Extreme-oriented sensitivity analysis using sparse polynomial chaos expansion
Application to train–track–bridge systems

Shang, Yue; Nogal, Maria; Teixeira, Rui; Wolfert, A. R. (Rogier) M.

DOI
10.1016/j.ress.2023.109818

Publication date
2023

Document Version
Final published version

Published in
Reliability Engineering and System Safety

Reliability Engineering and System Safety 243 (2024) 109818

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Extreme-oriented sensitivity analysis using sparse polynomial chaos expansion. Application to train–track–bridge systems

Yue Shang [a,*], Maria Nogal [a], Rui Teixeira [b], A.R. (Rogier) M. Wolfert [a]

[a] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, Delft, 2628 CN, The Netherlands
[b] School of Civil Engineering, University College Dublin, Richview Newstead, Belfield, Dublin 4, 695014, Ireland

ARTICLE INFO

Keywords:
Global sensitivity analysis
Extreme value
Polynomial chaos expansion
Train–track–bridge system
Optimization
Limit state

ABSTRACT

The use of sensitivity analysis is essential in model development for the purposes of calibration, verification, factor prioritization, and mechanism reduction. While most contributions to sensitivity methods focus on the average model response, this paper proposes a new sensitivity method focusing on the extreme response and structural limit states, which combines an extreme-oriented sensitivity method with polynomial chaos expansion. This enables engineers to perform sensitivity analysis near given limit states and visualize the relevance of input factors to different design criteria and corresponding thresholds. The polynomial chaos expansion is used to approximate the model output and alleviate the computational cost in sensitivity analysis, which features sparsity and adaptivity to enhance efficiency. The accuracy and efficiency of the method are verified in a truss structure, which is then illustrated on a dynamic train–track–bridge system. The role of the input factors in response variability is clarified, which differs in terms of the design criteria chosen for sensitivity analysis. The method incorporates multi-scenarios and can thus be useful to support decision-making in design and management of engineering structures.

1. Introduction

Sensitivity analysis (SA) provides an understanding of how a given model responds to changes in its input factors (i.e., variables and parameters), which allows for prioritizing the factors, reducing the model dimensionality, calibrating the model, and evaluating the consistency between the model input and output. Based on the factor space of interest, sensitivity methods can be classified into two categories: local and global approaches. Local measures assess the impact of variations in input factors near their nominal values, while global methods evaluate model output sensitivity to input factors over their entire domain of variation.

Various global sensitivity methods have been proposed, including variance-based methods (e.g., the Sobol method [1] and the Fourier amplitude sensitivity (FAST) method [2]), regression-based methods (e.g., standardized regression coefficient [3]), moment-independent methods (e.g., Borgonovo indices [4]), among others. This paper focuses on the variance-based approach, which leverages the functional decomposition of the variance measuring the contribution of each factor or their combinations to the output variance. Sobol's indices [1] and related total-effect indices are commonly used variance-based measures. These measures can be computed through sampling-based methods [5] or metamodels [6,7]. The sampling-based methods can come at a significant computational cost due to the large number of function evaluations required, which may render the methods infeasible for computationally demanding models.

Metamodels have received much attention for the computation of variance-based measures due to their potential to alleviate computational costs. Specifically, polynomial chaos expansion (PCE) and its variants have been used to derive Sobol's indices through post-processing of model coefficients. This approach is originally shown in [6], using the coefficients of generalized PCE [8], which is extended using sparse PCE [9,10] and partial least squares-driven PCE (PLS-PCE) [7] to deal with high-dimensional problems. Furthermore, advances have been made to account for the dependence in input factors [11], generalized modeling of both aleatory and epistemic uncertainty in input factors [12], and derivative-based sensitivity measures for efficient screening of unimportant factors [13]. These studies have demonstrated PCE to be a versatile and efficient tool for sensitivity analysis.

Variance-based sensitivity measures aim to evaluate the influence of input factors on the variance of a quantity of interest (QoI). Most often, the QoI refers to the average output of a model, as used in the Sobol indices and their extensions. These indices, therefore, are designed to capture the mean behavior of systems. However, they may

* Corresponding author.
E-mail address: y.shang@tudelft.nl (Y. Shang).

https://doi.org/10.1016/j.ress.2023.109818
Received 7 June 2023; Received in revised form 10 November 2023; Accepted 13 November 2023
Available online 15 November 2023

not be sufficient for analyzing safety-critical structures such as railway tracks.

To address this issue, this paper focuses on the extreme values of the structural response. Here, *extreme values* are the maximum and minimum levels the model output can achieve. *Response* refers to the output of a model that characterizes the structural behavior, which can be either static or dynamic. The condition of a structure can be described by a group of limit states, each associated with design criteria (or indicators) and thresholds. *Failure* occurs if the response quantity exceeds a given threshold.

This concept is closely linked to reliability analysis, making reliability-oriented sensitivity analysis (ROSA) the natural way to perform SA for safety-critical structures. ROSA is concerned with evaluating the sensitivity of the output of reliability analysis to the model inputs, where the part of the output space that yields failure is of interest (cf. [14–16]). These studies aim to identify critical input factors with more potential to reduce the probability of failure, instead of quantifying the influence of input factors on output extreme values. In parallel, some work directs sensitivity evaluation to quantile-based [17] and high-order measures (e.g., skewness and kurtosis decomposition [18]), which may provide an indication about input factors that drive the output towards its extreme values. However, they do not directly address sensitivity analysis near output extreme values.

Limited research has focused on sensitivity methods for extreme values. In [19], extremum sensitivities are computed by constraining the input distribution to regions leading to the output extremum. Monte Carlo filtering is used to model the conditional distribution, which is a sampling-based approach that partitions model realizations within or outside targeted regions in the input space. Another approach is optimization-based [20]. It involves discretizing an input factor within its range and searching for the output extremum at each discrete point by fixing the target factor while perturbing the rest (the so-called *A-1AT - all minus one at a time*). Then, a curve can be defined by the extreme values computed from all discrete points, and the variation of this curve is evaluated to quantify the main effect of the target factor on the output extreme values.

In this paper, we aim to advance the field of extreme-oriented SA by introducing a new sensitivity method called the *threshold-based sensitivity method*. This method allows for evaluating the model sensitivity near structural limit states by generalizing the optimization-based sensitivity method [20] to a wider range of targeted portions of the output space, from extreme values to thresholds. The threshold-based method offers flexibility by allowing the evaluation of multiple decision thresholds in SA. It also enables visualizing the structural performance (failure or non-failure) around limit states in the input space, highlighting critical regions of the input space that may lead to system failure under a given threshold. Additionally, we investigate the performance of PCE in extreme-based SA. While the use of PCE for Sobol's indices has been relatively well-studied, research on its effectiveness in extreme-based cases remains scarce. Therefore, we compare the performance of PCE with other commonly used metamodels, such as Kriging and PC-Kriging, to provide insights into its suitability for these types of analyses.

Further, we aim to apply this methodology in vehicle–structure interaction (VSI) systems, which are safety-critical and highly relevant to reliability-based design (cf. [21]). Assessing sensitivity near the extremes can provide valuable insights for decision-making in the design and management of such systems. A common approach for modeling VSI systems involves coupling a multibody (vehicle) system with a structure represented by a finite element (FE) formulation. This formulation involves numerous input factors related to geometry and material properties, which are inherently random due to material variability and manufacturing-induced tolerances, making it challenging to select appropriate values for these factors.

In most cases, the sensitivity of vehicle/structure responses to input factors is assessed using the sampling method, known as One At a Time (OAT) analysis. OAT implies the SA is approached by perturbing one single factor at each time, while the rest are fixed in a given value. This method is computationally efficient and well-suited for SA of VSI problems since the models involved can easily become computationally intensive. However, it may not provide reliable results when nonlinear terms are present in the model [22,23], which, in the VSI models, can be attributed to factors such as wheel–rail contact [24] and railpads [25].

A few studies have applied global sensitivity methods to evaluate VSI systems. Xu et al. [26] evaluate a nonlinear vehicle–track model to identify the factors dominating the system dynamics, with special attention on track irregularities. Later, Wan et al. [22] performed a dynamic global SA in a time-varying train–track–bridge system, comparing a Kriging-based approach against a Monte Carlo simulation scheme. Recently, sensitivity and uncertainty analyses were performed in tandem [27,28] to analyze the uncertainty propagation in a train–track–bridge system and identify factors that are most responsible for the response uncertainty. However, these studies only directed sensitivity analyses to the average system response using the Sobol method, while overlooking the part of the output space that yields failure.

The rest of the paper is organized as follows. Section 2 presents the PCE-based sensitivity method for extreme problems, including the A-1AT method with two formulations (Section 2.1), the basics of polynomial chaos expansion (Section 2.2), and the extension to the threshold-based sensitivity method (Section 2.3). The methodology is verified in a truss structure (Section 3) and further applied to a dynamic train–track–bridge system to demonstrate the applicability of the proposed method (Section 4). Section 5 discusses the sensitivity results along with evaluating the predictive performance of PCE metamodels and the impact of design thresholds on the sensitivity results. Final remarks and future research lines are drawn in Section 6.

2. Methodology

2.1. Extreme-oriented sensitivity analysis

In this section, the A-1AT sensitivity method with two strategies to formulate the extreme problems are presented: the original problem formulation in [20] is provided in Section 2.1.1 for readability and its extension to threshold-based sensitivity analysis is presented in Section 2.1.2.

2.1.1. A-1AT upon the output extrema

Let us consider a system whose behavior is described by $g(\cdot)$, where a set of input factors $\boldsymbol{X} = \{X_1, X_2, \dots, X_d\} \in \mathbb{R}^d$ yields a scalar output Y such that

$$Y = g(\boldsymbol{X}). \tag{1}$$

Each input factor is defined within an interval $X_i \in [X_i^l, X_i^u]$. First, when a factor X_i is fixed at a specific point x_i^0, the model defined in Eq. (1) can reach an extreme value (either maximum or minimum) by varying the remaining factors within their corresponding intervals. This extreme value represents an output extremum of the reduced (i.e., $d-1$) dimensional space, which can be denoted as $\bar{Y}_i$. Then, when considering the entire range of the factor X_i, we can determine a curve formed by the output extrema of the $d-1$ dimensional space (see Fig. A.12 for an illustrative example). The sensitivity of this curve at point x_i^0 can be defined in terms of finite differences, i.e.,

$$S_i^{Ext}\left(x_i^0; X_{\sim i}\right) \simeq \frac{\Delta \bar{Y}_i\left(x_i^0, X_{\sim i}\right)}{\Delta x_i}, \qquad x_i^0 \in \left[X_i^l, X_i^u\right], \ \forall X_i. \tag{2}$$

The procedure of evaluating the sensitivity of $\bar{Y}_i$ to X_i (Eq. (2)) is provided in the following steps.

