

# Multi-parameter identification of concrete dam using polynomial chaos expansion and slime mould algorithm

Li YiFei<sup>a,b</sup>, Cao MaoSen<sup>a,\*</sup>, H. Tran-Ngoc<sup>c</sup>, Samir Khatir<sup>d</sup>, Magd Abdel Wahab<sup>b,\*</sup>

<sup>a</sup> Department of Engineering Mechanics, Hohai University, Nanjing, China

<sup>b</sup> Soete Laboratory, Department of Electrical Energy, Metals, Mechanical Constructions, and Systems, Faculty of Engineering and Architecture, Ghent University, Ghent, Belgium

<sup>c</sup> Department of Bridge and Tunnel Engineering, Faculty of Civil Engineering, University of Transport and Communications, Hanoi, Viet Nam

<sup>d</sup> Center for Engineering Application & Technology Solutions, Ho Chi Minh City Open University, Ho Chi Minh City, Viet Nam

## ARTICLE INFO

### Article history:

Received 15 August 2022

Accepted 6 March 2023

Available online 16 March 2023

### Keywords:

Concrete dams

Parameters identification

Polynomial chaos expansion

Slime mould algorithm

Parameter sensitivity analysis

## ABSTRACT

This paper presents a novel methodology that combines polynomial chaos expansion and slime mould algorithm for multi-parameter identification of concrete dams. This methodology not only incorporates the merits of low computational cost in the polynomial chaos expansion and fast convergence of slime mould algorithm, but also considers the priori uncertainty in the input parameters by introducing statistical probability theory. By considering two examples with different complexity, this paper verifies the effectiveness of the proposed method with a univariate simply supported beam model, followed by a complex multivariate dam model to demonstrate its practicability in real engineering problems. In addition, parameter sensitivity analysis of the dam model is conducted at an extremely low cost by polynomial chaos expansion based on Sobol' indices. Furthermore, the conventional parameter identification methods based on optimization methods directly combined with the finite element model are employed for comparison, highlighting two distinct advantages of the proposed method: (i) the proposed method improves the computational efficiency by nearly 52 times while ensuring a high accuracy, and (ii) the classical non-population optimization algorithm, Bayesian optimization, is used for comparison, revealing the outstanding performance of slime mould algorithm in terms of convergence speed and robustness. The application of the proposed algorithm is not only limited to dams, but also it can be extended to any structure.

© 2023 Elsevier Ltd. All rights reserved.

## 1. Introduction

Water and electricity are essential resources for human survival. The construction of large-scale water and hydroelectric infrastructures not only alleviates the uneven spatial distribution of natural resources to a large extent, but also contributes powerfully to the country's Gross Domestic Product (GDP). As a major component of the hydroelectric infrastructure, dams are responsible for power generation, flood control and irrigation. In the event of a dam failure, extensive losses may take place, so it is of great importance to continuously assess the safety of dams.

Typically, the foremost task of structural safety assessment is to establish a high-precision numerical model that can adequately characterize the physical properties of the structure. However, the accuracy of the numerical model depends on various factors,

among which material uncertainty is one of the key factors affecting the modelling accuracy. Therefore, identifying the unknown material parameters in the numerical model to reduce material uncertainty is a crucial step in the structural safety assessment. In addition, there are numerical methods, such as finite element method [1], scaled boundary element methods [2] and meshless methods [3], etc. For engineering problems, the most well-known choice is the finite element method.

The traditional parameters identification methods are primarily based on analytical or numerical displacement back analysis [4,5,6]. These methods were developed at a stage when simulation software was in its infancy, and therefore mainly aimed at the back-analysis of simple structures. They often require constantly changing the material parameters in the finite element model and then adopting a first-order optimal approach, such as the gradient descent method, to minimize the discrepancy between numerical simulations and field measurements [7]. However, parameter identification is typically a high-dimensional, multi-peak optimization problem, and it is extremely easy to fall into

\* Corresponding authors.

E-mail addresses: [20110018@hhu.edu.cn](mailto:20110018@hhu.edu.cn) (C. MaoSen), [magd.abdelwahab@ugent.be](mailto:magd.abdelwahab@ugent.be) (M. Abdel Wahab).

the trap of local minimum solutions using this traditional approach. In the past two decades, this problem has been gradually solved owing to the sharp development of metaheuristic optimization algorithms, such as particle swarm optimization algorithm [8], Jaya algorithm [9], etc.

Generally, these optimization algorithms are applied to parameter identification, which is essentially an iterative optimization process that requires many calls to the finite element model [10]. Such a rough operation is inefficient for complex large-scale structures, such as concrete dams. Finite element models are sufficient to characterize these structural properties, but they are normally computationally expensive. To overcome this shortcoming, many researchers have resorted to popular machine learning algorithms, among which the data-driven surrogate model comes to the fore quickly.

The surrogate model is an approximate mathematical model that attempts to offset the incremental cost of stochastic models by substituting an expensive computational model at a low-cost [11]. In recent years, surrogate models have gained a lot of momentum in the field of inverse analysis, spawning a novel coupling algorithm, i.e. Surrogate model-Assisted Metaheuristic Optimization Algorithm, denoted as SAMOA, for parameter identification. The core of the SAMOA is to build a surrogate model that can characterize the mapping relationship between input parameters with stochastic characteristics and the system's response (or Quantity of Interests - QoIs) [12]. Many kinds of surrogate models have been developed, and some of the more popular ones are Polynomial Chaos Expansion (PCE) [13], Radial basis functions (RBF) [14], Artificial neural networks [15], and so on.

### 1.1. Application of PCE to dams

PCE has been widely used in dam engineering in the last few years. Ghanem et al. [16] first introduced the PCE model for surrogate modelling of embankment dams. Guo et al. [17,18] performed the reliability analysis on an embankment dam using the sparse PCE, and evaluated its sliding stability. Hariri et al. [19] comprehensively quantified the uncertainty in dam engineering problems based on PCE via four different case studies with various complexities, encompassing various aspects of analytical and numerical modelling, static and seismic analysis, etc. Amini et al. [20] applied the adaptive PCE for sensitivity and reliability analysis of aging dams and investigated the copula dependency among the random variables (RVs). Kalinina et al. [21] conducted the study on the uncertainty in instantaneous dam-break floods by adopting PCE to approximate the flood model and explored the effect of uncertain input parameters on output variability based on Sobol' sensitivity indices. Shahzadi et al. [22] combined PCE with deep neural networks to construct a surrogate model for rockfill dams and to assess the effect of constitutive soil parameters on the behaviours of a rockfill dam. Hariri et al. [23] coupled random forests with PCE to perform sensitivity analysis for symmetrical and asymmetrical arch dams and to identify the most critical locations. Sevieri et al. [24,25] developed a generalized PCE-based probabilistic procedure in a Bayesian framework, incorporating parameter identification and seismic fragility analysis for concrete gravity dams.

### 1.2. SAMOA-based parameter identification in dams

Monitoring data (e.g., measured displacements and frequencies) are commonly used for parameter identification in dams [26]. Liu et al. [27] proposed a coupled Unconstrained Lagrangian Support Vector Machine (ULSVM) and Cultural Genetic Algorithm (CGA) for optimizing the zoned elasticity module of a high arch dams during the initial impound period. Kang et al. [28,29] combined separately Kernel Extreme Learning Machine (KELM)-based

Response Surface Model (RSM) and Kriging model with Jaya algorithm to identify the unknown static and dynamic parameters of dams. Bao et al. [30] identified multi-parameters of the Jinping-I arch dam based on Multi-output Least Squares Support Vector Regression (MLSSVR) combined with Improved Differential Evolutionary algorithm (IDE). Li et al. [31] combined RSM with Genetic Algorithm (GA) to identify the dynamic elastic modulus of the dam and foundation with a prototype arch dam during the flooding. Liu et al. [32] utilized operational modal analysis techniques to extract the first three major frequencies and basic mode shapes of the dam and developed support vector regression assisted particle swarm optimization algorithm to determine its dynamic elastic modulus.

The literature mentioned above demonstrates the abundant applications of PCE models in dam engineering, but it mainly focuses on parameter sensitivity analysis and uncertainty propagation of dams, with few studies applying them to parameter identification. However, PCE models can accurately develop alternative models of dams at a very low computational cost, which is ideal for bypassing the expensive finite element models of dams. Moreover, Slime Mould Algorithm (SMA), a newly developed stochastic optimization algorithm, has been shown to have good performance in terms of global optimization and convergence speed, but has not been too deeply involved in the application of large structures. In this paper, SMA and PCE (SMA-PCE) are combined to explore their practicability and high performance for multi-parameter identification in dams. The main contributions of the work are summarized as follows:

- A framework for rapid identification of dam parameters is developed, covering how to construct and evaluate PCE models, and how to couple PCE models with SMA based on objective functions for rapid parameter identification.
- The effectiveness of the proposed algorithm was verified with a simply supported beam, and then the practicality was confirmed with a multi-parameter hyperbolic arch dam. The results show that the coupled SMA-PCE algorithm significantly improves the computational efficiency without any reduction of accuracy compared to the classical method based on traditional optimization algorithms.
- While the sensitivity analysis of the dam parameters is carried out at an extremely low-cost based on the PCE model, the influence of important factors in the PCE model and SMA on the accuracy of the coupled SMA-PCE algorithm is explored. A relatively optimal combination of factors is given to further improve the accuracy of the algorithm.

The subsequent sections in this paper are arranged as follows. A brief review of the PCE and SMA is provided in Section 2. The detailed description on the procedures and evaluation metrics of the SMA-PCE algorithm is presented in Section 3. The effectiveness of the SMA-PCE algorithm is validated by simply supported beam model in section 4, and then the multi-parameter identification of concrete dam using the SMA-PCE algorithm is introduced in section 5. Finally, the conclusions are given in Section 6.

## 2. Theoretical background

### 2.1. Polynomial chaos expansion

The classic PCE was first introduced by Wiener [33]. The key concept of PCE is to express the computational model on the basis of an orthogonal polynomial with input random variables [34] and is well-suited to solve the global smoothing problem, which is common in many engineering applications.

Considering a  $M$  dimensional random vector with independent components  $\mathbf{x} = \{x_1, x_2, \dots, x_M\}$  described by the joint PDF  $f_{\mathbf{x}}, j = 1, 2, M$ , the scalar  $Y$  resulted from this system is also a random variable, denoted  $Y = \mathcal{M}(\mathbf{x})$ .  $Y$  can be expressed exactly in an exact infinite expansion using the following equation [35,36]:

$$Y = \mathcal{M}(\mathbf{x}) \approx \mathcal{M}^{PCE}(\mathbf{x}) = \sum_{\boldsymbol{\kappa} \in \mathcal{A}} \xi_{\boldsymbol{\kappa}} \varphi_{\boldsymbol{\kappa}}(\mathbf{x}) \quad (1)$$

where  $\boldsymbol{\kappa} = \{\kappa_1, \dots, \kappa_n\} (\kappa_j \geq 0) \in \mathcal{A}$  is the multidimensional index notation vector and  $\mathcal{A} \in \mathbb{N}^M$  is the truncation set of multi-indices.  $\xi_{\boldsymbol{\kappa}} \in \mathbb{R}$  are the expansion coefficients to be determined and  $\varphi_{\boldsymbol{\kappa}}(\mathbf{x}) = \prod_{j=1}^M \psi_{\kappa_j}^{(j)}(x_j)$  are multivariate polynomials orthogonal with respect to  $f_{\mathbf{x}}$ , among them,  $\psi_{\kappa_j}^{(j)}$  is the univariate orthogonal polynomial in the  $j^{\text{th}}$  variable of corresponding polynomial degree  $\kappa_j$ .

