/15 along the river corresponding to three target locations. In addition, QoIs 4 to 6 present similar data calculated on 2017/7/15 (one month later). Technically, this operation can be expanded to other time intervals, which this paper ignores. The following observations can be drawn:

- For the given N_{DOE} , PCE and PCK models always have high accuracy, with error Err_{LOO} varying from $1e-4$ to $1e-2$.
- The error term Err_{LOO} for the Kriging model decreases by increasing the size of N_{DOE} , and the acceptable accuracy is reached when N_{DOE} is greater than 20. For one case (i.e., QoI 6), the accuracy of the Kriging model is worse than SVR for $N_{DOE} = 10$.
- The error term Err_{LOO} for the SVR surrogate model decreases first and then increases with the optimal N_{DOE} value of about 30–40. This can be attributed to the inadaptability of the SVR model to small data sets. Overall, the accuracy of the SVR model never reaches below the required threshold (i.e., 0.05). So, this method should be used with caution for complex dam models.

Based on the above discussion, we used the surrogate models based on $N_{DOE} = 30$ and $N_{val} = 20$ to build Bayesian inference of uncertain parameters. We will also compare the proposed SASOI algorithm with the classical inversion algorithm (based on iterative optimization) in terms of efficiency and accuracy. The Bayesian optimization (BO) algorithm is a step-wise iterative algorithm based on the Bayesian posterior distribution theory to find the global minima in the objective function. It does not require as much random search as the random group algorithm and has good robustness and fast convergence speed. Therefore, the BO algorithm is chosen for iterative optimization in this section and is implemented by real-time joint simulation of MATLAB and ANSYS. Two error metrics (i.e., MAPE and RMSE) are used to evaluate the accuracy of these inversion algorithms. Detailed results for different target points and surrogate methods are summarized in Table A.10. The results of parameter identification are summarized in Table 6, while the model accuracy is discussed in Fig. 9(a).

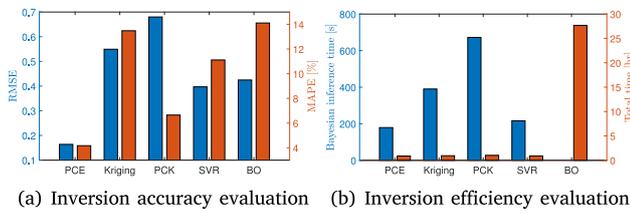


Fig. 9. Comparison of accuracy and efficiency between SASOI and BO-based inversion algorithms for Dayakou Dam.

Different error validation metrics for the same method may have inconsistent results. This necessitates using more than one metric to confirm the reliability of the results. PCE_SOI is the best model using both metrics. However, the PCK_SOI algorithm holds the second rank based on MAPE, while it has the lowest accuracy according to RMSE. The BO-based conventional inversion algorithm has mediocre performance among the five inversion methods. According to MAPE, the SASOI algorithms' accuracy is better than the traditional inversion algorithm, i.e., BO. However, in terms of RMSE, the BO algorithm competes with Kriging_SOI, PCK_SOI, and SVR_SOI. To ensure the accuracy of surrogate models compared to the conventional BO, it is recommended to use multiple surrogate techniques, and choose the one with the lowest error metrics.

Fig. 9(b) further discusses the inversion efficiency of both SASOI and BO algorithms. The total time consumed by the SASOI algorithm consists of two parts: (1) the computation time to construct the surrogate model data set and (2) the computation time for Bayesian inference of unknown input parameters. Based on the numerical environment described above, the calculation time of the first part to build four surrogate models is about 0.866 hr. However, the required time to make the Bayesian inference varies among the surrogate models:

- In terms of Bayesian inference time, the PCE_SOI algorithm has the best performance among the four surrogate models. Its Bayesian inference time is about one-third of PCK_SOI. Although the Bayesian inference time using SVR_SOI is close to PCE_SOI, the Err_{LOO} of SVR is much higher than PCE.
- The total time of the different SASOI algorithms does not differ much since the computation time of the first part is identical in all of them and much larger than the second part. However, the total time of the traditional inversion algorithm is 27.66 h, almost 30 times that of the SASOI algorithm.

In summary, the SASOI algorithm provides similar results to the traditional BO-based inversion algorithm in terms of accuracy; however, it is much faster. It is found that the SASOI algorithm solves the contradiction that the traditional inversion algorithm cannot balance computational accuracy and efficiency. In addition, the comprehensive analysis shows that the PCE_SOI algorithm is the best choice among the five methods in terms of accuracy and efficiency.

5.2. Dynamic parameter identification of a small-scaled dam

5.2.1. Model and experimental setup

This example aims to explore the capability of the SASOI algorithm to be used for dynamic parameter identification of small-scaled laboratory models. Therefore, a simplified scaled model of an ultra-high arch dam is built. The height of the scaled arch dam is 1.35 m, while the prototype dam is a 270 m high dam (i.e., the scale is 1:200). See Fig. 10(a). This model includes the dam and the foundation. The foundation is completely poured with commercial C30 concrete; however, the material of the dam body is modulated according to the principle of material ratio in the scale experiment.