Step a. Discretize the factor X_i in n_i points within its range $[X_i^l, X_i^u]$; denote each discrete point of X_i by $x_{i,j}, j = 1, \dots, n_i$.

Step b. For a discrete point $x_{i,j}$, the extreme value of Eq. (1) is obtained by fixing X_i at $x_{i,j}$ while varying the remaining factors $X_{\sim i}$ in their corresponding intervals. This poses an optimization problem for each discrete value of X_i, which is defined by

$$\begin{aligned} \min_{\mathbf{X}\in[\mathbf{X^l},\mathbf{X^u}]} \quad & g(\mathbf{X}) \\ \text{s.t.} \quad & X_i = \{x_{i,j}\}, \quad j = 1,\dots,n_i, \\ & \mathbf{X}_{\sim i} \in \left[\mathbf{X}^l_{\sim i}, \mathbf{X}^u_{\sim i}\right]. \end{aligned} \tag{3}$$

Step c. Save the optimal (i.e., extreme) value at each discrete value of X_i.

Step d. Once the optimal value is determined for all discrete values of X_i, the curve formed by the output extrema of the $d-1$ dimensional space, $\bar{Y}_i$, can be determined and the sensitivity of the curve is calculated according to Eq. (2).

Specifically, in Step *a*, different types of discretization can be selected for the input factors, such as equal-, log-, or randomly spaced discretization. The degrees of discretization, meaning the number of discrete points, can also vary for each input factor. Since the optimal search is defined for a specific factor (X_i) at a time, the search process is independent of the discretization strategy for the remaining non-fixed factors ($\mathbf{X}_{\sim i}$). The reader is referred to [20] for further details about the discretization process.

The optimization problem in Eq. (3) can be solved using either gradient-based methods or (meta)heuristics. As the method is optimization-based and our interest here is either the maximum or minimum, the term *optimal value* also refers to *extreme value* in Steps *c* and *d*. It is also worth noting that we consider the minimum value by default to conform to the standard form of defining an optimization problem, as shown in Eq. (3). A maximization problem can be treated by negating the objective function $g(\mathbf{X})$.

Further, the following importance measure [20] is defined to rank the input factors according to their contribution to the total variance introduced by the individual factors to the output extreme values,

$$I_i^{Ext} = \frac{\mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right)}{\sum_i \mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right)}, \tag{4}$$

where $\mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right)$ represents the variance of the output extreme values associated with factor X_i. $\sum_i \mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right)$ represents the total variance introduced by the individual factors to the output extreme values.

As mentioned earlier, the input factors can be discretized either uniformly or non-uniformly. If non-uniform discretization is used, the variance in Eq. (4) must be replaced by the weighted variance, i.e.,

$$\mathrm{Var}_{X_i}\left(\bar{Y}|x_i\right) = \sum_{j=1}^{n_i} \frac{w_j}{\sum_{j=1}^{n_i} w_j}\left(\bar{Y}_{i,j} - \bar{\mu}\right)^2, \quad \bar{\mu} = \sum_{j=1}^{n_i} \frac{w_j}{\sum_{j=1}^{n_i} w_j}\bar{Y}_{i,j}, \tag{5}$$

where $\bar{Y}_{i,j}$ is the output extremum at the discrete point $x_{i,j}$, and the weight w_j is given by

$$w_j = \frac{1}{2}\left(x_{i,j+1} - x_{i,j-1}\right), \quad \forall j \in [2, n_i - 1], \tag{6}$$

and for the first point ($j = 1$) of factor X_i, we have $w_1 = \frac{1}{2}\left(x_{i,2} - x_{i,1}\right)$. For the last point ($j = n_i$) of factor X_i, we have $w_{n_i} = \frac{1}{2}\left(x_{i,n_i} - x_{i,n_i-1}\right)$.

2.1.2. A-1AT upon the deviation between the output and given threshold

A limit state refers to a state of impending failure, beyond which a structure can no longer perform its intended function satisfactorily, in terms of either safety or serviceability. The basic idea of the limit-state design approach is to identify all possible modes of failure and determine acceptable levels of safety against the occurrence of each limit state. From this, one limit state function can be evaluated by

$$G_k(\mathbf{X}) = g(\mathbf{X}) - t_k \tag{7}$$

where $G_k(\mathbf{X})$ denotes the limit state function of a structure; $g(\mathbf{X})$ describes the actual performance of the structure measured by a design criterion, and $t_k (k = 1,\dots,\kappa)$ represents the kth allowable level (i.e., the threshold) of the corresponding criterion, e.g., maximum allowable displacement of a beam. κ denotes the number of decision thresholds considered for a single criterion.

The sign of $G_k(\mathbf{X})$ determines the structural reliability state, which is defined as

- failing state if $G_k(\mathbf{X}) > 0$;
- reliable state if $G_k(\mathbf{X}) < 0$;
- limit state if $G_k(\mathbf{X}) = 0$.

The structure needs to be checked for all groups of limit states to ensure sufficient margins between the actual structural behavior and given threshold. However, the determination of the structural state can be significantly affected by the attitudes of decision-makers towards the threshold levels. For example, ride comfort is relevant to structural serviceability, which is associated with a group of evaluation indices and rating scales (i.e., thresholds). The demand for traveling quality is however very subjective, and the decision on which rating scale to be included in the design process can influence the outcomes significantly, which may also affect the sensitivity analysis result.

To circumvent this issue, we propose to frame the preferences of the decision-makers into multiple decision thresholds, and the sensitivity of the model output to each input factor is evaluated for each level. For this purpose, the extreme problem is reformulated by modifying Step *b* in Section 2.1.1 as follows.

$$\begin{aligned} \min_{\mathbf{X}\in[\mathbf{X^l},\mathbf{X^u}]} \Delta(\mathbf{X}) \quad &= \min_{\mathbf{X}\in[\mathbf{X^l},\mathbf{X^u}]} \left(g(\mathbf{X}) - t_k\right)^2 \\ \text{s.t.} \quad & X_i = \{x_{i,j}\}, \quad j = 1,\dots,n_i, \\ & \mathbf{X}_{\sim i} \in \left[\mathbf{X}^l_{\sim i}, \mathbf{X}^u_{\sim i}\right], \end{aligned} \tag{8}$$

where $\Delta(\mathbf{X})$ represents the squared deviation of the model output from a given threshold, that is, $\Delta(\mathbf{X}) = \left[G_k(\mathbf{X})\right]^2$.

Recall that for a single design criterion, the output space is defined by $Y = g(\mathbf{X})$, the present sensitivity method essentially requires two basic steps: (1) identify the surface of interest in Y, and (2) analyze the sensitivity of that surface. In the first step, the surface can be characterized either by the output extrema through optimization (Eq. (3)) or by being related to a given threshold (Eq. (8)). If the surface related to a threshold is of interest, the idea is to search for the model output that approaches the reference threshold, and the goal is to identify which input factors are critical in reaching that threshold, i.e., the limit state of the structure. This process, which allows for analyzing the sensitivity of the surface around the given threshold, is referred to as *threshold-based sensitivity*.

An illustrative example of threshold-based sensitivity is shown in Fig. 1. The figure depicts how changes in a single input factor $X \in \left[X^l, X^u\right]$ affect the *threshold-based response* (calculated from Eq. (8)) near three different design thresholds ($t_k, k = 1,2,3$). Specifically, t_1 is easy to reach, t_2 is a middle level, and t_3 is extremely hard to achieve. These thresholds correspond to Scenario 1, 2, and 3, respectively. The vertical axis ($\tilde{Y}^k$) represents the threshold-based response, which is denoted by a solid line for each threshold level (the dotted line with the same color). The value of $\tilde{Y}^k$ varies depending on whether the maximum output exceeds the corresponding threshold. In other words, it poses a binary classification problem, where (1) for solid lines that fall below the corresponding threshold, the value of $\tilde{Y}^k$ represents the actual maximum output of a system, indicating a reliable state; (2) When a solid line remains stable at a given threshold, it suggests that the maximum output has either reached or exceeded the threshold, indicating the limit or failing state of the system. In such cases, the value of $\tilde{Y}^k$ is determined by the threshold value.

Note that the term *response* is used in the context of threshold-based SA, specifically referring to the threshold-based response, instead of

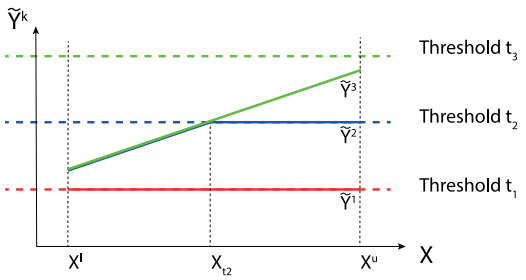

Fig. 1. Illustrative example of threshold-based sensitivity. The figure shows three scenarios: Scenario 1 (red lines) with threshold t_1 (dotted) and threshold-based response $\tilde{Y}^1$ (solid); Scenario 2 (blue lines) with threshold t_2 (dotted) and threshold-based response $\tilde{Y}^2$ (solid); and Scenario 3 (green lines) with threshold t_3 (dotted) and threshold-based response $\tilde{Y}^3$ (solid). In each scenario, the region above or below the corresponding threshold indicates the failure or non-failure domain, respectively. $\tilde{Y}^2$ is equal to $\tilde{Y}^3$ when $X \in [X^l, X_{t2})$, as they both represent the maximum output within this range, which is independent of the threshold values. Note: the figure is only illustrative. See Section 5.3 for a detailed analysis of a large-scale numerical example.

the *output* used in the general context of SA. This distinction is made because the threshold-based response is not exactly equivalent to the model output. Its value depends on whether the system reaches the failure threshold. Therefore, in this paper, the term *response* is used in two specific contexts: (1) in structural engineering problems where it refers to the reaction or behavior of a structure to the input it receives, as mentioned in Section 1, and (2) in threshold-based SA where the term threshold-based response is introduced to visualize the impact of input factors on the model output near the thresholds, which is elaborated in the following.