The sparsity-of-effects principle states that most models describing physical phenomena are governed by main effects and lower order interactions in the real world [37]. Hence, the accuracy of Eq. (1) depends on  $\mathcal{A}$  as well as  $\xi_{\boldsymbol{\kappa}}$ , which can be enhanced by truncation. There are two main truncation schemes, i.e., standard and hyperbolic [38,39], which are described in this section. The former one corresponds to all polynomials in the  $M$  input variables of total degree less than or equal to  $p$ :

$$\mathcal{A}^{M,p} = \{\boldsymbol{\kappa} \in \mathbb{N}^M : |\boldsymbol{\kappa}| \leq p\} \text{card.} \mathcal{A}^{M,p} \equiv P = \binom{M+p}{p} = \frac{(M+p)!}{p!M!} \quad (2)$$

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

$$\mathcal{A}^{M,p,q} = \{\boldsymbol{\kappa} \in \mathcal{A}^{M,p} : \|\boldsymbol{\kappa}\|_q \leq p\}, \|\boldsymbol{\kappa}\|_q = \left(\sum_{i=1}^M \kappa_i^q\right)^{1/q} \quad (3)$$

where  $q = 1$  for the standard truncation scheme in Eq. (2) and  $q < 1$  for hyperbolic truncation. Reducing the  $q$ -value decreases the number of polynomials of high interaction order included in the expansion, thus significantly reducing the number of model evaluations, but at the meantime, some errors may be caused by missing higher order terms [40]. Fig. 1 presents the variation of  $p$  and  $q$  for Hyperbolic truncation.

The infinite series in Eq. (1) can be rewritten as follows:

$$Y = \mathcal{M}(\mathbf{x}) = \sum_{j=0}^{P-1} \xi_j \varphi_j(\mathbf{x}) + \varepsilon_p \equiv \boldsymbol{\xi}^T \boldsymbol{\psi}(\mathbf{x}) + \varepsilon_p \quad (4)$$

where  $P$  is defined based on Eq. (2),  $\varepsilon_p$  is the truncation error, and superscript  $T$  means transpose.  $\boldsymbol{\xi} = \{\xi_0, \dots, \xi_{P-1}\}^T$  is a vector containing the coefficients and  $\boldsymbol{\psi}(\mathbf{x}) = \{\varphi_0(\mathbf{x}), \dots, \varphi_{P-1}(\mathbf{x})\}^T$  is the vector that assembles the values of all the orthogonal polynomials in  $\mathbf{x}$ .

There are multiple techniques to calculate the expansion coefficients  $\xi_{\boldsymbol{\kappa}}$  in Eq. (1), which might be divided into intrusive or non-

intrusive approaches. The former one is originally proposed in the context of the probabilistic finite element method [41], which is mainly used to discretize the constitutive equation both in physical space and random space. The coefficient calculation based on the latter is a post-processing result obtained from a set of model evaluations (also called the design of experiment DOE [42]), that are given on proper sampling of the random input variables. An advanced non-intrusive method, Least Angle Regression (LAR) [43], is adapted as a strategy to calculate the expansion coefficients.

The least angle regression (LAR) algorithm uses low-rank truncation schemes, which aims to find coefficient vectors with only a few non-zero entries (i.e., sparse solutions), while the other coefficients are set to zero. It can be formulated by expanding the least square minimization and adding a penalty term  $\lambda \|\boldsymbol{\xi}\|_1$  as:

$$\hat{\boldsymbol{\xi}} = \underset{\boldsymbol{\xi} \in \mathbb{R}^P}{\text{argmin}} \mathbb{E}[(\boldsymbol{\xi}^T \boldsymbol{\varphi}(\mathbf{x}) - Y)^2] + \lambda \|\boldsymbol{\xi}\|_1 \quad (5)$$

where  $\|\boldsymbol{\xi}\|_1 = \sum_{\boldsymbol{\kappa} \in \mathcal{A}} |\xi_{\boldsymbol{\kappa}}|$  is the regularization term that forces the minimization to favor low rank solutions.

### 2.1.1. PCE-based Sobol' indices for sensitivity analysis

Global sensitivity analysis can quantify the impact of input variables on the importance of model outputs, which aims to reduce the dimensions of the problem to minimize the number of model evaluations as much as possible [44]. Sobol' indices are common for sensitivity analysis based on variance decomposition, motivated by the idea of the extension of the computational model as a sum of increasing dimensions [45].

Suppose that all  $M$  dimensional input variables are  $\mathbf{x} = (x_1, x_2, \dots, x_M) \sim u[0, 1]^M$ , therefore, the Sobol' decomposition is defined as in Eq. (6):

$$\begin{aligned} \mathcal{M}(\mathbf{x}) &= \mathcal{M}_0 + \sum_{i=1}^M \mathcal{M}_i(x_i) + \sum_{1 \leq i < j \leq M} \mathcal{M}_{ij}(x_i, x_j) + \dots + \mathcal{M}_{12\dots M}(\mathbf{x}) \\ &= \mathcal{M}_0 + \sum_{\mathbf{v} \subset \{1, \dots, M\}} \mathcal{M}_{\mathbf{v}} \end{aligned} \quad (6)$$

where  $\mathbf{v} = \{i_1, \dots, i_s\}$  is a non-empty generic index set.

Hence, the total variance of the model  $M$  is described in terms of the sum of the variances as follows:

$$D \equiv \text{Var}[\mathcal{M}(\mathbf{x})] = \sum_{\mathbf{v} \subset \{1, \dots, M\}} \text{Var}[\mathcal{M}_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}})] \quad (7)$$

Eq. (7) naturally leads to the natural definition of Sobol' indices for sensitivity analysis:

$$S_{\mathbf{v}} = \frac{\text{Var}[\mathcal{M}_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}})]}{D} \quad (8)$$

Which represents the relative contribution of each group of variables  $\{x_{i_1}, \dots, x_{i_s}\}$  to the total variance, among them. The index concerning one input variable  $x_i$  is called the first order Sobol'

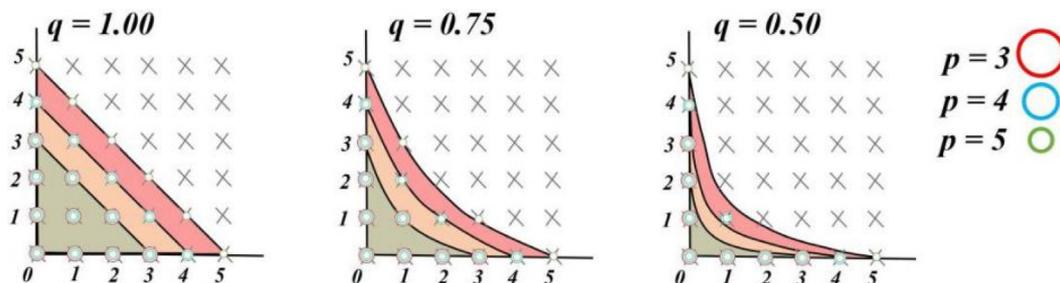


Fig. 1. Hyperbolic truncation set for  $p$  and  $q$ , adopted from [19]

index, which is widely used due to its capacity to quantify the additive effect of each input parameter individually, which can be defined as:

$$S_i = \frac{D_i}{D} = \frac{\text{Var}[\mathcal{M}_i(x_i)]}{D} \quad (9)$$

In addition, the sum of all Sobol' indices involving this variable  $x_i$  is termed as the total Sobol' index, denoted  $S_i^T$ :

$$S_i^T = \sum_{\mathbf{v} \supset i} S_{\mathbf{v}} \quad (10)$$

However, classical Monte Carlo-based Sobol' indices require extensive model evaluation, which makes them difficult to use with computationally expensive models. In order to bypass this problem, Sudret [13] has developed a PCE-based Sobol' indices for sensitivity analysis with low computational cost, which can be obtained by reordering the terms of the truncated PCE as in Eq. (2).

To conveniently reflect the decomposition into sums of increasing order, Eq. (6) can be reordered as:

$$\begin{aligned} \mathcal{M}^{\text{PCE}}(\mathbf{x}) &= \mathcal{M}_0 + \sum_{\mathbf{v} \subset \{1, \dots, M\}} \mathcal{M}_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}}), \text{ where } \mathcal{M}_{\mathbf{v}}(\mathbf{x}_{\mathbf{v}}) \\ &= \sum_{\kappa \in \mathcal{A}_{\mathbf{v}}} \zeta_{\kappa} \varphi_{\kappa}(\mathbf{x}) \end{aligned} \quad (11)$$

where  $\mathcal{A}_{\mathbf{v}} = \{\kappa \in \mathcal{A} : l \in \mathbf{v} \iff \kappa_l \neq 0\}$  is the set containing all multi-indices with non-zero components in the truncation set  $\mathcal{A}$ .

Due to the orthogonality of the polynomial chaos basis, the Sobol' indices are obtained analytically, at any order from the coefficients of the truncated PCE, hence, Eq. (8) can be rewritten as:

$$S_{\mathbf{v}} = D_{\mathbf{v}}/D = \sum_{\kappa \in \mathcal{A}_{\mathbf{v}}} \zeta_{\kappa}^2 / \sum_{\kappa \in \mathcal{A}} \zeta_{\kappa}^2 \quad (12)$$

## 2.2. Slime mould algorithm

In this section, the mathematical model of the slime mould algorithm is described in detail [46].

The Slime mould can approach food in response to airborne odours. In order to show its behaviours in a mathematical description, the following equation is introduced to reflect the reduction pattern.

$$\overrightarrow{X}(t+1) = \begin{cases} \overrightarrow{X}_b(t) + \overrightarrow{vb} \cdot (W \cdot \overrightarrow{X}_A(t) - \overrightarrow{X}_B(t)), r < p \\ \overrightarrow{vc} \cdot \overrightarrow{X}(t), r \geq p \end{cases} \quad (13)$$

where  $\overrightarrow{X}_b$  represents the individual location with the highest odor concentration currently found,  $t$  represents the current iteration, parameter  $\overrightarrow{vb} = [-a, a]$ ,  $a = \text{arctanh}(-(t/\text{max}_t) + 1)$ , and  $\overrightarrow{vc}$  linearly diminishes from 1 to 0.  $\overrightarrow{X}_A$  and  $\overrightarrow{X}_B$  represent two individuals randomly selected vectors from the swarm.  $\overrightarrow{X}$  represents the location of slime mould,  $r$  denotes the random value in the interval of  $[0, 1]$ ,  $p = \tanh|S(i) - DF|$ , among them,  $S_i, i \in 1, 2, \dots, n$  represents the fitness of  $\overrightarrow{X}$ .  $DF$  refers the best fitness obtained in all iterations and  $\overrightarrow{W}$  represents the weight of slime mould.

$\overrightarrow{W}$  can be determined by the following formula:

$$\overrightarrow{W}(\text{SmallIndex}) = \begin{cases} 1 + r \cdot \log\left(\frac{bF - S(i)}{bF - wF} + 1\right), \text{ condition} \\ 1 - r \cdot \log\left(\frac{bF - S(i)}{bF - wF} + 1\right), \text{ others} \end{cases} \quad (14)$$

where  $\text{SmallIndex} = \text{sort}(S)$  indicates that the fitness values are sorted ascending in the minimum value problem and  $bF$  and  $wF$ , respectively, denote the best and worst fitness value gained from

the current iteration. The term 'condition' indicates that  $S(i)$  ranks first half of the population.

Fig. 2(a) visualizes the effects of Eq.13 and illustrates the position change of the searching individual in 2D and 3D space, from which we can observe that the searching individuals can forage without any angular or directional constraints. This allows them to approach the optimal solution in all possible directions. It is also applicable to extend this concept to hyper-dimensional space.

The mathematical formula for updating the location of slime mould is written as follows:

$$\overrightarrow{X}^* = \begin{cases} \text{rand} \cdot (UB - LB) + LB, \text{ rand} < z \\ X_b(t) + \overrightarrow{vb} \cdot (W \cdot \overrightarrow{X}_A(t) - \overrightarrow{X}_B(t)), r < p \\ \overrightarrow{vc} \cdot \overrightarrow{X}(t), r \geq p \end{cases} \quad (15)$$

where  $LB$  and  $UB$  denote the lower and upper boundaries of search space, and their values are usually set artificially depending on the specific problem. In this paper, in order to explore the boundaries of search space,  $LB$  and  $UB$  are considered in conjunction with the prior probability space of the input parameters, denoted as  $LB = \mu - RF \times \sigma$  and  $UB = \mu + RF \times \sigma$ . Among these parameters,  $\mu$  and  $\sigma$  are, respectively, the mean and standard deviation of the probability distribution for the input parameters.  $RF$  is the range factor of the search boundary. Hence, scaling of the search space can be achieved by varying the value of  $RF$ .