Table 7

Material properties for the small-scaled experimental arch dam.

Parameter	Symbol	Unit	Model	Quantity
Dam modulus of elasticity	E_c	GPa	Normal	$N(1.5, 0.15)$
Dam mass density	ρ_c	kg/m ³	Normal	$N(2200, 220)$
Dam Poisson's ratio	ν_c	–	–	0.22
Foundation modulus of elasticity	E_f	GPa	Normal	$N(30, 3)$
Foundation mass density	ρ_f	kg/m ³	Normal	$N(2400, 240)$
Foundation Poisson's ratio	ν_f	–	–	0.19

Table 8

Parameter identification results for small-scaled experimental dam.

Parameter	Unit	PCE_SOI	Kriging_SOI	PCK_SOI	SVR_SOI
E_c	GPa	1.43	1.39	1.36	1.34
ρ_c	kg/m ³	2365.36	2333.50	2239.07	2109.60
E_f	GPa	27.05	28.04	26.87	27.01
ρ_f	kg/m ³	2446.54	2282.47	2438.14	2437.29

This figure also shows the required instrumentation for experimental modal analysis (EMA) of the scaled arch dam. There are eight 1A202E[®] low-frequency piezoelectric acceleration sensors on the dam body, of which seven are arranged at equal intervals on the crest and one at the upstream dam face. The vibration response of the model is obtained by hammer test, and the time-domain signal of the model system response is collected by DH5972N[®] online monitoring and analysis system.

5.2.2. Construction of dynamic parameter prior distribution space

Considering that the dynamic elastic modulus and mass density of the dam body and the foundation are unknown input parameters obeying a normal distribution, the measured mean values of the sampled specimen during concrete pouring were used as the average of the priori distribution. These material properties are listed in Table 7. Fig. 10(b) illustrates the finite element model of the simplified small-scaled arch dam. The finite element model of the small-scaled arch dam is constructed using the hexahedron and tetrahedron isoparametric elements. The QoIs are extracted in the form of natural frequencies under computational modal analysis, and the numerical calculation environment in this study is the same as in Section 4.1.

5.2.3. Comparison and remarks

The frequency response function of the model is obtained based on the FFT method, as shown in Fig. 11. This section extracts the first eight peak frequencies as the measured natural frequencies. These QoIs are tabulated in Table A.11.

The LHS method is used to sample the unknown input parameters $N_{sim} = 100$ times, and then the initial simulation data set is generated based on the same approach as discussed in Section 4.2. The following five batches with different sizes of N_{DOE} (i.e., 10, 20, 40, 60, and 80) and $N_{val} = 20$ are used to construct the surrogate models. The Err_{LOO} error is shown in Fig. 12. According to this figure, for the given N_{DOE} , PCE and PCK surrogate models always have higher accuracy compared to the other two models. This is, indeed, consistent with the observations of the previous three examples. The Err_{LOO} of the Kriging model decreases with the increase of N_{DOE} size. The PCE and PCK models even show better performance for lower frequencies, while the performance of the Kriging model is more or less constant with some fluctuations. The performance of the SVR model is noisy, and it seems that the error term is increased for higher frequency numbers. When $N_{DOE} = 60$ or higher, Err_{LOO} of the SVR model is acceptable for all frequencies except the last two ones.

Considering that the dynamic characteristics of the structures are mainly governed by the low-order modes [80], four surrogate models with $N_{DOE} = 60$ are selected to be the pilot model to accelerate the Bayesian inference of the unknown input parameters. The detailed



Fig. 10. Description of the small-scaled arch dam.

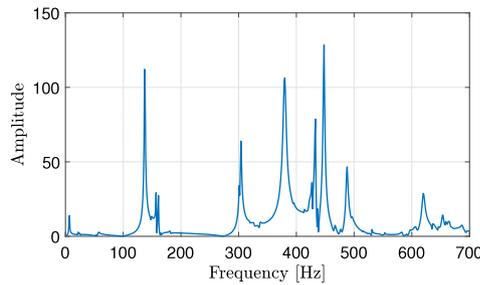


Fig. 11. Frequency response function of a small-scaled experimental arch dam by hammer test.

results for parameter identification are shown in Table 8. In addition, to explore the influence of different surrogate models on dynamic parameter identification results, this section selects N_{DOE} of 10 and 80 to be compared with 60. Results are shown in Figs. 13(a) and 13(b) with the following observations:

- According to MAPE and RSME, when $N_{DOE} = 10$, the accuracy of the SVR_SOI algorithm is significantly lower than others.
- When $N_{DOE} = 60$ or 80, the accuracy of the four SASOI algorithms is very similar. MAPE varies in [0.025, 0.028], while RSME falls in Alves and Hall [7], Tsitsiklis et al. [9]. Although the values for RSME seem to be high, they follow the same pattern for the true and predicted values, as reported in Table A.11.
- Similar to the previous observation, it seems that the accuracy of the SASOI algorithms does not correlate with N_{DOE} size (i.e., the accuracy of its corresponding surrogate model).