As shown in Fig. 1, in Scenario 1, the value of $\tilde{Y}^1$ (the red solid line) is constant at the level of t_1, indicating that the current threshold t_1 can always be reached for the factor X over its range of definition, $[X^l, X^u]$. Therefore, X is not a critical factor that would impede reaching the threshold at t_1. In Scenario 2, the response surface lies below t_2 when $X \in [X^l, X_{t2})$. This suggests that ensuring $X \in [X^l, X_{t2})$ can create a safe margin between the maximum output (the blue solid line) and the threshold (the blue dotted line). However, when $X \in [X_{t2}, X^u]$, the limit state is (reached or) exceeded, resulting in (impending) system failure. Therefore, X is critical in reaching the threshold t_2 over its entire range of definition $[X^l, X^u]$. Herein, we see that the threshold-based SA allows us to visualize how the input factors affect the model output near the limit states, indicating areas of the input space that may lead to undesirable outcomes, such as system failure. This knowledge can be useful in decision-making and risk management for the design and maintenance of engineering structures.

If the threshold is sufficiently large at t_3, the value of $\tilde{Y}^3$ is determined by the maximum output of the system (the green solid line). Varying $X \in [X^l, X^u]$ will not alter the system reliability state, as there are sufficient margins between the maximum and t_3. Note that in this case, the problem is equivalent to maximizing $g(\mathbf{X})$, and vice versa. This corresponds exactly to the original formulation in Eq. (3). The threshold-based sensitivity calculated through Eq. (8) essentially represents an alternative way of describing the extreme problems, while it extends the original setting by taking the limit state concept into account and is flexible to allow for assessing multiple decision-making scenarios, as measured by multi-thresholds, in the SA.

An overview of the extreme-based and threshold-based formulation is provided in Fig. 2. Both formulations require discretizing the input factors as the first step, and the discretization strategy remains consistent. However, they differ in terms of the target portion of the output space (Step 2), formulation of the optimization problem (Step 3), and the way the calculation results are post-processed (Step 4). The extreme-based formulation is mainly used to quantify and visualize the main effects of the input factors on the extreme values of the output. On the other hand, threshold-based sensitivity focuses on visualizing the impact of altering an individual input factor over a specified threshold by introducing the threshold-based response. This formulation clearly indicates the system reliability state around limit states in the input space, highlighting critical regions that may lead to system failure under a given threshold. This knowledge can be leveraged to support analyses centered around system reliability, such as reliability-based design optimization. Also, the evaluation results obtained from both formulations complement each other, providing valuable insight into the contribution of the input factors across different regions of the output space.

2.2. Polynomial chaos expansions

Let us recall the mapping $g(\cdot)$ in Eq. (1) to describe the behavior of a system. Assume that there is uncertainty associated with the input vector $\mathbf{X}$, which can be described by a random vector with joint PDF $f_{\mathbf{X}}(\mathbf{x})$ and marginal PDFs $f_{X_i}(x_i), i = 1, \ldots, d$. Here, we consider that the components of $\mathbf{X}$ are independent, which is the case for the models in the present study. Consequently, the model output Y is also a random variable that can be approximated using the PCE [29] such that

$$g(\mathbf{X}) \approx \hat{g}(\mathbf{X}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \mathbf{c}_{\boldsymbol{\alpha}} \boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\mathbf{X}), \tag{9}$$

where $\hat{g}(\mathbf{X})$ denotes the PCE approximation; $\mathcal{A}$ is a set of multi-indices $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_d\}$, $\mathbf{c} = \{\mathbf{c}_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathcal{A}\}$ are polynomial coefficients to be computed, and $\boldsymbol{\Psi} = \{\boldsymbol{\Psi}_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathcal{A}\}$ are multivariate polynomials that are orthogonal with respect to $f_{\mathbf{X}}(\mathbf{x})$. The independence of input variables allows us to construct the multivariate polynomials as the tensor product of univariate orthonormal polynomials with respect to $f_{X_i}(x_i)$, i.e.,

$$\boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\mathbf{X}) = \prod_{i=1}^{d} \Psi_{\alpha_i}(X_i), \tag{10}$$

where $\Psi_{\alpha_i}(X_i)$ is a polynomial of degree α_i in the ith input variable. The expansion is originally formulated with standard Gaussian random variables with Hermite polynomials [30]. It was later extended into a broader framework known as the *generalized* PCE [8] to employ basis functions from the Askey scheme of orthogonal polynomials with their underlying random variables. For example, Legendre polynomials can be associated with uniform random variables. If other types of random variables appear in the input vector $\mathbf{X}$, it is possible to perform an isoprobabilistic transform such that the generalized PCE can be applied to this variable [29,31].

For computational purposes, the PC expansion in Eq. (9) has to be truncated, where the polynomials $\{\boldsymbol{\Psi}_{\boldsymbol{\alpha}}, \boldsymbol{\alpha} \in \mathcal{A}\}$ are generally retained with total degree up to p such that

$$\begin{aligned} \mathcal{A}^{d,p} &= \left\{ \boldsymbol{\alpha} \in \mathbb{N}^d : \quad \|\boldsymbol{\alpha}\|_1 = \sum_{i=1}^{d} \alpha_i \leq p \right\}, \\ \mathbf{card}\, \mathcal{A}^{d,p} &\equiv P = \binom{d+p}{p}, \end{aligned} \tag{11}$$

where $\|\boldsymbol{\alpha}\|_1$ denotes the degree of the multi-indices $\boldsymbol{\alpha}$; $\mathbf{card}\, \mathcal{A}^{d,p}$ represents the number of multi-indices (i.e., the number of coefficients in the PCE). The coefficient vector $\mathbf{c}_{\boldsymbol{\alpha}}$ can be determined through regression approaches [9,29]. However, considering the size of the basis ($\mathbf{card}\, \mathcal{A}^{d,p}$) in Eq. (11), the computational effort in the regression method grows dramatically with the size of d or p, which makes the full PCE intractable in the high-dimensional problems (e.g., $d \geq 10$ or $p \geq 10$) [29].

This limitation was addressed by a hyperbolic truncation scheme [29], which defines a new set $\mathcal{A}_q^{d,p}$ of multi-indices as

$$\mathcal{A}_q^{d,p} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^d : \quad \|\boldsymbol{\alpha}\|_q = \left(\sum_{i=1}^{d} \alpha_i^q \right)^{1/q} \leq p \right\}. \tag{12}$$

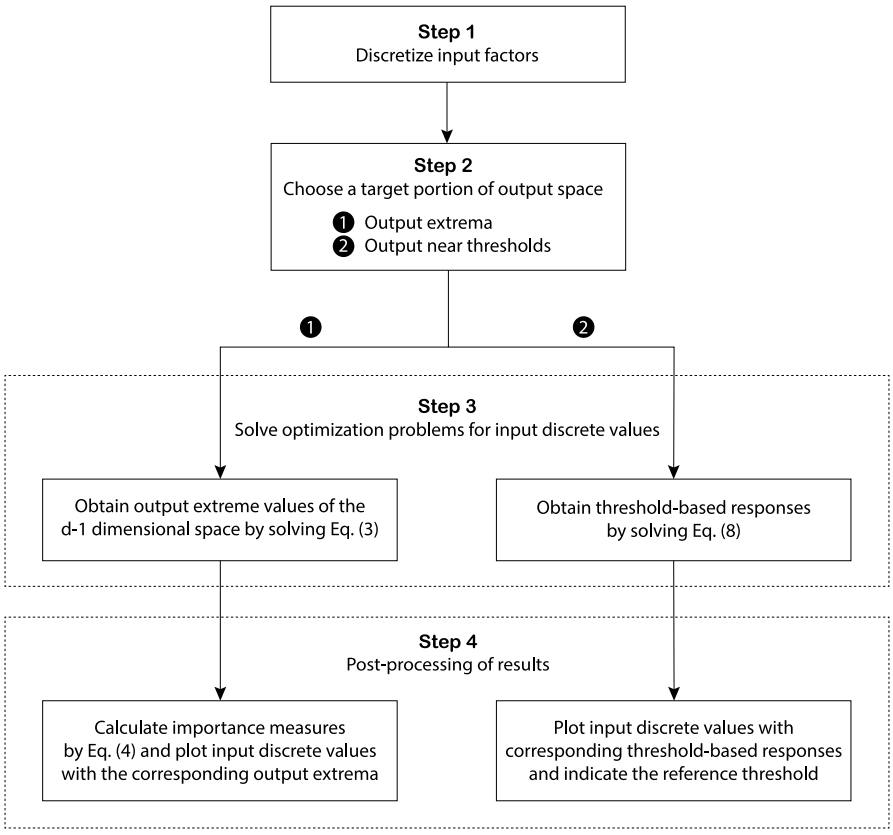

Fig. 2. Overview of the extreme-based (Path 1) and threshold-based (Path 2) formulation.

The multi-indices are determined based on q $(0 < q < 1)$ [29]. For $q = 1$, the hyperbolic truncation corresponds to the standard total-degree truncation degree in Eq. (11), where the polynomials of maximum total degree of p are retained. When $q < 1$, the truncation penalizes high-degree terms with many interacting variables, while favoring the main effects and low-order interactions. The reader is referred to [29] for further details.

Apart from the hyperbolic truncation scheme, an *adaptive* algorithm contributes further to an efficient procedure for the selection of polynomials [29]. The algorithm is based on the *Least Angle Regression* (LAR) that iteratively enhances the polynomials under construction [29]. In brief, it provides a collection of PC representations (metamodels) in such a way that terms in $\mathcal{A}_q^{d,p}$ are added one by one, and at each iteration, the metamodel under construction is given an error estimate ε. By defining $\varepsilon^* = \min(\boldsymbol{\varepsilon})$, one stops the algorithm if ε^* is less than the preset target error ε_{tgt}. The PC representation with ε^* is eventually retained, associated with the optimal subset $\mathcal{A}^*$. It is said to be *sparse* since it contains a reduced number of terms in $\mathcal{A}^*$ compared to a full representation in Eq. (9).

Once LAR provides a selected set of terms at each iteration, the coefficients of the related PC representation and the corresponding ε can be computed by ordinary least-square regression (OLS), which follows the so-called *hybrid* LAR proposed in [32].