Both  $\overrightarrow{vb}$  and  $\overrightarrow{vc}$  tend to zero gradually as the position of slime mould is updated continuously, meanwhile, the searching individuals will dock near the optimal location. A concise flowchart for describing the three different stages of SMA can be found in Fig. 2(b).

## 3. Proposed parameter identification Algorithm: SMA-PCE

The procedure of the proposed SMA-PCE algorithm for parameter identification is described in detail below.

**Phase One** - Define the problem (Step A): The first step is to build a finite element model to characterize the response of the real structure and to define a probabilistic input prior model (distribution type, mean and deviation) for the unknown parameters to be identified. The parameters of the real structural system are often inherently stochastic, and it is more realistic on account of uncertainty in the input parameters.

**Phase Two** - Constructing and evaluating PCE models (Step B-C): A certain number of input parameter datasets are generated using Latin Hypercube Sampling (LHS) and the corresponding output response datasets are then extracted based on probabilistic finite element models. Finally, the PCE surrogate model is constructed based on the 'input-output' sampling spectrum and its accuracy is evaluated by leave-one-out (LOO) cross-validation error, with resampling to generate more sampling points if accuracy is insufficient.

**Phase Three** - Coupling the PCE model with SMA and minimizing the objective function (Step D-E): Once an accurate and reliable PCE model has been constructed, it can be used as a direct substitute for the finite element model to calculate the output of the structural response, which directly bypasses a series of problems caused by the expensive computational cost of the Finite Element Analysis (FEA). By incorporating the predicted and measured output into the objective function, the coupling of the PCE model to SMA is implemented.

**Phase Four** - Parameter identification posterior verification (Step F): Post-process the optimal value of the parameter identification and treat it as a realistic value of the unknown input parameters for the model updating process.

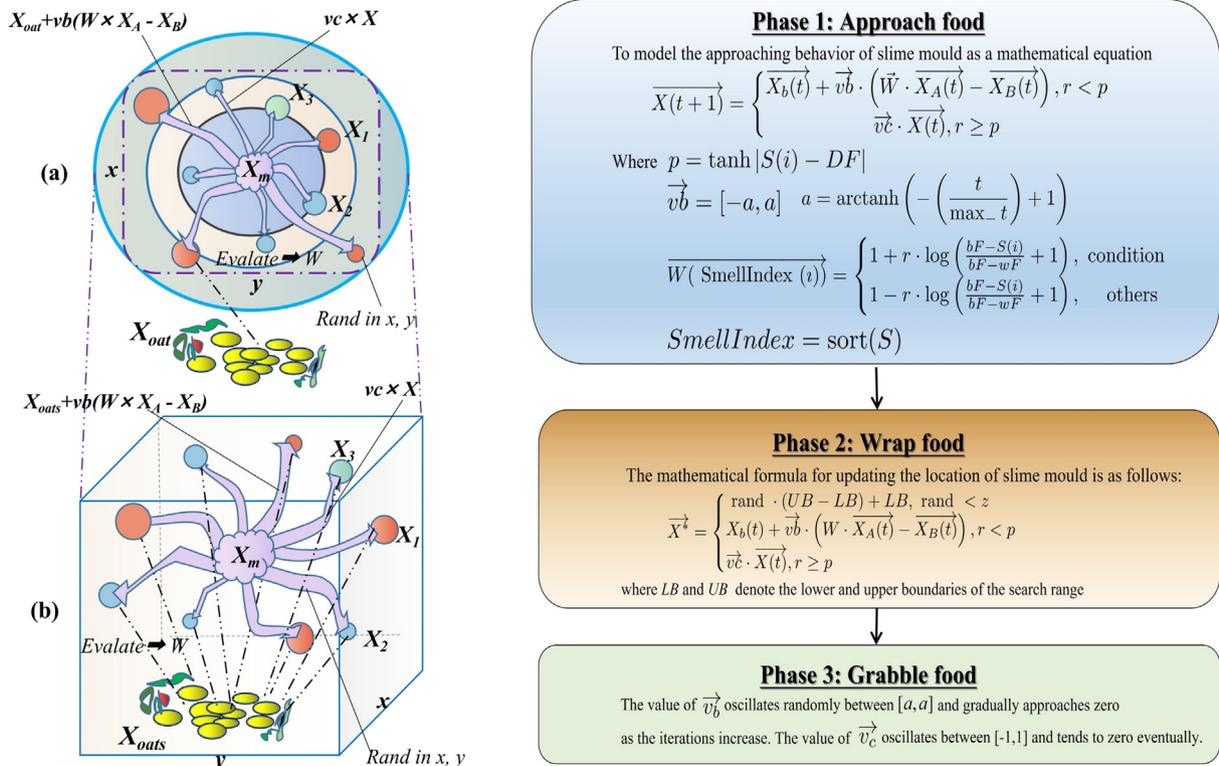


Fig. 2. Graphical and flowchart presentation of SMA: (a) Possible locations in 2D and 3D (adopted from [47]) and (b) flowchart describing the three phases of SMA.

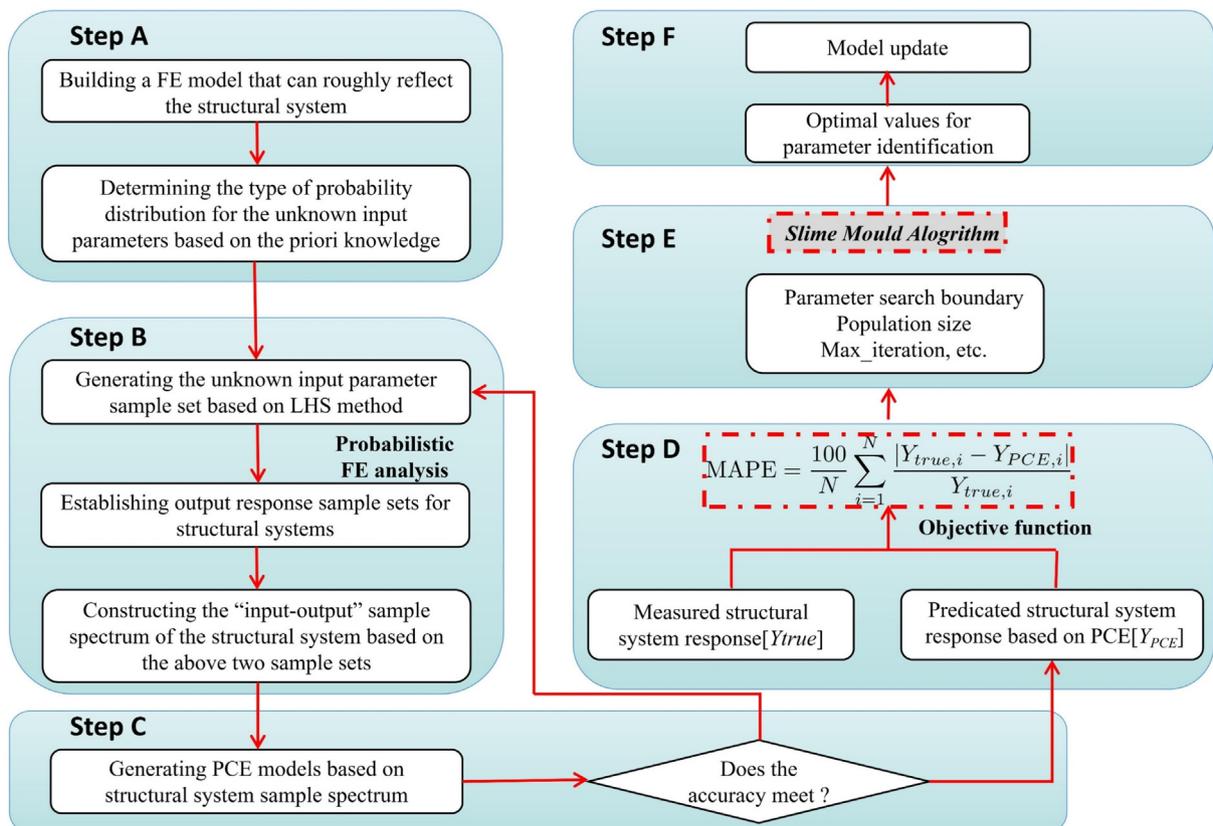


Fig. 3. Flowchart to describe the SMA-PCE method.

A detailed description of the flowchart of the SMA-PCE algorithm is shown in Fig. 3.

### 3.1. Metrics to evaluate SMA-PCE

The LOO cross-validation error,  $Err_{LOO}$ , is a model accuracy evaluation method developed based on statistical learning theory, which not only can largely reduce the risk of over-fitting the data set by cross-validation, but also, does not require running additional expensive model evaluations to generate an appropriate validation set [13].

After the PCE model is constructed (see Section 2.1), its accuracy and predictive quality can be quantified. This error method includes the construction of  $N$  PCE models  $\mathcal{M}^{PCE_i}$ , building on a left experimental design  $X \mathbf{x}^{(j)} = \{\mathbf{x}^{(k)}, k = 1, \dots, N, k \neq j\}$ , and then comparing the true value  $y^{(j)}$  with the predicted value on the excluded point [48]. It is defined as:

$$Err_{LOO} = \frac{\sum_{j=1}^N (\mathcal{M}(\mathbf{x}^{(j)}) - \mathcal{M}^{PCE_i}(\mathbf{x}^{(j)}))^2}{\sum_{j=1}^N (\mathcal{M}(\mathbf{x}^{(j)}) - \hat{\mu}_Y)^2} \quad (16)$$

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

The mean absolute percentage error (MAPE) is a common metric in statistics to assess the accuracy of the prediction methods. In terms of root mean square error, etc., it has a more intuitive interpretation in the form of a percentage. Hence, MAPE is selected as the objective function to evaluate the accuracy of the SMA-PCE method in this paper. Its formula is shown below:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_{true,i} - y_{PCE,i}|}{y_{true,i}} \quad (17)$$

where  $y_{true,i}$  is the actual measured value and  $y_{PCE,i}$  is the forecast value based on PCE model,  $n$  is the number of fitted points.

## 4. Validation of SMA-PCE by a simply supported beam

This section explores the validity of the SMA-PCE method for parameter identification by studying a simply supported beam.

### 4.1. Construction of parameter prior distribution space

This example evaluates a simply supported beam with a well-defined analytical expression for mid-span deflection, that is  $V_{mid} = \frac{5pl^4}{32Ebh^3}$ , having a uniformly distributed load  $p = 12,000$  N/m, length  $L = 5.0$  m, width  $b = 0.15$  m and height  $h = 0.30$  m. A set of  $N = 5$  independent experiments were carried out on this beam, and the measurements were [49]: 12.84, 13.12, 12.13, 12.19, and 12.67 mm. Due to measurement error, the measured deflections vary across experiments. Assuming that the modulus of elasticity,  $E$ , is an unknown input parameter and is subjected to a lognormal distribution,  $LN(30, 4.5)$  GPa. On this basis, the geometric model,

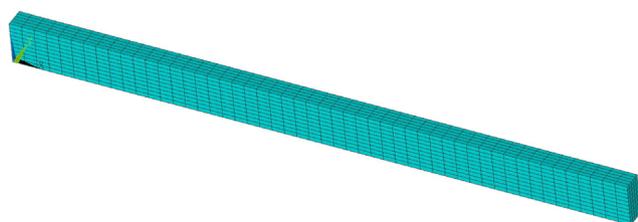
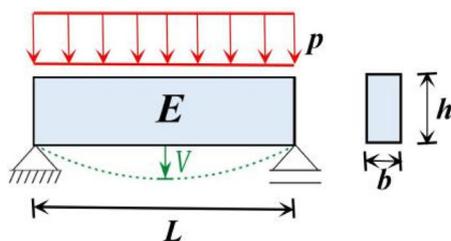


Fig. 4. Geometric and numerical models to describe simply supported Beam: (a) Geometric model and (b) finite element model.

and the finite element model of the simply supported beam are established as shown in Fig. 4. The finite element model is built directly based on ANSYS APDL using hexahedral SOLID185 elements, with the size of each element being  $0.1 \text{ m} \times 0.03 \text{ m} \times 0.03 \text{ m}$ , and a total of 2500 elements ( $N_{elements} = 2500$ ). In this paper, the finite element analysis assumes small deformation, and uses linear elastic constitutive relation for the quasi-static analysis of the structure.