Finally, Fig. 14 illustrates the results of the Bayesian inversion for the prior and posterior samples based on PCE_SOI and $N_{DOE} = 60$, as well as the frequency responses. Comparing the frequency curves for the first 30 natural frequencies, one can observe that the measured values are close to the posterior distribution. Therefore, the SASOI algorithm has good inversion performance.

6. Conclusions

To timely identify the abnormalities in dams under the changing environment, the inversion analysis from measured data can play a crucial role. However, this procedure is mainly based on deterministic analysis combined with engineering judgment. The computational burden of numerical analysis is the major obstacle in the rapid identification of dam parameters. In this study, the SASOI algorithm was presented as an effective technique for unknown parameter identification of dams. Four benchmark problems with different complexity were studied which cover both the static and dynamic cases, as well as numerical, experimental, and field measurements: (1) static parameter identification for

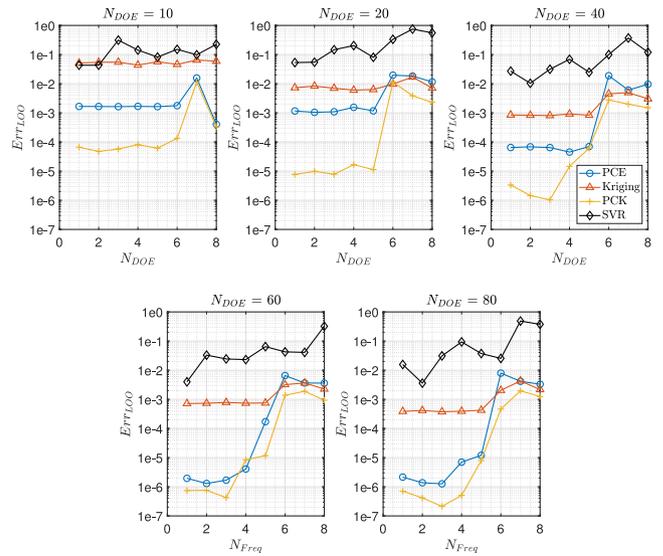


Fig. 12. Quantification of Err_{LOO} based on different surrogate models for small-scaled arch dam.

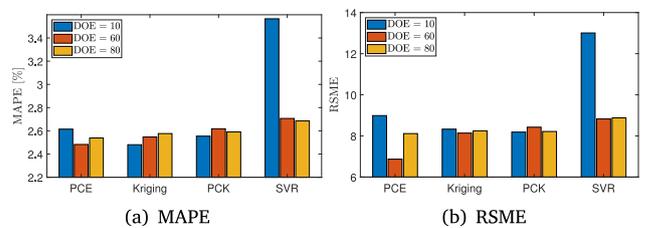


Fig. 13. Inversion accuracy evaluation based on different methods for small-scaled arch dam.

simply supported beam, (2) static parameter identification for Dayakou arch dam, (3) dynamic parameter identification for cantilever aluminum plate, and (4) dynamic parameter identification for small-scale arch dam experimental model.

Four types of surrogate models were used to analyze the relatively big database of various benchmark problems, and their capability in response prediction was evaluated using the leave-one-out error metric. The surrogate models were combined with accelerated Bayesian inference models, and the accuracy of the proposed SASOI algorithm was also evaluated using root mean squares error and mean absolute percentage error. The general conclusions are summarized as follows:

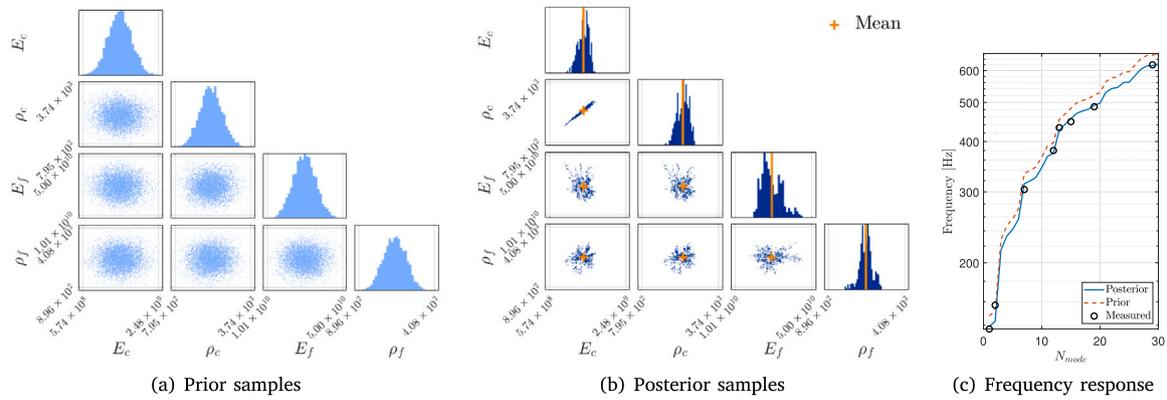


Fig. 14. Results of the Bayesian inversion on the prior and posterior sample with PCE_SOI and $N_{DOE} = 60$ for the small-scaled arch dam.

Table A.9 Comparison of measured and predicted values for cantilever plate.

Mode	Measured frequency [Hz]	PCE_SOI	Kriging_SOI	PCK_SOI	SVR_BOI
1	10	10.3	10.3	10.3	10.3
2	60	58.54	58.54	58.53	58.55
3	115	113.91	113.91	113.89	113.93
4	201	201.11	201.11	201.08	201.16
5	293	289.32	289.31	289.28	289.31
6	326	325.93	325.92	325.88	325.92
7	388	386.86	386.85	386.8	386.74
8	504	503.34	503.33	503.26	503.32
9	561	561.64	561.62	561.55	561.56
10	581	583.87	583.85	583.78	583.94
MAPE		0.875%	0.875%	0.875%	0.88%
RMSE		1.654	1.653	1.654	1.672

- In terms of the accuracy of the surrogate model: PCE and PCK models have higher performance. Even for a small experimental design data set, they always have high accuracy. The performance of the Kriging model is unstable under a small data set, but it improves by increasing the number of experimental points used to generate the surrogate model. As for the SVR model, which is not a suitable meta-model for small data sets with poor accuracy, its performance is inferior to that of the four surrogate models if only small data sets are used.
- In terms of the accuracy of the SASOI algorithm: Although the accuracy of four meta-models is similar and meets the requirements of engineering error control, the overall performance of the PCE_SOI is better. In addition, dynamic parameter identification has higher accuracy than the static one.
- The dam engineering applications showed that the total calculation time of the SASOI algorithm is similar for all surrogate models. However, compared with the classical inversion algorithm based on iterative optimization, the calculation efficiency of the proposed method is improved by about 27 times without affecting the accuracy.
- There is no correlation between the accuracy of the surrogate model and the accuracy of the SASOI algorithm.
- In general, it is recommended to use the adaptive sampling [81] to achieve the desired level of accuracy with the minimum possible sampling. Otherwise, the recommendation in this paper can be used to set up the size of DOE for dam engineering problems. Other generic recommendations can be found in Diaz et al. [82], Lin et al. [83] for the DOE size, such as $\frac{N \times (N+1)}{2}$ or $10N$, where N is the number of input parameters.

In summary, the SASOI algorithm presented in this study makes up for the shortcomings of the traditional inversion method in terms of computational accuracy and efficiency. It is an appropriate algorithm for the identification of the dynamic parameters in structural systems with random materials properties. The proposed algorithm can be used for any other infrastructure system subjected to static and dynamic loading. There are a few limitations and drawbacks to this algorithm:

- Construction of the surrogate model requires running the initial computational model several times. So, the efficiency of the entire algorithm depends on initial probabilistic simulation, and subsequent surrogate and optimization (i.e., SASOI). If the initial simulations are very heavy to compute, the efficiency of the entire algorithm is reduced.
- Due to the classical “curse of dimensionality”, the predictive performance of surrogate models in high-dimensional parameter space is greatly reduced.
- The number of the initial design of experiments is typically unknown. This may affect the performance of the surrogate model unless an adaptive sampling method is adapted [81].
- The developed surrogate models are not interpretable (i.e., they are not physics-informed).
- While the proposed algorithm provides accuracy using several error metrics, it does not provide the confidence level associated with each decision.

Future studies can be focused on the following cases:

- Extend the proposed SASOI algorithm for correlated multi-target problems [84]. For example, SASOI can be used to forecast the seasonal dam deformation under the hydrostatic and thermal loading where the displacement is recorded and analyzed in several locations simultaneously.
- Extend the proposed SASOI algorithm for seismic analysis of dams in which the time history of the response parameter needs to be estimated [85].
- Extend the proposed SASOI algorithm for time-varying models [21] such as aging and deterioration of dams.
- While the current SASOI algorithm covers only the classical uncertainty quantification, it can be extended to model the heterogeneity in material properties in the context of the random fields theory [86].
- The surrogate models used in this paper can be replaced by multiple machine learning or deep learning methods to examine the accuracy of the meta-models.

Table A.10
Comparison of the measured value and predicted value for Dayakou Dam.