Let $\{\mathbf{x}_{(1)}, \ldots, \mathbf{x}_{(N)}\}$ denote a set of N input realizations from an *experimental design* (ED), and $\{y_{(1)}, \ldots, y_{(N)}\}$ be the corresponding model evaluations, i.e., $\{y_{(\varrho)} = g(\mathbf{x}_{(\varrho)}), \varrho = 1, \ldots, N\}$. The PC approximation is calculated by Eq. (9), where the coefficients $\mathbf{c}$ are chosen by minimizing the mean-square error between the exact value and its PC approximation, i.e.,

$$\mathbf{c} = \arg \min_{\mathbf{c} \in \mathbb{R}^{\mathrm{card}\,\mathcal{A}}} \mathbb{E}\left[\left(g(\mathbf{X}) - \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \mathbf{c}_{\boldsymbol{\alpha}} \boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\mathbf{X})\right)^2\right], \tag{13}$$

where $\mathbb{E}$ is the mathematical expectation. The solution of Eq. (13) can be obtained based on the OLS estimates,

$$\hat{\mathbf{c}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{Y}, \tag{14}$$

where $\mathbf{A}$ is a data matrix of size $N \times P$ and its general entry is defined by

$$A_{\varrho j} = \Psi_{\alpha_j}(\mathbf{x}_{(\varrho)}), \quad \varrho = 1, \ldots, N;\ j = 0, \ldots, P-1. \tag{15}$$

The accuracy of each PC representation in the LAR is evaluated by the *relative leave-one-out error* estimate, denoted by ε as mentioned above. Let $\hat{g}_{(-\varrho)}$ be the metamodel that is constructed from the ED while removing the ϱth observation. The leave-one-out error is defined as

$$\varepsilon_{LOO} = \frac{1}{N}\sum_{\varrho=1}^{N}\left[g(\mathbf{x}_{(\varrho)}) - \hat{g}_{(-\varrho)}(\mathbf{x}_{(\varrho)})\right]^2, \tag{16}$$

where $g(\mathbf{x}_{(\varrho)})$ and $\hat{g}_{(-\varrho)}(\mathbf{x}_{(\varrho)})$ represents the model evaluation at $\mathbf{x}_{(\varrho)}$, and its prediction from $\hat{g}_{(-\varrho)}$, respectively. Then the relative leave-one-out error can be given by

$$\varepsilon = \frac{\varepsilon_{LOO}}{\mathrm{Var}_Y}, \tag{17}$$

where Var_Y denotes the empirical variance of the output Y, calculated by

$$\mathrm{Var}_Y = \frac{1}{N-1}\sum_{\varrho=1}^{N}(g(\mathbf{x}_{(\varrho)}) - \mu_Y)^2, \qquad \mu_Y = \frac{1}{N}\sum_{\varrho=1}^{N} g(\mathbf{x}_{(\varrho)}). \tag{18}$$

In case an independent dataset is available next to the training and validation set (used to construct metamodels), the *relative generalization error*, ε_{gen}, is a measure commonly used to quantify the accuracy and predictive quality of the metamodels, which is given by

$$\varepsilon_{gen} = \frac{\mathbb{E}[(g(\mathbf{x}) - \hat{g}(\mathbf{x}))^2]}{\mathrm{Var}_Y}. \tag{19}$$

2.3. Threshold-based sensitivity analysis using PCE

The proposed method seeks to efficiently evaluate the model sensitivity to the input factors under different design thresholds. The computation process is presented in Algorithm 1, including two main procedures: construction of PCE (Procedure 1) and threshold-based sensitivity evaluation (Procedure 2).

Algorithm 1: PCE-based simulation scheme for sensitivity analysis near limit states

Input: PDFs of input factors $(f_{X_i}(x_i), i = 1, ..., d)$
Output: Threshold-based response $(\tilde{Y}^k_{i,j})$
Procedure 1 Construct a PC approximation of the response for a design criterion
1 $\varepsilon_{tgt}, p_{max}, q, N \leftarrow$ Preset the algorithm parameters
2 $\{\mathbf{x}_{(\varrho)}, \varrho = 1, ..., N\} \leftarrow$ Read statistic inputs $\mathbf{X}$ with specified $f_{X_i}(x_i)$; select an ED with sample size N
3 $\{y_{(\varrho)}, \varrho = 1, ..., N\} \leftarrow$ Collect the model evaluations
4 **for** $p = 1 \rightarrow p_{max}$ **do**
5 $\mathcal{A}^{d,p}_q \leftarrow$ Gather a candidate set of p-order polynomials
6 $\mathcal{A}^{(p)}, \varepsilon^{(p)} \leftarrow$ Apply LAR to the candidate set $\mathcal{A}^{d,p}_q$
7 **if** $\varepsilon^{(p)} < \varepsilon_{tgt}$ **then**
8 $\mathcal{A}^* = \mathcal{A}^{(p)}$, $\varepsilon^* = \varepsilon^{(p)}$
9 **break**
10 **else** $p = p + 1$
11 **end if**
12 **end for**
13 $\mathbf{c} \leftarrow$ Compute the coefficients associated with $\mathcal{A}^*$ according to Eq. (13)
14 $\hat{g} \leftarrow$ Retain the final PC approximation with $\mathcal{A}^*, \mathbf{c}, \varepsilon^*$ according to Eq. (9)
End procedure 1
Procedure 2 Evaluate the threshold-based sensitivity on $\hat{g}$
15 $t_k, n_i, \kappa \leftarrow$ Preset the algorithm parameters
16 $X^l_i, X^u_i \leftarrow$ Define bounds for each input factor X_i according to $f_{X_i}(x_i)$
17 $x_{i,j} \leftarrow$ Discretize X_i in n_i points within $[X^l_i, X^u_i]$
18 **for** $j = 1 \rightarrow n_i$, $k = 1 \rightarrow \kappa$ **do**
19 $\bar{\chi}^k \leftarrow$ Minimize $((\hat{g} - t_k)^2, X_i = \{x_{i,j}\}, X_{\sim i} \in [X^l_{\sim i}, X^u_{\sim i}])$ according to Eq. (8)
20 $\tilde{Y}^k_{i,j} \leftarrow$ Collect $\bar{\chi}^k$
21 **end for**
End procedure 2

Procedure 1 mainly follows the techniques proposed in [29], as elaborated in Section 2.2. To enhance readability, the algorithmic framework for constructing the PCE is presented here. Procedure 1 starts with selecting values of the algorithm parameters (*line 1*). It chooses an ED (e.g., Latin hypercube sequence) and evaluates the model output at the corresponding sampled points (*line 2–3*). The least angle regression (LAR) is applied to select the optimal set of the basis $\mathcal{A}^*$, which requires an iterative procedure (*line 4–12*) in the following:

Initially, a candidate set of p-order ($p = 1$) polynomials is determined by the hyperbolic truncation scheme, i.e., the set $\mathcal{A}^{d,p}_q$ defined by q-norm according to Eq. (12) (*line 5*). The LAR is applied to $\mathcal{A}^{d,p}_q$ to select the optimal set $\mathcal{A}^{(p)}$ with the lowest error $\varepsilon^{(p)}$ according to Eq. (16) (*line 6*). Then, $\varepsilon^{(p)}$ is checked against the target error ε_{tgt}. If $\varepsilon^{(p)} < \varepsilon_{tgt}$, it stops the iterative process (*line 9*). Otherwise $p = p + 1$ (*line 10*), and repeat the process (*line 4–12*). The optimal set $\mathcal{A}^*$ is eventually retained, associated with the lowest error ε^* (*line 8*). The corresponding coefficients are determined by OLS estimators (*line 13*). The final PCE ($\hat{g}$) can be determined (*line 14*). Note that it is possible for the error $\varepsilon^{(p)}$ to increase from a certain order p, which can be attributed to overfitting. To ensure proper convergence of the algorithm, an early stop criterion is introduced. This criterion terminates the process if the error $\varepsilon^{(p)}$ increases for at least two subsequent iterations ($\varepsilon^{(p)} \geq \varepsilon^{(p-1)} \geq \varepsilon^{(p-2)}$). For further details, the reader is referred to [29]. For the implementation of this algorithm, this study sets the target error ε_{tgt} equal to 0 to minimize the error in the attained PCE metamodels.

Procedure 2 evaluates the threshold-based sensitivity (Section 2.1.2) on $\hat{g}$. It also begins by presetting the parameter values (*line 15*). Then, for non-uniformly distributed input factors, the lower (respectively upper) bound of X_i is determined by the 1st percentile (respectively 99th percentile) of the corresponding probability distribution (*line 16*). X_i is then discretized within these bounds into n_i points (*line 17*), as detailed in Section 2.1.1. For each discrete value of X_i and threshold level t_k, the squared deviation is minimized by fixing X_i at $x_{i,j}$ while varying the remaining factors $X_{\sim i}$ in their corresponding interval (*line 18&19*). Finally, save the threshold-based response, $\tilde{Y}^k_{i,j}$, for all the points n_i of X_i and thresholds $t_k, k = 1, \ldots, \kappa$ (*line 20*).

It is worth noting that the sensitivity method, whether in the extreme-based (Section 2.1.1) or the threshold-based formulation (Section 2.1.2), is independent of the probability distribution of input factors. This is because the sensitivity is evaluated based on the discretization of the input factors, as emphasized in [20]. However, in the proposed simulation scheme, the probability distribution is required as input because the PCE metamodels are developed based on this information.

In terms of computational efficiency, the original A-A1T method requires solving optimization problems in the order of $\mathcal{O}\{n \times d\}$, where n is the average number of discrete points in the non-fixed factor and d is the total number of input factors. When multiple design criteria are present (denote the number of criteria by N_c), the number of optimizations increases to $\mathcal{O}\{n \times d \times N_c\}$. Let N_a be the average function calls required by solving one optimization problem. Then, the total function evaluations required by the original A-A1T will be $\mathcal{O}\{n \times d \times N_c \times N_a\}$. On the other hand, the metamodel-based method requires N evaluations of the original model to construct the metamodels, in this case, PCE, which is independent of N_c since responses are generally returned from a single call.

3. Verification using a truss structure

This section verifies the PCE-based approach against the original method [20]. The goal is to assess the feasibility of using PCE to approximate the output of interest and compute sensitivity measures (Eq. (4)), considering both efficiency and accuracy. Furthermore, the performance of PCE is compared to other types of metamodels, which is provided in Appendix A.1.