### 4.2. Results and discussions

This section presents the results of constructing the PCE model and the accuracy of parameter identification based on the SMA-PCE algorithm for the simply supported beam. The unknown input parameter  $E$  is sampled by 100 times using the LHS method, and then, these “one-to-one” calculated mid-span deflections (outputs) are extracted based on the probabilistic FEA. The initial “input and output” dataset consists of two parts: the dataset for the design of the experiment,  $N_{DOE}$ , and the dataset for validation,  $N_{val}$ . The former is used to construct the PCE model, and the latter is used to verify its prediction accuracy by LOO cross-validation error.

The numerical computing environment in this paper is based on a high-performance UNIX workstation with 36-core CPU and 192 GB of memory, and the probabilistic FEA is executed in ANSYS APDL by calling 12-core CPU. The open-source software UQLab [50] is used to construct the PCE model, and then the algorithmic coupling of PCE and SMA is implemented in MATLAB.

To explore the effect of different truncation schemes and polynomial degrees on the accuracy of the PCE model, considering truncation norm  $q = [1.0, 0.75, 0.5]$  and polynomial degree  $p = [1, 2, 3, 4, 5, 6]$  in this paper, a sum of 18 PCE models based on  $N_{DOE} = 20$  and  $N_{val} = 20$  is constructed as shown in Fig. 5(a) with the following major observations:

- In the selected  $p$ -set, the accuracy of the PCE model is proportional to  $p$  and the PCE model has high reliability when  $p \geq 2$ .
- Different truncation schemes have almost no effect on the accuracy of the PCE model in a simple univariate structure. According to Fig. 1, when the evaluation point falls in the low-dimensional space (close to the origin of the coordinates), it always participates in the model evaluation even if the  $q$  value changes.

In addition, in order to further reveal the factors affecting the accuracy of the PCE model, we also consider different element sizes  $N_{elements} = [500, 1000, 2500, 5000]$  and  $N_{DOE} = [5, 10, 20, 40, 80]$ . Based on the above analysis,  $q = 0.75$  is chosen as the truncation scheme for this exploration, subsequently, a series of PCE models is constructed with the same validation dataset as shown in Fig. 5(b) and (c). Some interesting observations are as follows:

- From Fig. 5(b), we can notice two observations. (i) For different element sizes (different  $N_{elements}$ ), the accuracy of the PCE model increases with the number of polynomial degree  $p$  within the

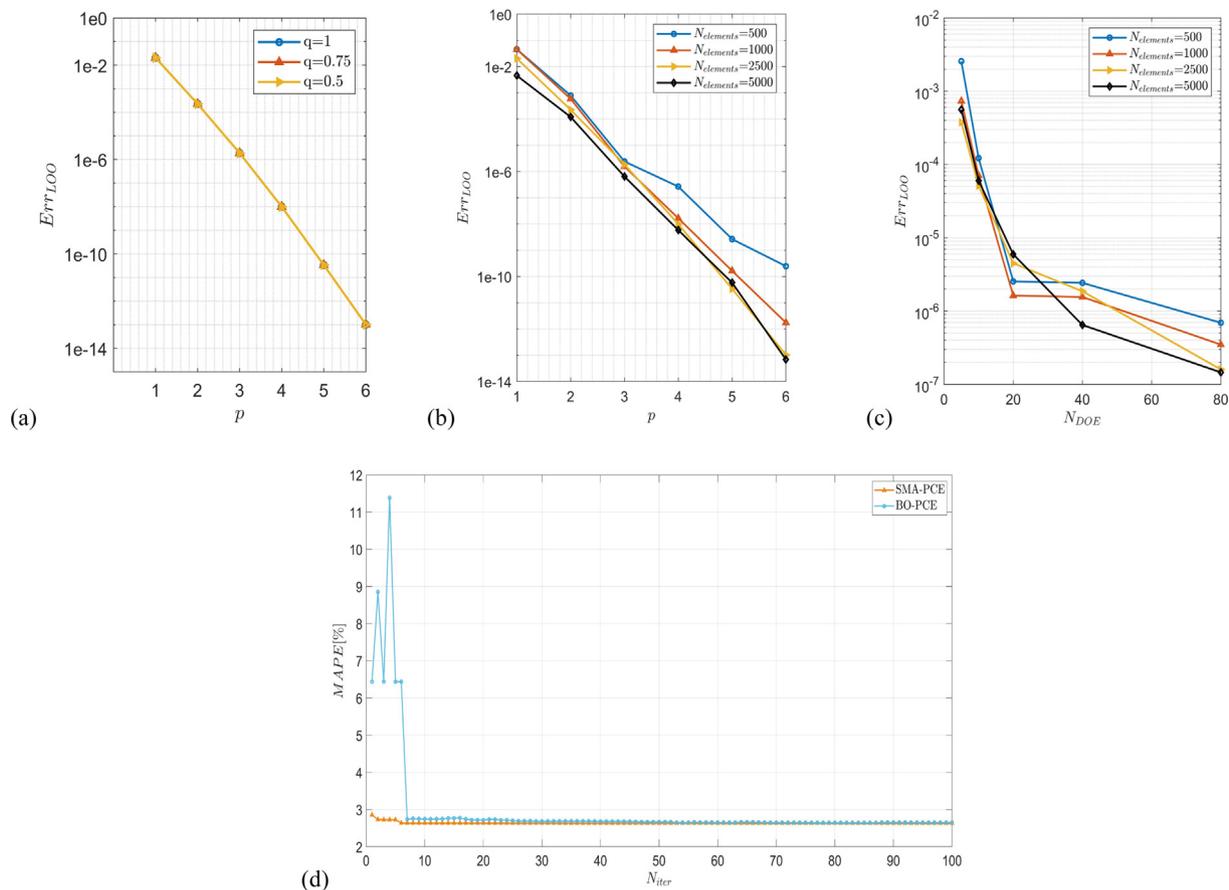


Fig. 5. Construction of PCE model and parameter identification for simply supported beam: (a)~(c) Accuracy of PCE model and (d) comparison between the accuracy of SMA-PCE and BO-PCE.

selected  $p$ -set, which is the same as in Fig. 5(a). (ii) the accuracy of the PCE models corresponding to the finite element models with different mesh densities shows some differences, which is due to the fact that the accuracy of the output response of the structural finite element model is closely related to its mesh density, and generally shows a positive correlation.

- From Fig. 5(c), we can observe that with the increase of  $N_{DOE}$ , the accuracy of the PCE model increases corresponding to the four finite element models with different  $N_{elements}$ . Especially before  $N_{DOE} = 20$ , the  $Err_{LOO}$  has a sharp decrease. After that, although it decreases, the rate has slowed down, and it can be predicted that the final convergence will occur.

As a rule of thumb, the surrogate model can be a perfect alternative for the original computational model when its accuracy reaches  $10^{-3}$ . While higher accuracy is desirable, it may not be efficient from a computational cost point of view. Therefore, keeping the surrogate model balanced in terms of accuracy and cost is essential. So, the PCE model with the moderate collection [ $p = 3$ ,  $q = 0.75$ ,  $N_{DOE} = 20$ ,  $N_{val} = 20$ ] is selected as the pilot model and coupled with SMA. Meanwhile, the well-known non-population stochastic optimization algorithm, Bayesian optimization (BO), is coupled with PCE for comparison with this algorithm. The following notations specify respectively each combination: SMA-PCE and BO-PCE. Their accuracy and convergence characteristics are reflected in Fig. 5(d), from which we can observe that SMA-PCE has faster convergence and stronger robustness than BO-PCE. However, both algorithms converge to almost the same value of computational accuracy after 30 iterations, i.e., MAPE is less than 3%,

which meets the acceptance criteria for this simple structure. This directly demonstrates the effectiveness of SMA-PCE for parameter identification.

In addition, a similar exploration into the accuracy of the PCE model has been applied to explore the accuracy of SMA-PCE. From Fig. 6, we can draw the following conclusions:

- A vertical comparison, for the same  $q$  with different  $p$ , where the convergence rate of SMA-PCE in the pre-convergence oscillation phase, does not correlate with the accuracy of the PCE model.
- A horizontal comparison, for different  $q$  with the same  $p$ , even though the accuracy of the PCE model is the same, shows that the accuracy of SMA-PCE still has a slight difference in the pre-convergence oscillation stage. This is caused by the stochastic search characteristic of SMA.
- Overall, SMA-PCE basically converges consistently when the number of iterations,  $N_{iter}$ , exceeds 10, which further demonstrates that SMA-PCE is well-suited for parameter identification with a very fast convergence rate and strong robustness.

## 5. Application of SMA-PCE to dam structure

The previous section verifies the excellent effectiveness of the SMA-PCE method based on the univariate simply supported beam. In this section, the same algorithm is applied to the multivariate complex structure, a concrete hyperbolic arch dam, to explore the applicability of this algorithm for large civil engineering structures.

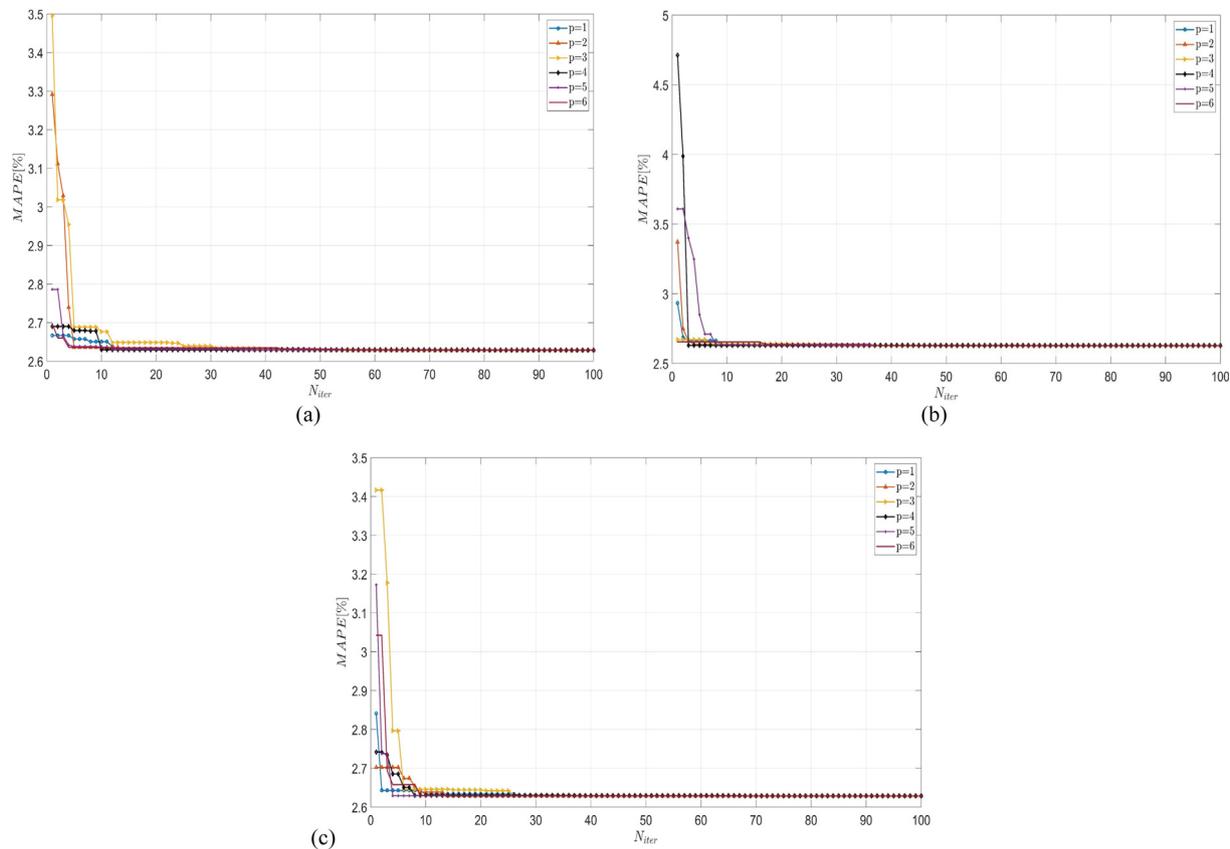


Fig. 6. Quantifying the accuracy of the SMA-PCE method in simply supported Beam for different values of  $p$  and  $q$ : (a)  $q = 1.0$ , (b)  $q = 0.75$  and (c)  $q = 0.5$ .