Working condition	Measured point	Uy [mm]	PCE_SOI	Kriging_SOI	PCK_SOI	SVR_SOI	BO
2017/6/15	1	7.44	7.47	8.23	7.63	7.94	7.55
	2	3.82	4.01	4.11	3.88	4.06	3.7
	3	1.68	1.64	1.4	1.69	1.46	1.15
2017/7/15	1	8.22	8.14	8.93	8.51	8.53	8.5
	2	4.36	4.09	4.27	4.07	4.23	4.04
	3	2.04	1.83	1.32	1.53	1.35	1.27
MAPE			4.17%	13.48%	6.65%	11.11%	14.11%
RSME			0.164	0.549	0.68	0.397	0.424
Bayesian inference time [s]			179	390	672	216	–
Total time [h]			0.883	0.942	1.02	0.893	27.664

Table A.11
Comparison of the measured value and predicted value for small-scaled arch dam.

Mode	Measured frequency [Hz]	PCE_SOI	Kriging_SOI	PCK_SOI	SVR_BOI
1	137	139.54	138.26	139.55	138.61
2	157	143.22	141.88	143.23	142.17
3	304	313.94	311.02	313.97	311.76
4	379	374.64	371.39	374.76	372.01
5	433	427.94	423.97	427.97	424.62
6	448	453.78	452.07	456.55	454.13
7	480	490.33	503.33	492.38	492.99
8	620	618.77	615.51	618.36	616.14
MAPE		2.48%	2.55%	2.62%	2.71%
RMSE		6.87	8.14	8.42	8.83

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Part of data related to monitoring dam data is confidential; however, the rest will be available upon reasonable request from the corresponding author.

Acknowledgments

This work was partially funded by the Fundamental Research Funds for the Central Universities, China (B200203123); the Postgraduate Research & Practice Innovation Program of Jiangsu Province, China (Grant No. KYCX20-0428); Nanjing Science and Technology Project, China (No. 202002014); and the Nantong Science and technology opening cooperation project, China in 2021 (No. BW2021001).

Appendix. Detailed results

See Tables A.9–A.11.