The verification case concerns a truss structure shown in Fig. 3. The structure consists of 23 bars, and the upper section is subjected to six vertical loads. This particular structure has been studied in many works (e.g., [31–33]) for different purposes, where SA has been conducted using the Sobol method and not in the extreme case.

Ten random variables ($d = 10$) are considered in the example, including the applied loads (P_m, $m = 1, \ldots, 6$); Young's modulus and cross sections of the horizontal and diagonal elements (respectively

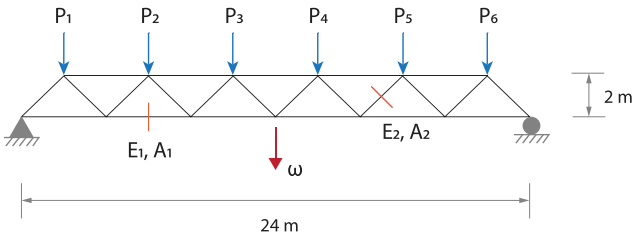

Fig. 3. Truss structure with 23 members [33].

Table 1
Truss structure: definition range and discretization of the input factors.

Variable	Data proposed by [33]			Range	Discretization
	Distribution	Mean	Stdv.	$[P_{0.01}, P_{0.99}]$	
E_1, E_2 (Pa)	Lognormal	2.10×10^{11}	2.10×10^{10}	$[1.54 \times 10^{11}, 2.84 \times 10^{11}]$	50 points (log spaced)
A_1 (m^2)	Lognormal	2.0×10^{-3}	2.0×10^{-4}	[0.0015, 0.0027]	50 points (equally spaced)
A_2 (m^2)	Lognormal	1.0×10^{-3}	1.0×10^{-4}	[0.0007, 0.0014]	50 points (equally spaced)
P_m ($m = 1...6$) (N)	Gumbel	5.0×10^4	7.5×10^3	[26 475, 64 677]	50 points (log spaced)

denoted by E_1 and A_1 for the horizontal; E_2 and A_2 for the diagonal bars). Accordingly, the input random vector is defined as $\mathbf{X} = \left[E_1, E_2, A_1, A_2, P_1, \ldots, P_6\right]^{\mathrm{T}}$. Table 1 provides the range of definition and the number of discrete points n_i for each factor X_i. For input factors with large ranges, the discrete points are distributed using a log scale. This choice is based on the suggestion in [20] where, compared to the linear and random discretization methods, using the log-spaced discretization for factors with large ranges ensures the robustness of the importance measures with a relatively small number of n_i.

The quantity of interest is the midspan deflection ω (counted positively downwards). It is obtained by evaluating the FE model, denoted as g^{truss}, using the Matlab code [34], i.e., $\omega = g^{truss}(\mathbf{X})$. PCE metamodels ($\hat{g}^{truss}$) are constructed to approximate the value of ω based on $\mathbf{X}$. The random vector $\mathbf{X}$ is transformed into a standard Gaussian distributed vector to use Hermite polynomials [30] in constructing the PCE in the form [29,31],

$$\xi_i = \Phi^{-1}\left(F_{X_i}\left(x_i\right)\right), \quad i = 1, \ldots, 10, \tag{20}$$

where Φ is the standard Gaussian CDF and $F_{X_i}\left(x_i\right)$ is the CDF of X_i. This results in the following PCE according to Eq. (9),

$$\omega \approx \hat{g}^{truss}(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \mathbf{c}_{\boldsymbol{\alpha}} \boldsymbol{\Psi}_{\boldsymbol{\alpha}}(\boldsymbol{\xi}). \tag{21}$$

Latin Hypercube Sampling (LHS) is used to generate the ED. The base scenario considers a sample size of $N = 1000$, while the effect of different sample sizes is evaluated (see Appendix A.1). The FE model is queried at the sampled points to generate the corresponding response. The input–output pairs form a dataset that is randomly divided into training, validation, and test sets with the respective percentage of 64%, 16%, and 20%, where five iterations were performed, resulting in five candidate models.

All PCE models are constructed by varying the q-norm from 0.5 to 1 and the maximum degree p from 3 to 15. The model performance is evaluated based on the test set using ε_{gen} (Eq. (19)), where the model with the smallest ε_{gen} yields 2.54×10^{-7} for predicting ω. The optimal degree p is 9 and q-norm is 0.5, associated with $\varepsilon = 1.15 \times 10^{-7}$. This model includes 181 polynomial basis elements, while the size of basis elements for $q = 0.5$ and $p = 9$ is 571, and the size of full basis elements can be 92 378 for $q = 1$ and the same p.

Two functions, g^{truss} and $\hat{g}^{truss}$, are applied to the A-1AT method (Section 2.1.1). The importance measures I_i^{Ext} to each factor X_i are reported in Table 2. It can be observed that I_i^{Ext} returned from the PCE agree well with those calculated from directly calling the FE model (the reference). The reference case also shows differences with the result in [31]. This is because [31] employs the Sobol method, which measures the impact of input factors based on the average value of the midspan deflection.

The number of function calls to the FE model is also tracked for both methods. As mentioned earlier, an optimization problem is present (Eq. (3)) for each discrete point of X_i to search for the output extremum. In this example, the Matlab algorithm 'patternsearch' [35] has been used. The average number of model runs per optimization N_a was about 165, and the average number of discrete points n for input factors was 50. Therefore, the reference method requires 82 242 ($= n \times d \times N_a$) function calls, which took 42.5 min on a desktop with an 8-core CPU and 16 GB of RAM. In contrast, the PCE-based method needs $N = 1000$ function evaluations to construct and validate the metamodel. The evaluation of $\hat{g}^{truss}$ in the SA took 5.5 min with the same computing condition.

Table 2
Truss structure: comparison of importance measures calculated from the PCE ($\hat{g}^{truss}$) and actual FE model evaluation (g^{truss}) when $N = 1000$.

Variable	Importance measure I_i^{Ext}	
	Reference	PCE-based
E_1	0.3952	0.3965
A_1	0.3703	0.3706
P_3	0.0652	0.0652
P_4	0.0652	0.0649
P_5	0.0314	0.0315
P_2	0.0314	0.0310
A_2	0.0180	0.0177
E_2	0.0154	0.0149
P_6	0.0040	0.0038
P_1	0.0040	0.0038
Function calls	82 242	1000

4. Application to train–track–bridge interaction system

4.1. Modeling of train–track–bridge dynamics

As a typical example of vehicle–structure interaction problems, the vibration of the train–track–bridge (TTB) system is a fundamental concern in railway engineering, frequently used to evaluate running safety, riding comfort, and performance of railway tracks and bridges. Here, we demonstrate the applicability of the proposed method using a coupled TTB model [36]. The model, implemented in Matlab, is formulated based on the FE method. It is capable of simulating the vertical dynamic interaction between the subsystems, namely the train, track, and bridge. Fig. 4 provides a schematic representation of the model, where an articulated train is shown traveling over a ballast bridge at a specific speed. Equations of motion are defined for each subsystem, and their features are summarized below.

The train is represented by a succession of vehicles, with each vehicle consisting of one carbody, two bogie frames, and four wheelsets, treated as rigid bogies. Each bogie is connected to two wheelsets through the primary suspension, and the main body rests on two bogies via the secondary suspension. Both vertical and rotational movements of the carbody (respectively, z_c and θ_c) and bogies (respectively, z_{bi} and $\theta_{bi}, i = 1, 2$) are considered. A rigid contact is assumed between the rail and wheelsets. This results in 6 Degrees of Freedom (DOF) for each vehicle of the train and the displacement vector of a vehicle can be denoted as $\mathbf{u}_v = \left[z_c, \theta_c, z_{b1}, \theta_{b1}, z_{b2}, \theta_{b2}\right]^{\mathrm{T}}$.

The railway track considers a ballast track that includes a combination of components, i.e., rail, pads, sleepers, ballast, and subgrade. The rail is modeled as an Euler–Bernoulli beam, with each element having 4 DOFs in terms of vertical and rotational motions. The remaining components are represented as layers of masses and viscoelastic supports, which is a conventional simplification for modeling the dynamic behavior of railway tracks. The model used here is a three-layer track model. Depending on the particular bridge configuration, a bridge structure can be accurately modeled by finite elements of solid, shell, and beam. For simplicity and without loss of generality, this demonstration considers a simply-supported bridge meshed with Euler–Bernoulli beam elements.

Each subsystem is defined by a set of equations of motion, and their coupling can be formulated in the following general matrix form,

$$\mathbf{M}_g \ddot{\mathbf{U}}_g + \mathbf{C}_g \dot{\mathbf{U}}_g + \mathbf{K}_g \mathbf{U}_g = \mathbf{F}_g \tag{22a}$$

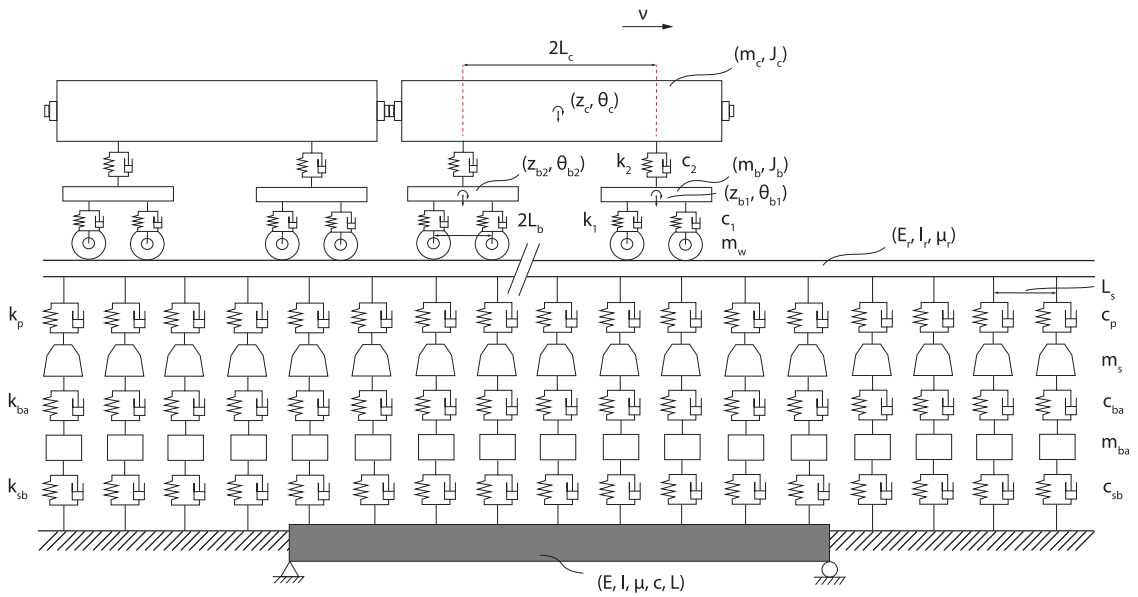

Fig. 4. Schematic representation of a train–track–bridge model.