### 5.1. Experimental measurements on dam

Dayakou dam is a concrete double-curved arch dam, for which the top and bottom elevations are 653 m and 558 m, respectively, and the maximum dam height is 95 m. The dam is designed with four transverse-joints in the form of a through-seam and the maximum spacing is about 70 m. In addition, three overflow surface holes are placed in the middle of the dam crest, with the weir crest elevation of 643.5 m and the net width of each hole is 10.0 m. The layout of these constructions can be clearly seen in Fig. 7(a). The dam was completed in November 2015 and began to store water in February 2017 until the end of July 2017.

For more reliable characterization of the actual structural response, a high-precision “dam-foundation” three-dimensional finite element model was constructed based on hexahedral and tetrahedral SOLID185 elements in ALTAIR HyperMesh, as shown in Fig. 7(b). The whole model was divided into six components, where the dam model consists of the dam body and the weak layer, and the foundation model made up of bottom and top bedrock as well as geological faults, in between, the base surface for buffering is provided to enhance the slip and seepage resistance of the dam. In addition, the foundation model is extended to 1.5 and 2.5 times the dam height in the upstream and downstream directions, respectively, and it is also extended to 2 times the dam height to the left, right, and bottom. The classical mass-less foundation method was used to simulate the infinite boundary, and the quasi-static analysis executed based on small deformation assumptions and linear elastic constitutive relation. Their basic properties are reported in Table 1.

The measured points on the A1 measurement line in the middle of the dam to monitor the arch dam deformation are shown in Fig. 7(a). The measured points 1 and 2 are arranged on the positive

vertical line, and the measured point 3 is arranged on the inverted vertical line. Since the base point of the inverted vertical is positioned deep in the rock foundation, the default absolute displacement is 0. By calculating the measured data of each measuring point, the absolute displacement at the three measured points can be obtained, i.e., the measured displacement along the river,  $U_y$ , for subsequent parameter identification.

### 5.2. Load analysis based on measured data

Since there are significant distinctions in the load situation of the dam at different periods, the impounding period after the completion of the dam was selected as the target of this paper. Typically, the load situation on the dam during this period is dominated by hydrostatic pressure and temperature loads, where hydrostatic pressure depends on the reservoir water level at the upstream and downstream of the dam. Temperature loads are mainly caused by the difference between the temperature field of the dam surface and the temperature field of the seal arch grouting. Fig. 8 illustrates the distribution of the latter at upstream and downstream of the dam.

Based on the relevant measured data, the upstream reservoir water level of the dam as well as the average monthly air temperature and surface water temperature over time are summarized in Fig. 9. From Fig. 9, we can observe that the dam started to store water on February 15, 2017, and then, the upstream reservoir water level was increased to the normal water level of 648 m on July 15, 2017. For the sake of simplifying the complexity of the problem, the two-time points (2017/6/15 and 2017/7/15) with the closest reservoir water levels are selected as references. The corresponding upstream water levels are 647.3 m and 648 m, and the downstream water level is always 571 m, upon which

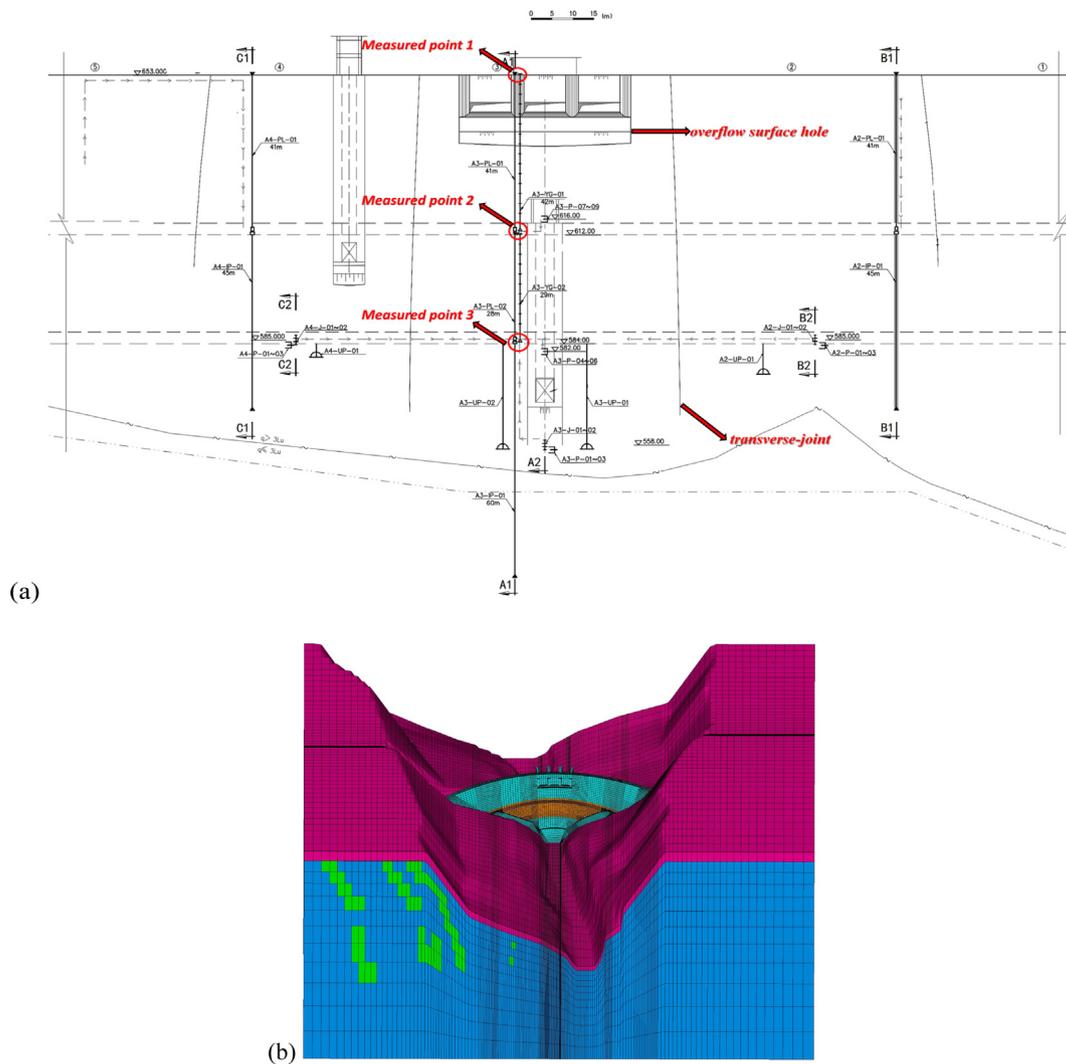


Fig. 7. Description of Dayakou arch dam: (a) Dam downstream plane and the layout of measured points and (b) numerical model.

Table 1  
Basic parameters of the finite element model for Dayakou arch dam.

Component Order and Name	No. of Elements	Elastic modulus [GPa]	Density [kg/m <sup>3</sup> ]	Poisson's ratio [-]	Coefficient of linear thermal expansion [1/°C]
1-Dam body	47,788	20	2400	0.19	$7.0 \times 10^{-6}$
2-Base surface	2304	20	2400	0.19	$7.0 \times 10^{-6}$
3-Weak layer of dam	4040	16	2400	0.21	$7.0 \times 10^{-6}$
4-Bottom bedrock	129,569	12	-	0.22	-
5-Top bedrock	22,160	6	-	0.26	-
6-Geologic fault	4983	5	-	0.3	-

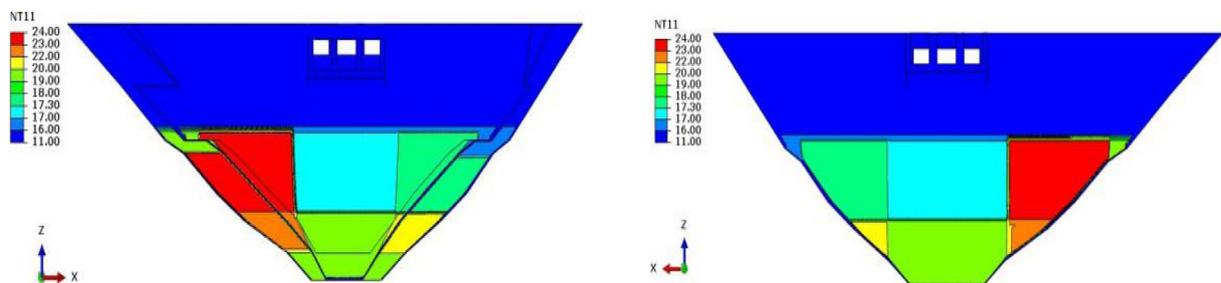


Fig. 8. Sealing arch grouting temperature field of Dayakou arch dam: (a) Downstream and (b) upstream.

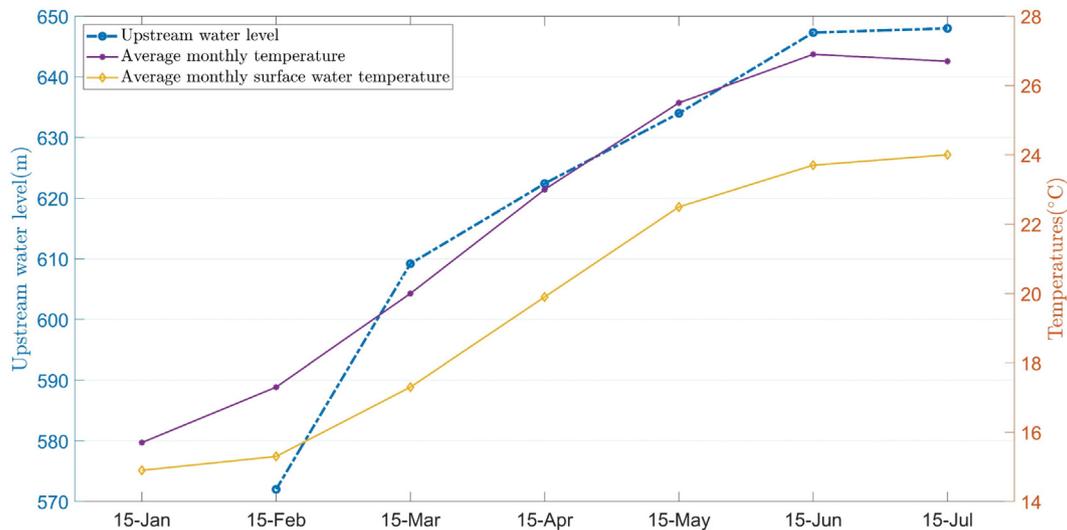


Fig. 9. Measured reservoir water level and temperature of Dayakou arch dam.

the hydrostatic pressure of each node can be calculated. In addition, the measured displacements at these two time points can be obtained based on the measured data, which are shown in Table 2.

However, the calculation for temperature loads is more complex, in which the surface temperature field of the dam is divided into two parts. The first part is the boundary temperature field of the dam in contact with the air, which adopts an optimization method to determine its value, after considering that the concrete surface temperature is impacted by light. The optimization strategy is to generate 5 sets of sample data in the temperature interval with or without sunshine, as in shown Table 2, where W1 is sunshine considered, W2 is sunshine not considered, and W3-7 is random sampling in both intervals based on the LHS method. According to the above work conditions, the FEA was carried out separately, and then the Root Mean Square Error (RMSE) was selected to estimate the difference between the measured and calculated displacements, and the temperature value with the smallest RMSE was chosen as the optimal boundary temperature. The second part is the boundary temperature field of the dam in contact with water, which can be obtained by linear interpolation based on a small amount of measured data from the underwater surface temperature of the dam with consideration of light.

Considering that the sealing arch grouting temperature of the dam is below 20 °C in most regions, which is lower than the calculated surface temperature of the dam in this case, the applied loads are dam self-weight, hydrostatic pressure (from upstream and downstream reservoirs), and temperature loads.

Table 2 Optimization for the temperature field at the boundary between the dam and the air.