References

- [1] M. Hariri-Ardebili, Risk, Reliability, Resilience (R3) and beyond in dam engineering: A state-of-the-art review, *Int. J. Disaster Risk Reduct.* 31 (2018) 806–831.
- [2] A. De Sortis, P. Paoliani, Statistical analysis and structural identification in concrete dam monitoring, *Eng. Struct.* 29 (1) (2007) 110–120.
- [3] R. Arditto, G. Maier, G. Massalongo, Diagnostic analysis of concrete dams based on seasonal hydrostatic loading, *Eng. Struct.* 30 (11) (2008) 3176–3185.
- [4] S. Dou, J. Li, F. Kang, Parameter identification of concrete dams using swarm intelligence algorithm, *Eng. Comput.* 34 (7) (2017) 2358–2378.
- [5] Y. Yu, B. Zhang, H. Yuan, An intelligent displacement back-analysis method for earth-rockfill dams, *Comput. Geotech.* 34 (6) (2007) 423–434.
- [6] J. Deng, C. Lee, Displacement back analysis for a steep slope at the Three Gorges Project site, *Int. J. Rock Mech. Min. Sci.* 38 (2) (2001) 259–268.
- [7] S. Alves, J. Hall, System identification of a concrete arch dam and calibration of its finite element model, *Earthq. Eng. Struct. Dyn.* 35 (11) (2006) 1321–1337.
- [8] S.P. Neuman, G.E. Fogg, E.A. Jacobson, A statistical approach to the inverse problem of aquifer hydrology: 2. Case study, *Water Resour. Res.* 16 (1) (1980) 33–58.
- [9] J. Tsitsiklis, D. Bertsekas, M. Athans, Distributed asynchronous deterministic and stochastic gradient optimization algorithms, *IEEE Trans. Automat. Control* 31 (9) (1986) 803–812.
- [10] G. Sevieri, M. Andreini, A. De Falco, H.G. Matthies, Concrete gravity dams model parameters updating using static measurements, *Eng. Struct.* 196 (2019) 109231.
- [11] H.-R. Bae, R.V. Grandhi, R.A. Canfield, Sensitivity analysis of structural response uncertainty propagation using evidence theory, *Struct. Multidiscip. Optim.* 31 (4) (2006) 270–279.
- [12] M. Pelikan, D.E. Goldberg, E. Cantú-Paz, et al., BOA: The Bayesian optimization algorithm, in: *Proceedings of the Genetic and Evolutionary Computation Conference GECCO-99*, 1, Citeseer, 1999, pp. 525–532.
- [13] P.I. Frazier, Bayesian optimization, in: *Recent Advances in Optimization and Modeling of Contemporary Problems*, Informa, 2018, pp. 255–278.
- [14] S. Razavi, B.A. Tolson, D.H. Burn, Review of surrogate modeling in water resources, *Water Resour. Res.* 48 (7) (2012).
- [15] Y. Jin, Surrogate-assisted evolutionary computation: Recent advances and future challenges, *Swarm Evol. Comput.* 1 (2) (2011) 61–70.
- [16] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, *Reliab. Eng. Syst. Saf.* 93 (7) (2008) 964–979.
- [17] R. Ghanem, G. Saad, A. Doostan, Efficient solution of stochastic systems: application to the embankment dam problem, *Struct. Saf.* 29 (3) (2007) 238–251.
- [18] X. Guo, D. Dias, C. Carvajal, L. Peyras, P. Breul, Reliability analysis of embankment dam sliding stability using the sparse polynomial chaos expansion, *Eng. Struct.* 174 (2018) 295–307.
- [19] X. Guo, Q. Sun, D. Dias, E. Antoinet, Probabilistic assessment of an earth dam stability design using the adaptive polynomial chaos expansion, *Bull. Eng. Geol. Environ.* 79 (9) (2020) 4639–4655.
- [20] M. Hariri-Ardebili, B. Sudret, Polynomial chaos expansion for uncertainty quantification of dam engineering problems, *Eng. Struct.* (2020) 109631.
- [21] A. Amini, A. Abdollahi, M. Hariri-Ardebili, U. Lall, Copula-based reliability and sensitivity analysis of aging dams: Adaptive Kriging and polynomial chaos Kriging methods, *Appl. Soft Comput.* (2021) 107524.
- [22] X. Guo, D. Dias, Kriging based reliability and sensitivity analysis—Application to the stability of an earth dam, *Comput. Geotech.* 120 (2020) 103411.
- [23] A. Kalinina, M. Spada, D.F. Vetsch, S. Marelli, C. Whealton, P. Burgherr, B. Sudret, Metamodeling for uncertainty quantification of a flood wave model for concrete dam breaks, *Energies* 13 (14) (2020) 3685.
- [24] G. Shahzadi, A. Soulaïmani, Deep neural network and polynomial chaos expansion-based surrogate models for sensitivity and uncertainty propagation: An application to a rockfill dam, *Water* 13 (13) (2021) 1830.
- [25] M.A. Hariri-Ardebili, G. Mahdavi, A. Abdollahi, A. Amini, An RF-PCE hybrid surrogate model for sensitivity analysis of dams, *Water* 13 (3) (2021) 302.
- [26] G. Sevieri, A. De Falco, M. Andreini, H.G. Matthies, Hierarchical Bayesian framework for uncertainty reduction in the seismic fragility analysis of concrete gravity dams, *Eng. Struct.* 246 (2021) 113001.
- [27] T. Fengjie, T. Lahmer, Shape optimization based design of arch-type dams under uncertainties, *Eng. Optim.* 50 (9) (2018) 1470–1482.
- [28] Y.-Q. Wang, R.-H. Zhao, Y. Liu, Y.-Z. Chen, X.-Y. Ma, Shape optimization of single-curvature arch dam based on sequential Kriging-genetic algorithm, *Appl. Sci.* 9 (20) (2019) 4366.
- [29] Y. Wang, Y. Liu, X. Ma, Updated Kriging-assisted shape optimization of a gravity dam, *Water* 13 (1) (2021) 87.
- [30] A. Abdollahi, A. Amini, M. Hariri-Ardebili, An uncertainty-aware dynamic shape optimization framework: Gravity dam design, *Reliab. Eng. Syst. Saf.* (2022) 108402.