where $\mathbf{M}_g$, $\mathbf{C}_g$, and $\mathbf{K}_g$ denote, respectively, the mass, damping, and stiffness matrices of the global system. $\mathbf{U}_g$, $\dot{\mathbf{U}}_g$, and $\ddot{\mathbf{U}}_g$ are the displacement, velocity, and acceleration vectors of the global system, respectively. $\mathbf{F}_g$ is the global force vector. Their expressions are given by

$$\mathbf{M}_g = \begin{bmatrix} \mathbf{M}_v & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_b \end{bmatrix}, \quad \mathbf{C}_g = \begin{bmatrix} \mathbf{C}_v & \mathbf{C}_{vt} & \mathbf{0} \\ \mathbf{C}_{tv} & \mathbf{C}_t & \mathbf{C}_{tb} \\ \mathbf{0} & \mathbf{C}_{bt} & \mathbf{C}_b \end{bmatrix},$$
$$\mathbf{K}_g = \begin{bmatrix} \mathbf{K}_v & \mathbf{K}_{vt} & \mathbf{0} \\ \mathbf{K}_{tv} & \mathbf{K}_t & \mathbf{K}_{tb} \\ \mathbf{0} & \mathbf{K}_{bt} & \mathbf{K}_b \end{bmatrix}, \quad \mathbf{U}_g = \begin{bmatrix} \mathbf{U}_v \\ \mathbf{U}_t \\ \mathbf{U}_b \end{bmatrix}, \quad \mathbf{F}_g = \begin{bmatrix} \mathbf{F}_v \\ \mathbf{F}_t \\ \mathbf{F}_b \end{bmatrix}. \tag{22b}$$

where the subscripts v, t, and b denote respectively the train, track, and bridge subsystems.

The coupled equations of motion, Eq. (22a), is solved by direct numerical integration with the Newmark-β method to obtain the dynamic responses of the subsystems [36], corresponding to the design criteria considered in SA.

4.2. Characteristics of input factors and design criteria

The variability space of input factors is defined to perform SA. This includes determining parameters (that are subject to small deviations and thereby being treated as deterministic) and variables (that may have relevant non-deterministic properties which can lead to a significant change in model response) in the TTB system. We consider the variability in factors related to the structural resistance and excitation source. The former is relevant to the train, track, and bridge, while the excitation source focuses on the effect of track geometry irregularities on the TTB dynamics. Table 3 provides an overview of the random variables defined in the system, along with their corresponding probability distributions.

Trains exhibit significant variability in terms of their dimensions and properties. The ranges of variation for the properties of train catalogs are location specific and have to be estimated according to the operating context in which the analysis is taking place. For demonstration purposes, the Manchester benchmark model is considered with variations in factors including the carbody mass (m_c), primary suspension stiffness (k_1), and secondary suspension stiffness (k_2) (see Table 3). These are selected according to [22], given the fact that the mass of vehicles varies depending on the occupancy rate of passengers and the suspension stiffness shows variability during its service life. The rest of the properties are treated as constants according to the benchmark model [39], as presented in Table 4.

Regarding the track structure, the dispersion of supporting components, namely, the railpad, sleeper, and ballast, is high. The railpad and sleeper are often associated with a wide range of design alternatives, while the ballast properties are very likely to change due to geometry issues and maintenance works. Therefore, they are more appropriately described by bounding limits, i.e., uniform distributions shown in Table 3.

It is worth mentioning that for the ballast, it is not straightforward to find equivalent properties to capture its behavior using the multi-layer track model (e.g., the present three-layer model). To this end, Zhai et al. [38,40] proposed a ballast model to analyze its vibration based on a hypothesis that the load transmission from the sleepers to the ballast follows approximately a cone distribution. This model defines a function mapping between the input factors relevant to the ballast material properties and the output regarding the equivalent parameter values for the ballast block used in the multi-layer track model (see Fig. 4). It has been validated against a field measurement [40] and applied to relevant works (e.g., [21]). For details of the mathematical formulation, the reader is referred to [38,40].

As shown in Table 3, we consider the variability in ballast properties, including the density (ρ_{ba}), elastic modulus (E_{ba}), depth (h_{ba}), and load distribution angle (α_{ba}), which are specified according to [21]. Then, the above ballast model is applied to determine the equivalent parameter values for the ballast, namely, the vibrating mass of ballast under a sleeper support (m_{ba}) and the support stiffness k_{ba} (see Fig. 4). The viscous damping coefficient c_{ba} is treated as a random variable referring to [21,27] to cover the variability in energy dissipation mechanisms. Besides, the ballast model includes dimensions of the sleeper support such as the sleeper spacing, L_s, to determine m_{ba} and k_{ba}. Since the dimensions are normally better defined in design specifications, they are treated as constants according to [40].

The rail is modeled as Euler–Bernoulli beam elements, whose behavior is defined by the elastic modulus (E_r), moment of inertia (I_r), and mass per unit length (μ_r), as presented in Fig. 4. Compared with the ballast, the properties of the steel rail are relatively easy to determine from nominal design values, and the subgrade properties (k_{sb} and c_{sb}) are also better defined according to the required bearing capacity. These properties are therefore considered deterministic and the corresponding values are referred to [40].

Table 3
Input random variables. Note: the fourth and fifth columns depend on the distribution type. For Gaussian or lognormal distributed random variables, mean values and coefficient of variation (CV) are used. For uniformly distributed variables, minimum and maximum values are used.

Variable	Unit	Distribution	Mean or Min.	CV or Max.	Reference
Train random variables					
Carbody mass factor	–	Lognormal	1	0.15	[22]
Primary suspension stiffness factor	–	Lognormal	1	0.1	[22]
Secondary suspension stiffness factor	–	Lognormal	1	0.1	[22]
Track random variables					
Railpad stiffness (k_p)	N/m	Uniform	1e8	1.5e9	[37]
Railpad damping (c_p)	N s/m	Uniform	1e4	7e4	[37]
Sleeper mass (m_s)	kg	Uniform	220	325	[21]
Ballast density (ρ_{ba})	kg/m^3	Uniform	1500	2100	[21]
Ballast elastic modulus (E_{ba})	N/m^2	Uniform	8e7	1.6e8	[21]
Ballast depth (h_{ba})	m	Uniform	0.3	0.6	[21]
Ballast load distribution angle (α_{ba})	°	Uniform	15	35	[21]
Ballast damping (c_{ba})	N s/m	Uniform	4e4	2.8e5	[37]
Irregularity amplitude (I_{rr})	10^{-7} rad m	Uniform	4.032	10.80	[38]
Bridge random variables					
Concrete density weight (μ)	kg/m	Gaussian	69 000	4%	[21]
Concrete elastic modulus (E)	N/m^2	Gaussian	35e9	8%	[21]
Structural damping ratio (c)	%	Gaussian	2	0.3 stdv.	[21,27]
Second moment of area (I)	m^4	Gaussian	51.3	5%	[28]

Table 4
Vehicle parameters.

Parameter	Notation	Value
Carbody mass	m_c	32000 kg
Bogie mass	m_b	2615 kg
Wheelset mass	m_w	1813 kg
Primary suspension stiffness	k_1	1220 kN/m
Secondary suspension stiffness	k_2	430 kN/m
Primary suspension damping	c_1	4 kN s/m
Secondary suspension damping	c_2	20 kN s/m
Distance between centers of car and bogie	L_c	9.5 m
Distance between centers of bogie and wheel	L_b	1.28 m
Full length of vehicle	L_v	25 m
Velocity	v	140 km/h

The properties of railway bridges are case-specific, and it is challenging to find a set of properties that can be applied to describe large catalogs of bridge structures in general. Herein, we consider a 50 m long concrete bridge referring to a specific case [41]. This also suggested the default parameter setting for the bridge in the adopted model [36]. The mean values of bridge properties are first determined according to the selected case. Then, Gaussian distributions are assigned to describe the dispersion of these properties, as reported in Table 3.

Track irregularities are deviations from the ideal track geometry that can significantly affect the dynamic behavior of trains and structures. It is therefore essential to account for their effect in the dynamic analysis of such systems. Random track irregularities are often characterized by power spectral density (PSD) functions, which describe the severity of track irregularities as a function of the spatial frequency $\Omega = 1/\lambda$, where λ represents the wavelength in meters. Various PSD functions have been developed by different railway authorities. Given its extensive use in the field of railway engineering, the German track spectrum for the vertical track profile is adopted with the PSD function defined as (e.g., [38]),

$$S_v(\Omega) = \frac{A_v \Omega_c^2}{\left(\Omega^2 + \Omega_r^2\right)\left(\Omega^2 + \Omega_c^2\right)}, \tag{23}$$

where the unit of $S_v(\Omega)$ is m^2/rad/m; Ω is the spatial frequency; Ω_r and Ω_c are cut-off frequencies, set to 0.0206 and 0.8246 rad/m, respectively. The magnitude of A_v (unit: m^2 rad/m) quantifies the track quality and varies between 4.032×10^{-7} and 1.08×10^{-6} to represents respectively the limit for low disturbance and high disturbance of track irregularities. Here, we assume that the track quality is represented by a uniformly distributed random variable within the range. The function considers wavelengths ranging from 3 to 150 m (referring to EN 13848-5 [42] ranges D1-D3). To perform time-domain analyses, track irregularity profiles are generated from Eq. (23) using inverse fast Fourier transforms.

The performance of the TTB system is controlled by four design criteria, namely Sperling index (W_z), vertical rail displacement (U_r), vertical sleeper acceleration (A_s), and vertical deck deflection (U_d), where the dynamic response of each subsystem is accounted for to capture the overall level of vibration in the system. Of interest for the present SA are the maximum values of the responses, which are relevant to the occurrence of 'system failure', and therefore reflect the most unfavorable condition of the system. While the remaining response quantities can be directly obtained from the dynamic simulation, W_z is a synthetic index calculated based on the carbody acceleration. This index measures the level of riding comfort and follows the computation procedure proposed in [43].