Time	Measured point	Measured displacement (mm)		Boundary temperature of the dam in contact with the air (°C)						
		Relative value	Absolute value	W1	W2	W3	W4	W5	W6	W7
2017/6/15	1	3.62	7.44	4.510	9.930	8.772	5.829	7.277	7.724	7.077
	2	2.14	3.82	2.411	4.370	3.949	2.887	3.409	3.571	3.337
	3	1.68	1.68	0.517	1.210	1.058	0.684	0.868	0.925	0.843
	RMSE			1.994	1.496	0.852	1.219	0.534	0.487	0.596
2017/7/15	4	3.86	8.22	4.510	9.930	8.772	5.829	7.277	7.724	7.077
	5	2.32	4.36	2.411	4.370	3.949	2.887	3.409	3.571	3.337
	6	2.04	2.04	0.517	1.210	1.058	0.684	0.868	0.925	0.843
	RMSE			2.574	1.102	0.692	1.801	1.027	0.839	1.123

### 5.3. Construction of prior distribution parameter space

In this example, the modulus of elasticity and the Coefficient of Linear Thermal Expansion (CLTE) are considered as unknown input parameters and obey Gaussian distribution model, with the average of a few field measurements as the mean and given a 10% standard deviation. The probabilistic input priori model of the nine unknown parameters is shown in Table 3, and then, the calculated displacement along the river at these measured points is extracted as outputs (QoIs). The numerical calculation environment is the same as in Section 4.

### 5.4. Parameter sensitivity analysis based on PCE

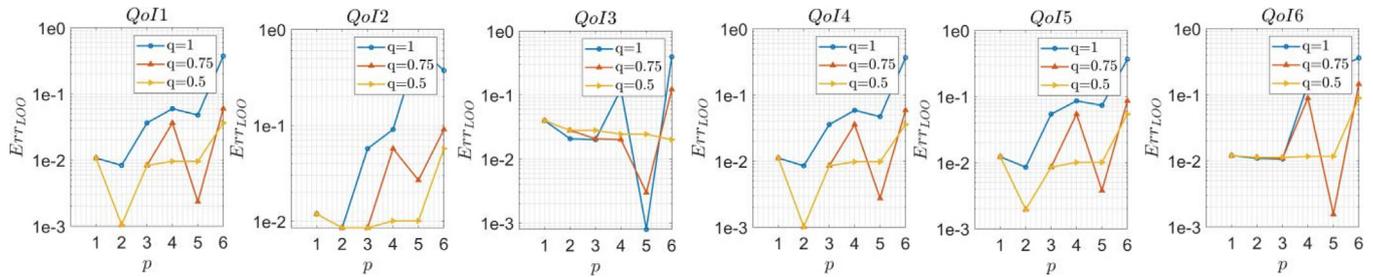
This section also provides a comprehensive exploration on the impact of different truncation schemes and polynomial degrees on the accuracy of the PCE model. Then, a PCE model with reliable accuracy is selected for sensitivity analysis on the input parameters as shown in Table 3, to reveal their effects on structural output response.

In this case, 40 sets “input-output” datasets are extracted for constructing the PCE model in the same way as in Section 4.2. Fig. 10 shows the accuracy of all PCE models for the Dayakou arch dam, from which we can come to some conclusions that differ from univariate modelling:

- Generally, Hyperbolic truncation ( $q < 1$ ) improves the accuracy of the PCE model.

**Table 3**  
Unknown parameter probabilistic input priori model for Dayakou arch dam.

Component Number	Unknown Parameters	Symbol	Model	Quantity
1	Elastic Modulus	$E1$	Gaussian	$N(20, 2)$
	CLTE	$\alpha1$	Gaussian	$N(7.0 \times 10^{-6}, 7.0 \times 10^{-7})$
2	Elastic Modulus	$E2$	Gaussian	$N(20, 2)$
	CLTE	$\alpha2$	Gaussian	$N(7.0 \times 10^{-6}, 7.0 \times 10^{-7})$
3	Elastic Modulus	$E3$	Gaussian	$N(20, 2)$
	CLTE	$\alpha3$	Gaussian	$N(7.0 \times 10^{-6}, 7.0 \times 10^{-7})$
4	Elastic Modulus	$E4$	Gaussian	$N(12, 1.2)$
5	Elastic Modulus	$E5$	Gaussian	$N(6, 0.6)$
6	Elastic Modulus	$E6$	Gaussian	$N(5, 0.5)$



**Fig. 10.** Quantification of  $Err_{LOO}$  in the Dayakou arch dam based on the PCE model. Note: QoIs 1–3 and QoIs 4–6 separately correspond to three measured points on 2017/6/15 and 2017/7/15.

- In multi-parameter, the polynomial degree,  $p$ , is not necessarily positively correlated with the accuracy of PCE model.
- In a trade-off comparison, the PCE model has a high accuracy for all three different truncation schemes when  $p = 2$ .

Consequently, the six PCE models corresponding to  $q = 0.75$  and  $p = 2$  are selected to characterize the Dayakou arch dam model. The accuracy of each PCE model is close to or below  $1E-2$ , which satisfies the requirement of prediction accuracy for large civil engineering structures. Accordingly, PCE-based Sobol' indices sensitivity analysis concerning the input parameters of the Dayakou arch dam was carried out, and the results are illustrated in Fig. 11.

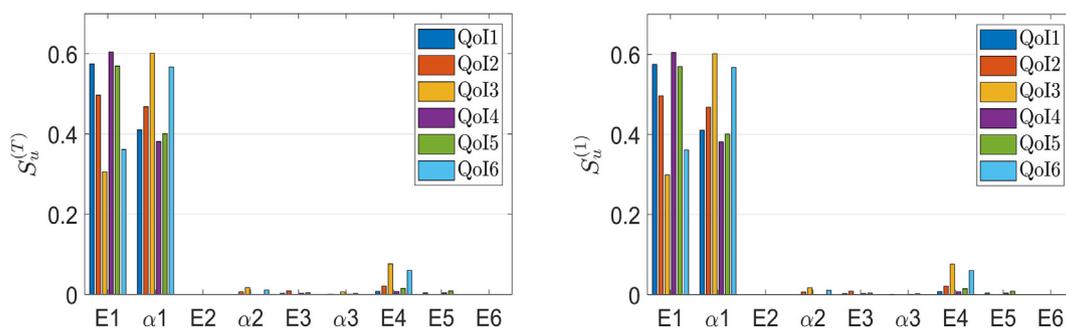
- It is obvious that  $E_1$  and  $\alpha_1$  have the paramount importance on the structural output response, since Component 1 in Table 1 is the absolute-dominant part of the dam.
- Total Sobol' indices and First order Sobol' indices have a slight difference in results, which is caused by that higher-order interactions between input parameters having extremely less impact on the overall results of parameter sensitivity analysis.

5.5. Comparison and remarks

These crucial factors affecting the accuracy of the PCE model are discussed in detail in Section 4.2. However, two important factors affecting the accuracy of SMA, that is, the number of search agents  $N$  and the search boundary range factor  $RF$ , which have not been explored. Hence, this section will further explore the impact of these important on the accuracy of the SMA-PCE method for Dayakou dam.

This issue is explored using the control variables method. First, the important factors in SMA are fixed to explore the effect of varying  $p$  and  $q$  on the accuracy of the SMA-PCE method, and their results are shown in Fig. 12, from which we can draw the following conclusions:

- For multivariate parameter identification, different PCE model accuracy has a more significant impact on the accuracy of the SMA-PCE algorithm, which is distinctly different from the previous conclusions in the univariate Simply Supported Beam model.
- The accuracy of the PCE model is not proportional to the accuracy of the SMA-PCE method. For example, the PCE model has the highest accuracy when  $p = 2$  in a compromise, but currently the accuracy of the SMA-PCE algorithm is not so satisfactory.



**Fig. 11.** PCE-based Sobol' indices for the Dayakou arch dam: (a) Total Sobol' indices and (b) first order Sobol' indices.

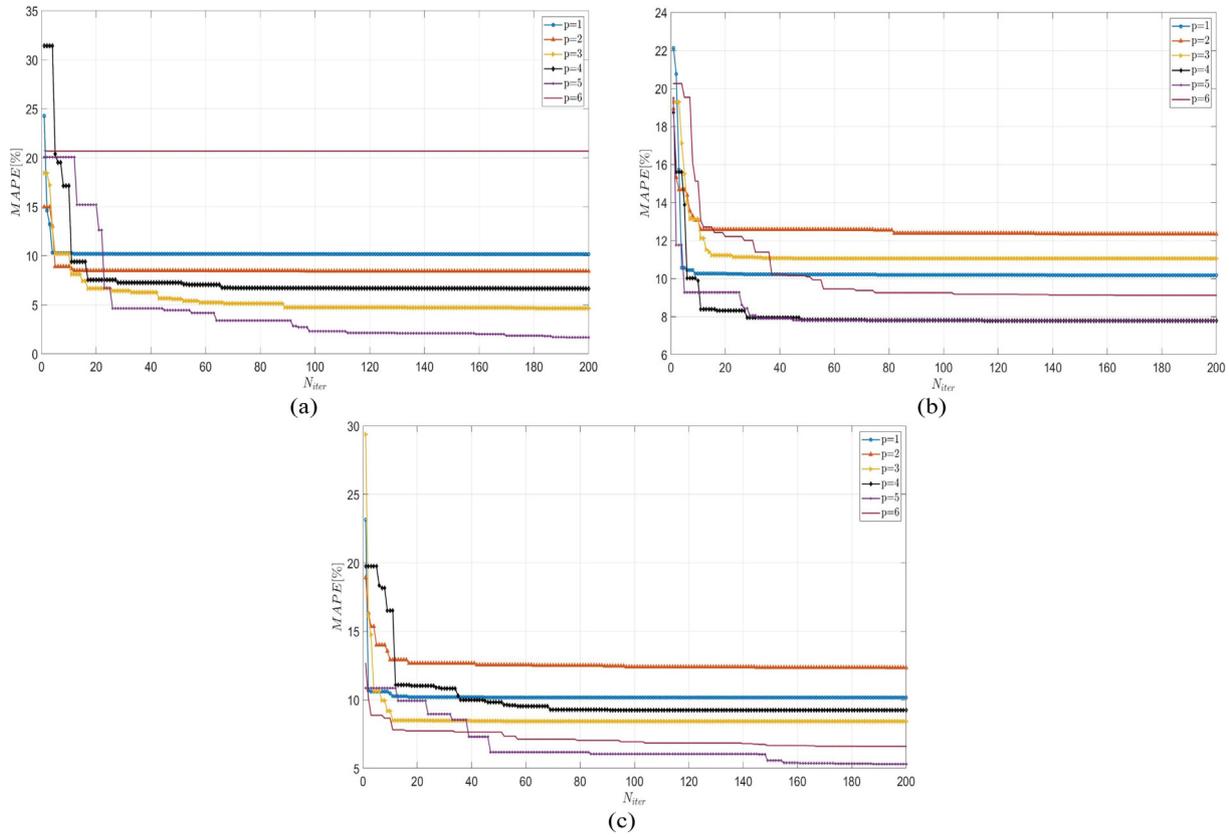


Fig. 12. The accuracy of the SMA-PCE method for different values of  $p$  and  $q$ ,  $N = 10$  and  $RF = 3$ : (a)  $q = 1.0$ , (b)  $q = 0.75$  and (c)  $q = 0.5$ .