- [31] H.-z. Su, Z.-p. Wen, C.-s. Gu, An early-warning model of dam safety based on rough set theory and support vector machine, in: 5th International Conference on Machine Learning and Cybernetics, IEEE, Dalian, China, 2006, pp. 3455–3460.
- [32] H. Su, Z. Chen, Z. Wen, Performance improvement method of support vector machine-based model monitoring dam safety, *Struct. Control Health Monit.* 23 (2) (2016) 252–266.
- [33] V. Ranković, N. Grujović, D. Divac, N. Milivojević, Development of support vector regression identification model for prediction of dam structural behaviour, *Struct. Saf.* 48 (2014) 33–39.
- [34] M.M.R. Tabari, H.R.Z. Sanayei, Prediction of the intermediate block displacement of the dam crest using artificial neural network and support vector regression models, *Soft Comput.* 23 (19) (2019) 9629–9645.
- [35] B. Chen, Z. Wu, J. Liang, Y. Dou, Time-varying identification model for crack monitoring data from concrete dams based on support vector regression and the Bayesian framework, *Math. Probl. Eng.* 2017 (2017).
- [36] M. Hariri-Ardebili, F. Pourkamali-Anaraki, Support vector machine based reliability analysis of concrete dams, *Soil Dyn. Earthq. Eng.* 104 (2018) 276–295.
- [37] C. Lin, T. Li, S. Chen, X. Liu, C. Lin, S. Liang, Gaussian process regression-based forecasting model of dam deformation, *Neural Comput. Appl.* 31 (12) (2019) 8503–8518.
- [38] Y. Zhou, Y. Zhang, R. Pang, B. Xu, Seismic fragility analysis of high concrete faced rockfill dams based on plastic failure with support vector machine, *Soil Dyn. Earthq. Eng.* 144 (2021) 106587.
- [39] Q. Ren, M. Li, S. Bai, Y. Shen, A multiple-point monitoring model for concrete dam displacements based on correlated multiple-output support vector regression, *Struct. Health Monit.* (2022) 14759217211069639.
- [40] H. Su, J. Li, Z. Wu, Feedback analysis for mechanical parameters of dam and its foundation with optimization algorithm, *J. Hydraul. Eng.-ASCE* 10 (2007) 129–134.
- [41] I. Karimi, N. Khaji, M. Ahmadi, M. Mirzayee, System identification of concrete gravity dams using artificial neural networks based on a hybrid finite element-boundary element approach, *Eng. Struct.* 32 (11) (2010) 3583–3591.
- [42] Y. Chen, C. Gu, B. Wu, C. Shao, Z. Wu, B. Dai, Inversion modeling of dam-zoning elasticity modulus for heightened concrete dam using ICS-IPSO algorithm, *Math. Probl. Eng.* 2019 (2019).
- [43] L. Yang, H. Su, Z. Wen, Improved PLS and PSO methods-based back analysis for elastic modulus of dam, *Adv. Eng. Softw.* 131 (2019) 205–216.
- [44] R. Fedele, G. Maier, B. Miller, Identification of elastic stiffness and local stresses in concrete dams by in situ tests and neural networks, *Struct. Infrastructure Eng.* 1 (3) (2005) 165–180.
- [45] F. Kang, X. Liu, J. Li, H. Li, Multi-parameter inverse analysis of concrete dams using kernel extreme learning machines-based response surface model, *Eng. Struct.* 256 (2022) 113999.
- [46] C. Liu, C. Gu, B. Chen, Zoned elasticity modulus inversion analysis method of a high arch dam based on unconstrained Lagrange support vector regression (support vector regression arch dam), *Eng. Comput.* 33 (3) (2017) 443–456.
- [47] T. Bao, J. Li, Y. Lu, C. Gu, IDE-MLSSVR-based back analysis method for multiple mechanical parameters of concrete dams, *J. Struct. Eng.* 146 (8) (2020) 04020155.
- [48] V.H. Vu, M. Thomas, A. Lakis, L. Marcouiller, Operational modal analysis by updating autoregressive model, *Mech. Syst. Signal Process.* 25 (3) (2011) 1028–1044.
- [49] H. Li, G. Wang, B. Wei, Y. Zhong, L. Zhan, Dynamic inversion method for the material parameters of a high arch dam and its foundation, *Appl. Math. Model.* 71 (2019) 60–76.
- [50] F. Kang, Y. Wu, J. Li, H. Li, Dynamic parameter inverse analysis of concrete dams based on Jaya algorithm with Gaussian processes surrogate model, *Adv. Eng. Inform.* 49 (2021) 101348.
- [51] Y. Liu, L. Wenbo, G. Wang, L. Linmei, Y. Chen, Structural parameter inversion of a gravity dam based on the dynamic response induced by an underwater explosion, *J. Vib. Control* (2022) 10775463211066992.
- [52] N. Wiener, The homogeneous chaos, *Amer. J. Math.* 60 (4) (1938) 897–936.
- [53] K. Cheng, Z. Lu, C. Ling, S. Zhou, Surrogate-assisted global sensitivity analysis: an overview, *Struct. Multidiscip. Optim.* (2020) 1–27.
- [54] C. Soize, R. Ghanem, Physical systems with random uncertainties: chaos representations with arbitrary probability measure, *SIAM J. Sci. Comput.* 26 (2) (2004) 395–410.
- [55] N. Fajraoui, S. Marelli, B. Sudret, Sequential design of experiment for sparse polynomial chaos expansions, *SIAM/ASA J. Uncertain. Quant.* 5 (1) (2017) 1061–1085.