5. Results and discussion

5.1. PCE representations of the model output

As the first part of the methodology, PCE metamodels are constructed to approximate the model output. This is implemented using the uncertainty quantification toolbox UQLab [44]. Following the workflow in Section 3, LHS is employed to generate a ED $\left\{\boldsymbol{x}_{(\varrho)}, \varrho = 1, \ldots, N\right\}$ with a fixed size of $N = 3000$. The choice of a relatively large sample size aims to ensure the accuracy of the metamodels, as SA is performed around extreme values that may be omitted by global sampling approaches conducted in a single step. When the ED is fixed, the LHS technique has been preferred in reliability applications due to its global representation of the input space [45]. This aligns with the goal of metamodeling here, which is to develop a globally accurate metamodel to approximate the original function and enable the measurement of each input factor's effect on the output.

The TTB model is queried at the sampled points to obtain the response $\left\{\boldsymbol{y}_{(\varrho)}, \varrho = 1, \ldots, 3000\right\}$. For a specific sample ϱ, the response vector is defined by $\boldsymbol{y} = \left\{W_z, U_r, A_s, U_d\right\}^{\mathrm{T}}$. The sample data is split randomly into training, validation, and test sets with the respective percentage of 64%, 16%, and 20%, where for each design criterion, ten iterations were performed and the resulting candidate metamodels are evaluated based on the test set using ε_{gen} (Eq. (19)). Fig. 5 compares the PCE predictions with the actual model evaluations at the test set,

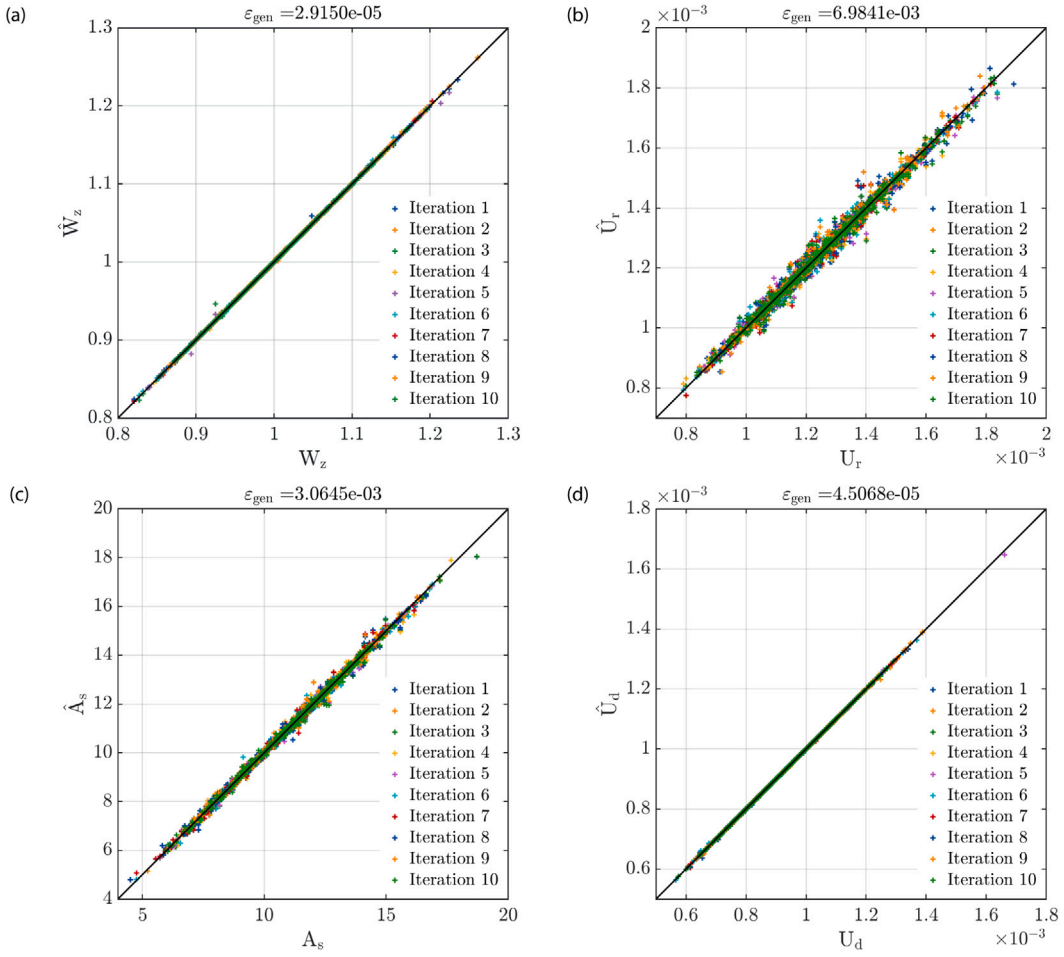

Fig. 5. Comparison of PCE with the actual model evaluation at the test set: (a) Sperling index W_z; (b) vertical rail displacement U_r (unit: m); (c) vertical sleeper acceleration A_s (unit: m/s^2); and (d) vertical deck deflection U_d (unit: m).

and ϵ_{gen} obtained from the optimal PC approximation is indicated for each criterion.

The PCE metamodels are developed by varying the q-norm from 0.5 to 1 and the maximum degree p from 3 to 15. The one yielding the smallest ϵ_{gen} to predict W_z comprises 371 polynomial basis elements, with $p = 15$ and $q = 0.5$. By comparison, for the same p, the size of full basis elements is 9081 when $q = 0.5$ and around 1.55×10^8 when $q = 1$. The index of sparsity, defined in [29] as the ratio of the number of elements in the sparse representation to the full size of elements for the same p and q, is $371/9081 \approx 0.041$. The small ratio indicates the computational gain that the sparse PCE can bring to the sensitivity analysis, compared with the full PC representation.

The optimal sparse PCE for predicting U_r consists of 136 basis elements, with $p = 15$ and $q = 0.5$. It has an index of sparsity of $136/9081 \approx 0.015$. As for the response A_s and U_d, the optimal PCE includes 223 and 442 elements, respectively, with the same p and q as in the model for approximating U_r. The corresponding index of sparsity is $223/9081 \approx 0.025$ for A_s and $442/9081 \approx 0.049$ for U_d.

Additionally, the accuracy and efficiency of PCE are compared with Kriging and PC-Kriging approaches, which are provided in Appendix A.2.

Note that the performance of metamodels depends on various factors, such as the shape and complexity of the function being approximated, the ED size, and the sampling strategies employed. In this study, we employ the PCE approach in the proposed sensitivity method, which allows us to efficiently capture the global stochastic behavior of the system [45,46] and measure the effect of input factors on the output. In the truss example, PCE performs particularly well with limited ED sizes (see Appendix A.1), which is advantageous when dealing with expensive models. However, it should be noted that PCE is most effective for functions that can be well-approximated by global smooth polynomials. In two engineering problems, PCE outperformed Kriging (see Appendices A.1 and A.2). Kriging is generally suitable for managing local variability of the output [46]. This indicates that the functions involved show global smoothness rather than (highly) local nonlinearity, thereby contributing to the better performance of PCE. Furthermore, we assess the sensitivity of a 2-dimensional analytical function with globally nonlinear behavior to visually compare these metamodels. The detailed comparison is provided in Appendix A.3.

As mentioned earlier, this study uses LHS with a fixed ED to provide a global description of the input space. Instead of sampling the ED at once, it is possible to apply adaptive or sequential sampling techniques to refine sampling in specific regions of interest. This approach allows for a more efficient allocation of computational resources, balancing the exploration and exploitation of the input space while constructing the metamodel. Note that in this case, the sampling is more focused on achieving accuracy in the proximity of specific regions, depending on the degree of exploitation, rather than aiming for a globally accurate metamodel throughout the entire domain, which is more exploration-based. The choice of sampling methods can depend on function complexities. For functions with highly local nonlinearity, where the extremes are usually local phenomena, a balance of exploration and exploitation of the input space would be necessary. Considering the functions in the current engineering problems demonstrate global smoothness, LHS, which captures the global description of the ED, is considered suitable for this study.

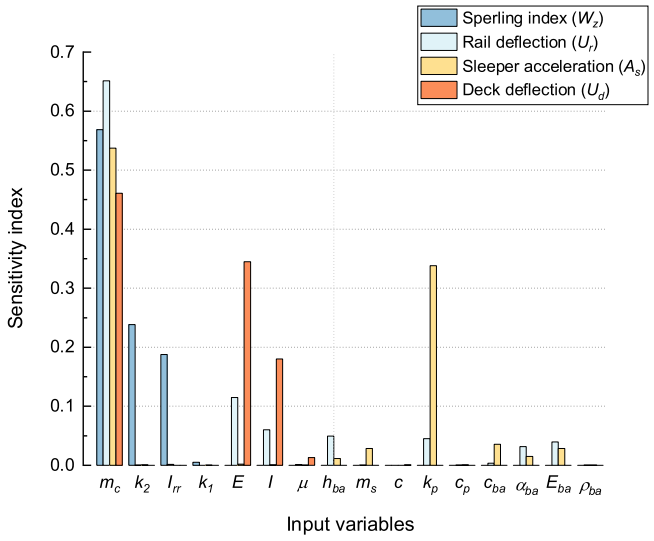

Fig. 6. Comparison of sensitivity index of each factor in terms of W_z, U_r, A_s, and U_d.

5.2. Ranking of input factors

We employ the PCE models developed in Section 5.1 to assess the impact of input factors on the variability of extreme responses. This evaluation is performed on the original sensitivity method (Section 2.1.1). The input factors are discretized using an equal number of points. Similar to the discretization strategy discussed in Section 3, a log-spaced discretization method is applied for factors with ranges exceeding 10^3. After evaluating various discretization options, $n_i = 100$ is chosen for each factor to balance the computational cost and the accuracy of sensitivity indices, which is applied to all design criteria to ensure consistency in the comparison.

As shown in Fig. 6, importance measures of the input factors are compared in terms of response W_z, U_r, A_s, and U_d. The indices are first presented in descending order for W_z. Then, different rankings regarding the relevance of the factors are observed for the remaining criteria. The variability in carbody mass (m_c) shows high relevance to all the criteria considered, indicating that the loading magnitude plays a significant role in the vibration of the TTB system.