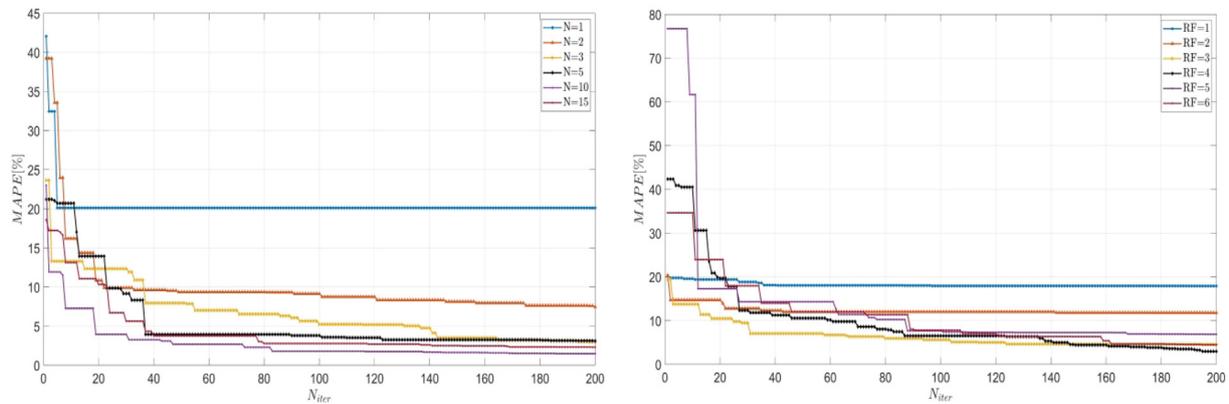


Fig. 13. The accuracy of the SMA-PCE method;  $q = 0.75$  and  $p = 5$ . (a) varying values of  $N$ ,  $RF = 3$ ; (b) varying values of  $RF$ ,  $N = 5$ .

- No matter how much  $q$ , the SMA-PCE algorithm always has good accuracy when  $p = 5$ .

Comprehensively, considering the accuracy of the PCE model and SMA-PCE method,  $p = 5$  and  $q = 0.75$  may be deemed to be a preferable choice. On this basis, the effects of varying  $N$  and  $RF$  on the accuracy of SMA-PCE are explored separately. The results are shown in Fig. 13.

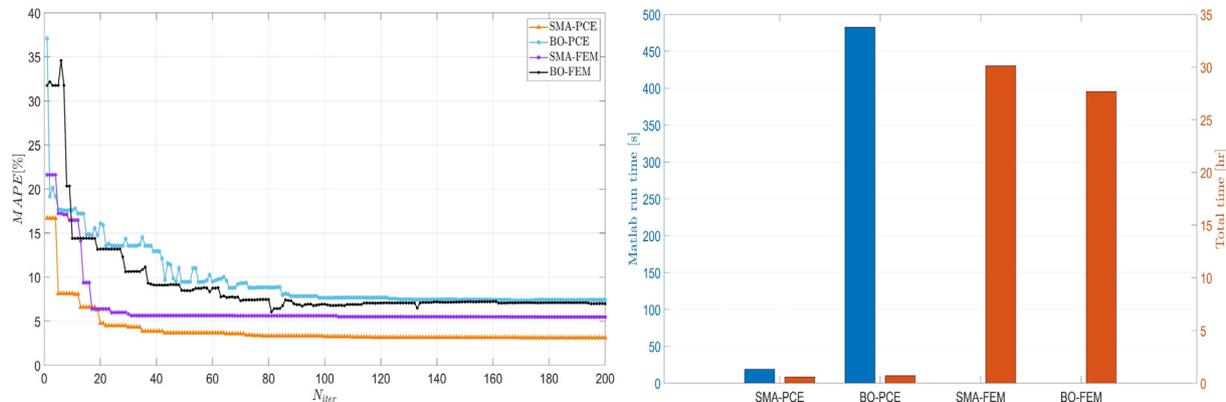
- Overall, varying the values of both  $N$  and  $RF$  has a moderate effect on the accuracy of the SMA-PCE method.
- In theory, the larger the value of  $N$  is, the greater the possibility of convergence to the local optimal solution. But in the view of

Fig. 13(a), convergence occurs when  $N \geq 5$ , and if a too large value of  $N$  is chosen, it will instead greatly increase the consumption of computational resources.

- Varying  $RF$  value has a greater impact on the SMA-PCE accuracy than varying  $N$  value. This can be well explained by the fact that when  $RF$  is fixed, increasing the value of  $N$  will improve the efficiency of SMA search for local optimum. Fig. 13(a) shows that when  $N \geq 5$ , the algorithm tends to converge after 40 iterations. However, when  $N$  is fixed, but increasing the value of  $RF$ , i.e., the search boundary is extended, the speed of SMA in finding the local optimum may decrease. Fig. 13(b) shows that the accuracy of the SMA-PCE method not only shows a larger oscillation when  $RF \geq 4$ , but also decreases compared to  $RF = 3$ .

**Table 4**  
Comparison between the measured and predicted parameters for Dayakou Dam.

Working condition	Measured point	Uy [mm]	SMA-PCE	BO-PCE	SMA-FEM	BO-FEM
2017/6/15	1	7.44	7.438	7.498	7.476	7.487
	2	3.82	3.825	4.881	3.911	3.942
	3	1.68	1.634	1.380	1.535	1.390
2017/7/15	1	8.22	8.528	8.564	8.583	8.504
	2	4.36	4.180	4.488	4.261	4.477
	3	2.04	1.850	1.594	1.768	1.611
MAPE			3.343%	8.202%	5.253%	8.043%
Part 1[h]			0.578	0.578	-	-
Part 2[s]			11.92	488.32	-	-
Total time [h]			0.581	0.715	30.06	27.66



**Fig. 14.** Comparison of accuracy and efficiency between SMA-PCE and SMA-FEM inversion algorithms for Dayakou Dam: (a) Inversion accuracy evaluation and (b) inversion efficiency evaluation.

**Table 5**  
Parameter identification results and calculation time of each algorithm for Dayakou Dam.

Methods	SMA-PCE	BO-PCE	SMA-FEM	BO-FEM
E1	25.8	20.2	26	25.3
$\alpha 1$	$7.69 \times 10^{-6}$	$8.96 \times 10^{-6}$	$5.96 \times 10^{-6}$	$5.99 \times 10^{-6}$
E2	14	15.7	14	16.9
$\alpha 2$	$7.27 \times 10^{-6}$	$8.73 \times 10^{-6}$	$5.57 \times 10^{-6}$	$6.71 \times 10^{-6}$
E3	11.2	11.6	20.8	19.2
$\alpha 3$	$6.40 \times 10^{-6}$	$7.16 \times 10^{-6}$	$6.00 \times 10^{-6}$	$5.58 \times 10^{-6}$
E4	11	12.8	8.4	8.43
E5	7.28	7.12	4.2	4.3
E6	6.42	6.26	3.51	4.63

Considering the above discussions, a set of important factors was chosen, i.e.  $[q = 0.75, p = 5, N = 10, RF = 3]$ , to construct the PCE model and the SMA-PCE method. Finally, the SMA-PCE method is compared with the classical method based directly on iterative optimization. The following notations: SMA-FEM and BO-FEM, specifically refer to the direct combination of SMA or BO with finite element model, which is performed employing co-simulation in MATLAB and ANSYS APDL, so that their total time can be simply recorded directly in MATLAB.

It is worth mentioning that the total computation time of the coupled algorithms in this paper is divided into two parts. (1) Part 1 is the time for building the “input–output” dataset of the PCE model, so it is the same both for SMA-PCE and BO-PCE (2) Part 2 is the running time of the coupled algorithms in MATLAB.

The comparative results of the computational accuracy and efficiency of the four methods are summarized in Table 4 and Fig. 14. In addition, their parameter identification results are given in Table 5, from which we can draw some conclusions as follows:

- Compared with BO-PCE and BO, SMA-PCE and SMA offer significant improvements, not only in terms of faster convergence

and better robustness, but also in terms of higher accuracy, which is reflected in Fig. 14(a);

- In terms of computational cost, Table 4 and Fig. 14(a) show that the computational time of the SMA-PCE method is nearly 52 times shorter than that of the SMA-FEM method.
- A further comparison between SMA-PCE and BO-PCE shows a significant difference in Part 2, i.e., the former is almost 40 times shorter than the latter. Although the difference in their total computation time is not significant, it is sufficient to further demonstrate the superiority of the SMA-PCE method.

## 6. Conclusions

Due to the uniqueness of their function and their large volumes, dams are often built in high mountain and valley river sections. These areas usually have complex geological and climatic conditions, and with the variable operating conditions of the dams themselves. Therefore, timely assessment of dam safety in a changing environment is a pressing challenge, and rapid parameter identification provides a feasible alternative to this dilemma.

However, most classical parameter identification methods for dams are based on an optimization algorithm directly combined with the finite element model for iterative back-analysis, which requires multiple invocations of the finite element model. Hence, the computational efficiency of such methods is severely restricted by the computational cost of the finite element model. In this paper, the efficient and accurate PCE models are developed as a direct alternative to the computationally expensive finite element models to predict their structural response in real-time, and then coupled with the novel and powerful slime mould algorithm. The proposed SMA-PCE method can be used as an effective technique to rapidly identify the structural unknown parameters, as demonstrated in the above two examples with different complexities. The general conclusions are summarized as follows:

- In terms of constructing PCE models, the PCE models corresponding to the univariate simply supported beam model are much more accurate than the multivariate Dayakou arch dam model. The same size “input–output” datasets (40 sets) are used to construct the PCE model, the  $Err_{LOO}$  of the former is close to  $1 \times 10^{-6}$ , while the  $Err_{LOO}$  of the latter is at most  $1 \times 10^{-3}$ . This is associated with the classical “curse of dimensionality” in machine learning algorithms. Therefore, for the multi-parameter identification problem, the predictive accuracy of the surrogate model can be improved by appropriately increasing the sample size.
- In multivariate models, the accuracy of the PCE model is closely related to the truncation norm  $q$  and polynomial degree  $p$ . In general, choosing the hyperbolic truncation scheme,  $q < 1$ , will improve the accuracy of the PCE model, while increasing  $p$  is not positively correlated with the accuracy of the PCE model. So, in practical applications, the relatively optimal combination of  $q$  and  $p$  should be chosen to construct a reliable PCE model for predicting the structural response, which is a prerequisite for the SMA-PCE method to identify the parameters.
- In terms of SMA, the number of search agents  $N$  and the search boundary range factor  $RF$  have a significant impact on the convergence speed as well as the accuracy of SMA-PCE. This feature can be generalized to other population intelligence optimization algorithms. In addition, SMA has faster convergence speed and stronger robustness compared to the classical non-population Bayesian optimization algorithm.
- In the example of the Dayakou arch dam model, compared with traditional parameter identification methods, such as the SMA-FEM method, the SMA-PCE method not only reduces the computational time by more than 50 times, but also improves the computational accuracy to a certain extent by choosing the suitable combination of important factors [ $p$ ,  $q$ ,  $N$ ,  $RF$ ].

In summary, the SMA-PCE method presented in this paper compensates for the low computational efficiency of traditional parameter identification methods without compromising computational accuracy, and its application can be extended to other structures. There are a few limitations and drawbacks to this algorithm:

- The initial “input–output” dataset for constructing the surrogate model is sourced from the initial computational model (finite element model, etc.), since the efficiency of the entire algorithm is greatly limited by the computational cost of the initial computational model.
- Due to the classical “curse of dimensionality”, the predictive performance of surrogate models in high-dimensional parameter spaces will be greatly reduced, so that an appropriate increase in modelling cost is essential in such problems.

- The appropriate size of the initial design-of-experiment datasets ( $N_{DOE}$ ) used to construct the surrogate model are normally unknown, and thus it is necessary to use different  $N_{DOE}$  to construct the surrogate model for comparison.
- The developed surrogate models are not interpretable, i.e. they are not physics-informed.

Future studies can be focused on the following cases:

- Extend the proposed SMA-PCE algorithm in the identification for structural dynamic parameters considering vibrations [51].
- Extend the proposed SMA-PCE algorithm for seismic analysis of dams, in which the time history of the response parameter needs to be estimated [52].
- Extend the proposed SASOI algorithm for time-varying models such as aging and deterioration of dams [53].

### Data availability

The data that has been used is confidential.

### 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.

### Acknowledgment

The authors are grateful for the Jiangsu International Joint Research and Development Program (No. BZ2022010); the Nanjing International Joint Research and Development Program (No. 202112003); the Fundamental Research Funds for the Central Universities (B220204002).