- [56] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, et al., Least angle regression, *Ann. Statist.* 32 (2) (2004) 407–499.
- [57] J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, Design and analysis of computer experiments, *Statist. Sci.* (1989) 409–423.
- [58] T.J. Santner, B.J. Williams, W.I. Notz, B.J. Williams, The design and analysis of computer experiments, vol. 1, Springer, 2003.
- [59] V. Dubourg, B. Sudret, F. Deheeger, Metamodel-based importance sampling for structural reliability analysis, *Probab. Eng. Mech.* 33 (2013) 47–57.
- [60] C. Lataniotis, D. Wicaksono, S. Marelli, B. Sudret, UQLab user manual – Kriging (Gaussian process modeling), Technical Report, Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, 2019, Report UQLab-V1.3-105.
- [61] C. Lataniotis, S. Marelli, B. Sudret, UQLab user manual–Kriging (Gaussian process modelling), Report, (3) Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Zurich, Switzerland, 2017.
- [62] N. Hansen, A. Ostermeier, Completely derandomized self-adaptation in evolution strategies, *Evol. Comput.* 9 (2) (2001) 159–195.
- [63] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, J. Wiart, A new surrogate modeling technique combining Kriging and polynomial chaos expansions–Application to uncertainty analysis in computational dosimetry, *J. Comput. Phys.* 286 (2015) 103–117.
- [64] R. Schobi, B. Sudret, J. Wiart, Polynomial-chaos-based Kriging, *Int. J. Uncertain. Quantif.* 5 (2) (2015).
- [65] V. Vapnik, The Nature of Statistical Learning Theory, Springer science & business media, 2013.
- [66] V. Cherkassky, Y. Ma, Practical selection of SVM parameters and noise estimation for SVM regression, *Neural Netw.* 17 (1) (2004) 113–126.
- [67] Y.M. Marzouk, H.N. Najm, L.A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, *J. Comput. Phys.* 224 (2) (2007) 560–586.
- [68] K.-V. Yuen, S.-C. Kuok, Bayesian methods for updating dynamic models, *Appl. Mech. Rev.* 64 (1) (2011).
- [69] A.M. Stuart, Inverse problems: a Bayesian perspective, *Acta Numer.* 19 (2010) 451–559.
- [70] J.B. Nagel, B. Sudret, A unified framework for multilevel uncertainty quantification in Bayesian inverse problems, *Probab. Eng. Mech.* 43 (2016) 68–84.
- [71] P. Wagner, J. Nagel, S. Marelli, B. Sudret, UQLab user manual–Bayesian inference for model calibration and inverse problems, Report No. UQLab-V1, 2019, pp. 3–113.
- [72] C.L. Segura Jr., S. Sattar, M.A. Hariri-Ardebili, Quantifying material uncertainty in seismic evaluations of reinforced concrete bridge column structures, *ACI Struct. J.* 119 (3) (2022) 141–152.
- [73] B. Echard, N. Gayton, M. Lemaire, N. Relun, A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models, *Reliab. Eng. Syst. Saf.* 111 (2013) 232–240.
- [74] M. Hariri-Ardebili, P. Boodagh, Taguchi design-based seismic reliability analysis of geostuctures, *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards* (2018) 1–19.
- [75] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, *Probab. Eng. Mech.* 25 (2) (2010) 183–197.
- [76] S. Marelli, B. Sudret, UQLab: A framework for uncertainty quantification in Matlab, in: *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, 2014, pp. 2554–2563.
- [77] MATLAB, version 9.11 (R2021b), The MathWorks Inc., Natick, Massachusetts, 2021.
- [78] ANSYS, ANSYS software reference manuals, release notes, mechanical APDL, elements reference, commands reference and theory reference, version release 18, 2017.
- [79] B. Shi, P. Qiao, A new surface fractal dimension for displacement mode shape-based damage identification of plate-type structures, *Mech. Syst. Signal Process.* 103 (2018) 139–161.
- [80] Y. Uchita, T. Shimpo, V. Saouma, Dynamic centrifuge tests of concrete dams, *Earthq. Eng. Struct. Dyn.* 34 (2005) 1467–1487.
- [81] C.G. Bucher, Adaptive sampling - an iterative fast Monte Carlo procedure, *Struct. Saf.* 5 (2) (1988) 119–126.
- [82] P. Diaz, A. Doostan, J. Hampton, Sparse polynomial chaos expansions via compressed sensing and D-optimal design, *Comput. Methods Appl. Mech. Engrg.* 336 (2018) 640–666.
- [83] D.K. Lin, T.W. Simpson, W. Chen, Sampling strategies for computer experiments: design and analysis, *Int. J. Reliab. Appl.* 2 (3) (2001) 209–240.
- [84] S. Chen, C. Gu, C. Lin, M.A. Hariri-Ardebili, Prediction of arch dam deformation via correlated multi-target stacking, *Appl. Math. Model.* 91 (2021) 1175–1193.
- [85] M.A. Hariri-Ardebili, S. Chen, G. Mahdavi, Machine learning-aided PSDM for dams with stochastic ground motions, *Adv. Eng. Inform.* 52 (2022) 101615.
- [86] F. Salazar, M.A. Hariri-Ardebili, Coupling machine learning and stochastic finite element to evaluate heterogeneous concrete infrastructure, *Eng. Struct.* 260 (2022) 114190.