### References

- [1] Nanthakumar S, Lahmer T, Zhuang X, et al. Detection of material interfaces using a regularized level set method in piezoelectric structures. *Inverse Prob Sci Eng* 2016;24(1):153–76. <https://doi.org/10.1080/17415977.2015.1017485>.
- [2] Jamison C, Navarro C, Turner C, et al. The inverse problem utilizing the boundary element method for a nonstandard female torso. *IEEE Trans Biomed Eng* 2010;58(4):876–83. <https://doi.org/10.1109/TBME.2010.2093525>.
- [3] Samaniego E, Anitescu C, Goswami S, et al. An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications. *Comput Methods Appl Mech Eng* 2020;362:.. <https://doi.org/10.1016/j.cma.2019.112790>112790.
- [4] Neuman SP, Fogg GE, Jacobson EA. A statistical approach to the inverse problem of aquifer hydrology: 2. case study. *Water Resour Res* 1980;16:33–58. <https://doi.org/10.1029/WR016i001p0033>.
- [5] Deng J, Lee C. Displacement back analysis for a steep slope at the three gorges project site. *Int J Rock Mech Min Sci* 2001;38:259–68. [https://doi.org/10.1016/S1365-1609\(00\)00077-0](https://doi.org/10.1016/S1365-1609(00)00077-0).
- [6] Banan MR, Banan MR, Hjelmstad K. Parameter estimation of structures from static response. ii: Numerical simulation studies. *J Struct Eng* 1994;120:3259–83. [https://doi.org/10.1061/\(ASCE\)0733-9445\(1994\)120:11\(3259\)](https://doi.org/10.1061/(ASCE)0733-9445(1994)120:11(3259)).
- [7] Alves S, Hall J. System identification of a concrete arch dam and calibration of its finite element model. *Earthq Eng Struct Dyn* 2006;35:1321–37. <https://doi.org/10.1002/eqe.575>.
- [8] Khatir S, Dekemele K, Loccufer M, et al. Crack identification method in beam-like structures using changes in experimentally measured frequencies and Particle Swarm Optimization. *Comptes Rendus Mécanique* 2018;346(2):110–20. <https://doi.org/10.1016/j.crme.2017.11.008>.
- [9] Khatir S, Boutchicha D, Le Thanh C, et al. Improved ANN technique combined with Jaya algorithm for crack identification in plates using XIGA and experimental analysis. *Theor Appl Fract Mech* 2020;107:.. <https://doi.org/10.1016/j.tafmec.2020.102554>102554.
- [10] Tsitsiklis J, Bertsekas D, Athans M. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans Autom Control* 1986;31:803–12. <https://doi.org/10.1109/TAC.1986.1104412>.
- [11] Razavi S, Tolson BA, Burn DH. Review of surrogate modeling in water resources. *Water Resour Res* 2012;48. <https://doi.org/10.1029/2011WR011527>.

- [12] Jin Y. Surrogate-assisted evolutionary computation: Recent advances and future challenges. *Swarm Evol Comput* 2011;1(2):61–70. <https://doi.org/10.1016/j.swevo.2011.05.001>.
- [13] Li YF, Hariri-Ardebili MA, Deng TF, et al. A surrogate-assisted stochastic optimization inversion algorithm: Parameter identification of dams[J]. *Adv Eng Inform* 2023;55:101853.
- [14] Samir K, Brahim B, Capozucca R, et al. Damage detection in CFRP composite beams based on vibration analysis using proper orthogonal decomposition method with radial basis functions and cuckoo search algorithm. *Compos Struct* 2018;187:344–53. <https://doi.org/10.1016/j.compstruct.2017.12.058>.
- [15] Wang C, Li Y, Tran NH, et al. Artificial neural network combined with damage parameters to predict fretting fatigue crack initiation lifetime[J]. *Tribology Int* 2022;175:107854.
- [16] Ghanem R, Saad G, Doostan A. Efficient solution of stochastic systems: application to the embankment dam problem. *Struct Saf* 2007;29:238–51. <https://doi.org/10.1016/j.strusafe.2006.07.015>.
- [17] Guo X, Dias D, Carvajal C, Peyras L, Breul P. Reliability analysis of embankment dam sliding stability using the sparse polynomial chaos expansion. *Eng Struct* 2018;174:295–307. <https://doi.org/10.1016/j.engstruct.2018.07.053>.
- [18] Guo X, Sun Q, Dias D, Antoinet E. Probabilistic assessment of an earth dam stability design using the adaptive polynomial chaos expansion. *Bull Eng Geol Environ* 2020;79:4639–55. <https://doi.org/10.1007/s10064-020-01847-2>.
- [19] YiFei L, Minh HL, Khatir S, et al. Structure damage identification in dams using sparse polynomial chaos expansion combined with hybrid K-means clustering optimizer and genetic algorithm[J]. *Eng Struct* 2023;283:115891.
- [20] Amini A, Abdollahi A, Hariri-Ardebili M, Lall U. Copula-based reliability and sensitivity analysis of aging dams: Adaptive kriging and polynomial chaos kriging methods. *Appl Soft Comput* 2021;109:. <https://doi.org/10.1016/j.asoc.2021.107524>107524.
- [21] Kalimna A, Spada M, Vetsch DF, Marelli S, Whealton C, Burgherr P, et al. Metamodeling for uncertainty quantification of a flood wave model for concrete dam breaks. *Energies* 2020;13(14):3685. <https://doi.org/10.3390/en13143685>.
- [22] Shahzadi G, Soulaïmani A. Deep neural network and polynomial chaos expansion-based surrogate models for sensitivity and uncertainty propagation: An application to a rockfill dam. *Water* 2021;13(13):1830. <https://doi.org/10.3390/w13131830>.
- [23] Hariri-Ardebili MA, Mahdavi G, Abdollahi A, Amini A. An RF-PCE hybrid surrogate model for sensitivity analysis of dams. *Water* 2021;302. <https://doi.org/10.3390/w13030302>.
- [24] Sevieri G, Andreini M, De Falco A, Matthies HG. Concrete gravity dams model parameters updating using static measurements. *Eng Struct* 2019;. <https://doi.org/10.1016/j.engstruct.2019.05.072>109231.
- [25] Sevieri G, De Falco A, Andreini M, Matthies HG. Hierarchical Bayesian framework for uncertainty reduction in the seismic fragility analysis of concrete gravity dams. *Eng Struct* 2021;246:. <https://doi.org/10.1016/j.engstruct.2021.113001>113001.
- [26] Vu VH, Thomas M, Lakis A, Marcouiller L. Operational modal analysis by updating auto regressive model. *Mech Syst Sig Process* 2011;25:1028–44. <https://doi.org/10.1016/j.ymsp.2010.08.014>.
- [27] Liu C, Gu C, Chen B. 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* 2017;33:443–56. <https://doi.org/10.1007/s00366-016-0483-9>.
- [28] Kang F, Liu X, Li J, Li H. Multi-parameter inverse analysis of concrete dams using kernel extreme learning machines- based response surface model. *Eng Struct* 2022;256:. <https://doi.org/10.1016/j.engstruct.2022.113999>113999.
- [29] Kang F, Wu Y, Li J, Li H. Dynamic parameter inverse analysis of concrete dams based on jaya algorithm with gaussian processes surrogate model. *Adv Eng Inf* 2021;49:. <https://doi.org/10.1016/j.aei.2021.101348>101348.
- [30] Bao T, Li J, Lu Y, Gu C. IDE-MLSSVR-based back analysis method for multiple mechanical parameters of concrete dams. *J Struct Eng* 2020;146(8):04020155. [https://doi.org/10.1061/\(ASCE\)ST.1943-541X.0002602](https://doi.org/10.1061/(ASCE)ST.1943-541X.0002602).
- [31] Li H, Wang G, Wei B, Zhong Y, Zhan L. Dynamic inversion method for the material parameters of a high arch dam and its foundation. *App Math Model* 2019;71:60–76. <https://doi.org/10.1016/j.apm.2019.02.008>.
- [32] Liu Y, Wenbo L, Wang G, Linmei L, Chen Y. Structural parameter inversion of a gravity dam based on the dynamic response induced by an underwater explosion 10775463211066992. *J Vibration Control* 2022. <https://doi.org/10.1177/10775463211066992>.
- [33] Wiener N. The homogeneous chaos. *Am J Math* 1938;60:897–936. <https://doi.org/10.2307/2371268>.
- [34] Xiu D, Karniadakis GE. The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 2002;24(2):619–44. <https://doi.org/10.1137/S1064827501387826>.
- [35] Ernst OG, Mugler A, Starkloff H-J, Ullmann E. On the convergence of generalized polynomial chaos expansions. *ESAIM. Mathematical Modelling and Numerical Analysis* 2012;46:317–39. <https://doi.org/10.1051/m2an/2011045>.
- [36] Soize C, Ghanem R. Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM J Sci Comput* 2004;26:395–410. <https://doi.org/10.1137/S1064827503424505>.
- [37] Wei DL, Cui ZS, Chen J. Uncertainty quantification using polynomial chaos expansion with points of monomial cubature rules. *Comput Struct* 2008;86(23–24):2102–8. <https://doi.org/10.1016/j.compstruc.2008.07.001>.
- [38] Blatman G, Sudret B. Adaptive sparse polynomial chaos expansion based on least angle regression. *J Comput Phys* 2011;230:2345–67. <https://doi.org/10.1016/j.jcp.2010.12.021>.
- [39] Fajraoui N, Marelli S, Sudret B. Sequential design of experiment for sparse polynomial chaos expansions. *SIAM/ASA Journal on Uncertainty Quantification* 2017;5:1061–85. <https://doi.org/10.1137/16M1103488>.
- [40] Luthen N, Marelli S, Sudret B. Sparse polynomial chaos expansions: literature survey and benchmark. *SIAM/ASA Journal on Uncertainty Quantification* 2021;9:593–649. <https://doi.org/10.1137/20M1315774>.
- [41] Stefanou G. The stochastic finite element method: past, present and future. *Comput Methods Appl Mech Eng* 2009;198(9–12):1031–51. <https://doi.org/10.1016/j.cma.2008.11.007>.
- [42] Weissman SA, Anderson NG. Design of experiments (DoE) and process optimization. A review of recent publications. *Org Process Res Dev* 2015;19(11):1605–33. <https://doi.org/10.1021/op500169m>.
- [43] Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *Ann Stat* 2004;32(2):407–99. <https://doi.org/10.1214/009053604000000067>.
- [44] Song X, Zhang J, Zhan C, Xuan Y, Ye M, Xu C. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *J Hydrol* 2015;523:739–57. <https://doi.org/10.1016/j.jhydrol.2015.02.013>.
- [45] Nossent J, Elsen P, Bauwens W. Sobol'sensitivity analysis of a complex environmental model. *Environ Model Softw* 2011;26(12):1515–25. <https://doi.org/10.1016/j.envsoft.2011.08.010>.
- [46] Li S, Chen H, Wang M, Heidari AA, Mirjalili S. Slime mould algorithm: A new method for stochastic optimization. *Futur Gener Comput Syst* 2020;111:300–23. <https://doi.org/10.1016/j.future.2020.03.055>.
- [47] Tiachacht S, Khatir S, Thanh CL, et al. Inverse problem for dynamic structural health monitoring based on slime mould algorithm. *Eng Comput* 2022;38(3):2205–28. <https://doi.org/10.1007/s00366-021-01378-8>.
- [48] Blatman G, Sudret B. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probab Eng Mech* 2010;25:183–97. <https://doi.org/10.1016/j.probenmech.2009.10.003>.
- [49] Wagner P, Nagel J, Marelli S, Sudret B. Uqlab user manual-bayesian inference for model calibration and inverse problems. Report No UQLab-V 2019;1:3–113.
- [50] Marelli S, Sudret B. UQLab: A framework for uncertainty quantification in Matlab[M]. *Am Soc Civil Eng* 2014. <https://doi.org/10.3929/ethz-a-010238238>.
- [51] Wei Q, Xu H, Li Y, et al. Identification of Damage on Sluice Hoist Beams Using Local Mode Evoked by Swept Frequency Excitation. *Sensors* 2021;21(19):6357. <https://doi.org/10.3390/s21196357>.
- [52] Hariri-Ardebili MA, Chen S, Mahdavi G. Machine learning-aided PSDM for dams with stochastic ground motions. *Adv Eng Inf* 2022;52:. <https://doi.org/10.1016/j.aei.2022.101615>101615.
- [53] Amini A, Abdollahi A, Hariri-Ardebili MA, et al. Copula-based reliability and sensitivity analysis of aging dams: Adaptive Kriging and polynomial chaos Kriging methods. *Appl Soft Comput* 2021;109:. <https://doi.org/10.1016/j.asoc.2021.107524>107524.