Specifically, the factors relevant to the vehicle (m_c and k_2) and track geometry quality (I_{rr}) contribute to most of the variability (about 99.47%) in the maximal W_z. Employing the condition $I_i^{Ext} < 0.01$ to sort out unimportant factors allows one to consider 13 out of 16 input factors as unimportant when the design focus lies on the critical riding condition that passengers may experience. This implies that those factors could be given a deterministic value to reduce the model complexity without essentially affecting the extreme response W_z. On the other hand, the above three factors are deserving of further analysis or measurement to assign the appropriate values in the model for a more accurate representation of the riding quality condition.

At the wheel–rail interface, more input factors are involved that have non-negligible effects on the variability of maximal U_r. These are m_c, E, I, h_{ba}, k_p, E_{ba}, and α_{ba}, with the first two being dominant and accounting for about 76.60% of variability in maximal U_r. The accuracy of these factors is of high relevance to the quality of the extreme response U_r, which should be carefully defined if the rail deflection is of concern in the design process.

For underlayers of the track structure, the variability of the maximal vertical sleeper acceleration (A_s) is significantly influenced by m_c and k_p (about 87.55%). Apart from m_c, the effect of the railpad stiffness k_p can be explained by the fact that the railpads are elastic components introduced to the track structure. Stiff pads contribute to a reduction in noise and vibration from wheel–rail contact, while soft pads allow for a lower effect of loads transmitted to underlayers and therefore mitigate damage and vibrations to the sleepers and ballast [47]. This indicates if the designer focuses on the critical condition of the supporting layers (for example, in the case of transition zones to the railway bridges where issues with settlement often exist), the values of m_c and k_p should be chosen with careful consideration in the modeling process, especially for k_p since it often comes with a wide range of values and is also influenced by the frequency and environmental factors such as temperature, preload, and aging [25]. If possible, a more advanced model that captures the effect of those factors on the railpad properties should be employed for a more accurate representation of the railpad behavior.

It can also be observed from Fig. 6 that m_s, as the only property relevant to the sleeper in the modeling, plays a minor role (about 2.83%) in the variability of maximum sleeper accelerations. Despite the common use of mass elements to represent sleepers in railway structures, it may be inferred that, in the adopted TTB model, the sleeper modeling might be over-simplified if the condition of track supporting layers is of particular concern. Instead, beams or solids are alternatives to improve the modeling of sleepers. This highlights the value of using sensitivity analysis based on extremes for the purpose of validating models.

Track ballast provides a supporting layer to the sleepers, and it becomes reasonable that the ballast properties (ρ_{ba}, E_{ba}, h_{ba}, α_{ba}, and c_{ba}) play a relatively important role in the sleeper response A_s. However, in Fig. 6, the sensitivity indices of those factors with regard to A_s are more spread out with the individual importance measure no greater than 5%. This distributed effect may be attributed to the ballast model [38,40] adopted in the dynamic analysis (see Section 4.2), where the five ballast properties were aggregated into two equivalent parameters (m_{ba} and k_{ba}) that were actually used in the dynamic simulation. It becomes clear when measuring the total effect of the ballast properties, which account for 9.10% of the variability in maximal A_s.

The maximum deck deflection shows the highest sensitivity to the carbody mass (m_c) and bridge properties (E, I, μ and c), which jointly accounts for almost all of the response variability (about 99.99%). Specifically, factors E and I contribute about 52.51% of the variability, which is physically reasonable, given the importance of EI in beam deflection. Therefore, it is important to carefully define these factors before performing dynamic analysis if the goal is to evaluate the bridge response.

Note that train speed is not considered a variable in the current SA. It is fixed in its given value since the present work aims at evaluating the impact of input factors with high variability on the extreme dynamic response, while the operational speed of the trains

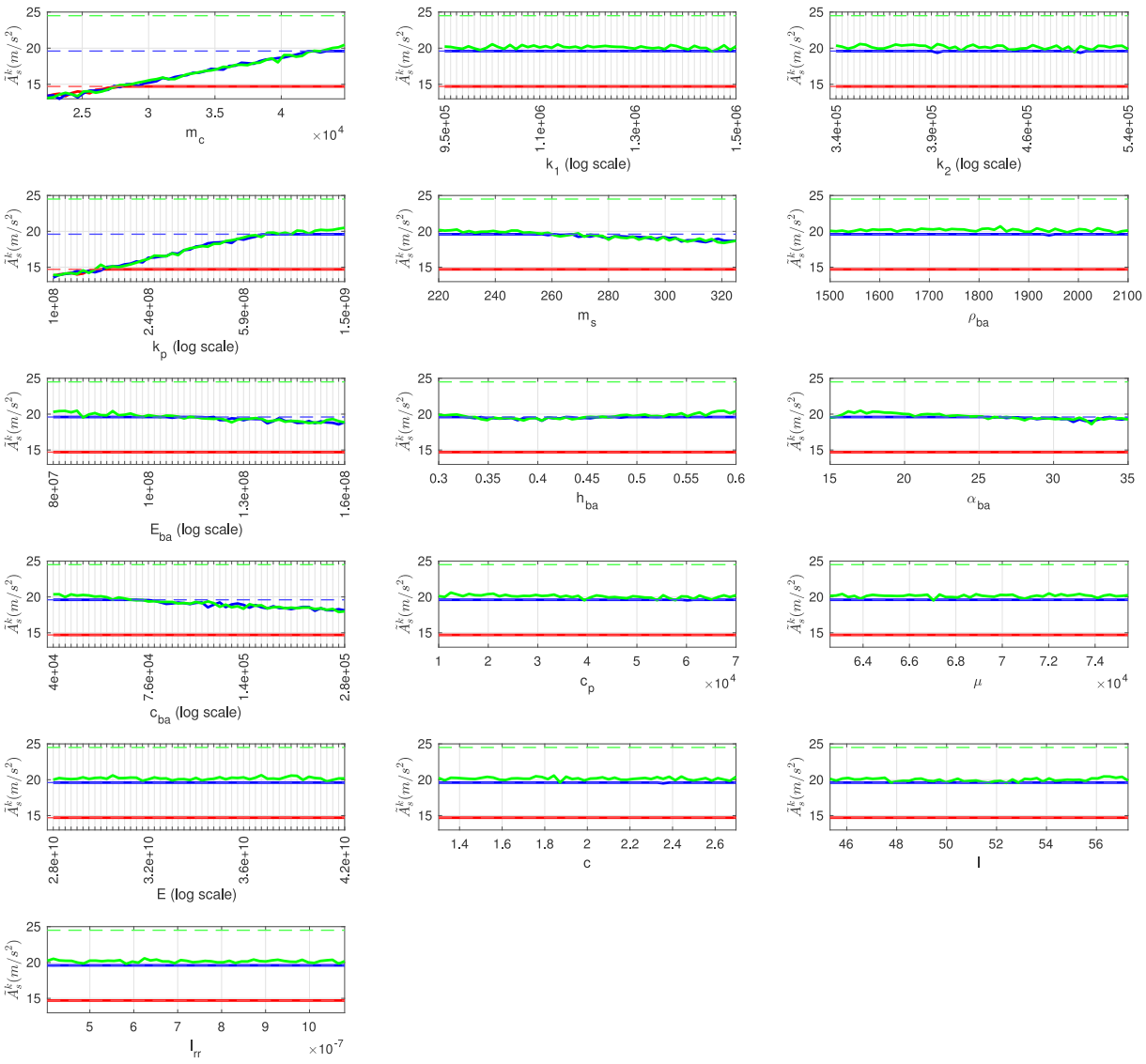

Fig. 7. Threshold-based response surface of A_s for each fixed factor. Scenario 1 (red lines) with threshold A_s = 1.5 g m/s^2 (dotted) and threshold-based response $\tilde{A}_s^1$ (solid); Scenario 2 (blue lines) with threshold A_s = 2 g m/s^2 (dotted) and threshold-based response $\tilde{A}_s^2$ (solid); and Scenario 3 (green lines) with threshold A_s = 2.5 g m/s^2 (dotted) and threshold-based response $\tilde{A}_s^3$ (solid). In each scenario, the region above or below the corresponding threshold indicates the failure or non-failure domain, respectively. This also applies to Fig. 8.

is normally predefined with less uncertainty. The Kelvin foundation adopted in the TTB model implies a limitation of the train speed to subcritical velocities. For speeds that are above the critical velocity, dynamic amplification effects of the response can be observed. This can cause significant changes in the extreme response surface and therefore affects the result of sensitivity analysis.

5.3. Impact of design thresholds

The threshold-based sensitivity analysis (Section 2.3) is applied in the TTB case. Fig. 7 shows the threshold-based sleeper acceleration ($\tilde{A}_s^k$) for each fixed factor, considering all input factors for demonstration purposes. Fig. 8 presents the threshold-based response for the remaining criteria, specifically focusing on the most significant input factors to maintain brevity.

In Fig. 7, the thresholds are defined referring to [48], where sleeper accelerations for the plain line are observed in the range of ±2 g (g = 9.8 m/s^2, gravitational acceleration). Accordingly, the thresholds of 1.5 g, 2 g, and 2.5 g m/s^2 are determined for Scenarios 1, 2, and 3, respectively.

When A_s = 1.5 g m/s^2, it can be seen that only m_c and k_p cause changes in the response curves (the solid red lines in the corresponding subplots), implying their critical contributions to the exceedance of the current limit state. Guaranteeing either m_c < 27 500 kg or k_p < 140 MN/m, the unfavorable system condition can be avoided. Note that a combination of other insignificant factors may also influence the system reliability state. The current SA aims at quantifying the univariate effect of input factors on the model response near the limit states, while the joint effect will require further analysis.

In Scenario 2 (A_s = 2 g m/s^2), there are more input factors that influence whether the reliable state of the system can be attained, i.e., m_c, k_p, m_s, E_{ba}, h_{ba}, α_{ba}, and c_{ba}. Compared with Scenario 1, m_c or k_p can take the value in a larger range without taking the risk of 'failure', since the design threshold is less restrictive. However, when A_s = 2.5 g m/s^2, the maximal response is always below the threshold. In this case, none of the input factors are considered important, as none of the combinations of the factors taking values in their definition range will cause 'failure'. From this, it is highlighted that the chosen threshold has a substantial impact on the contribution of input factors (whether critical or not) to the variability of the extreme model response.

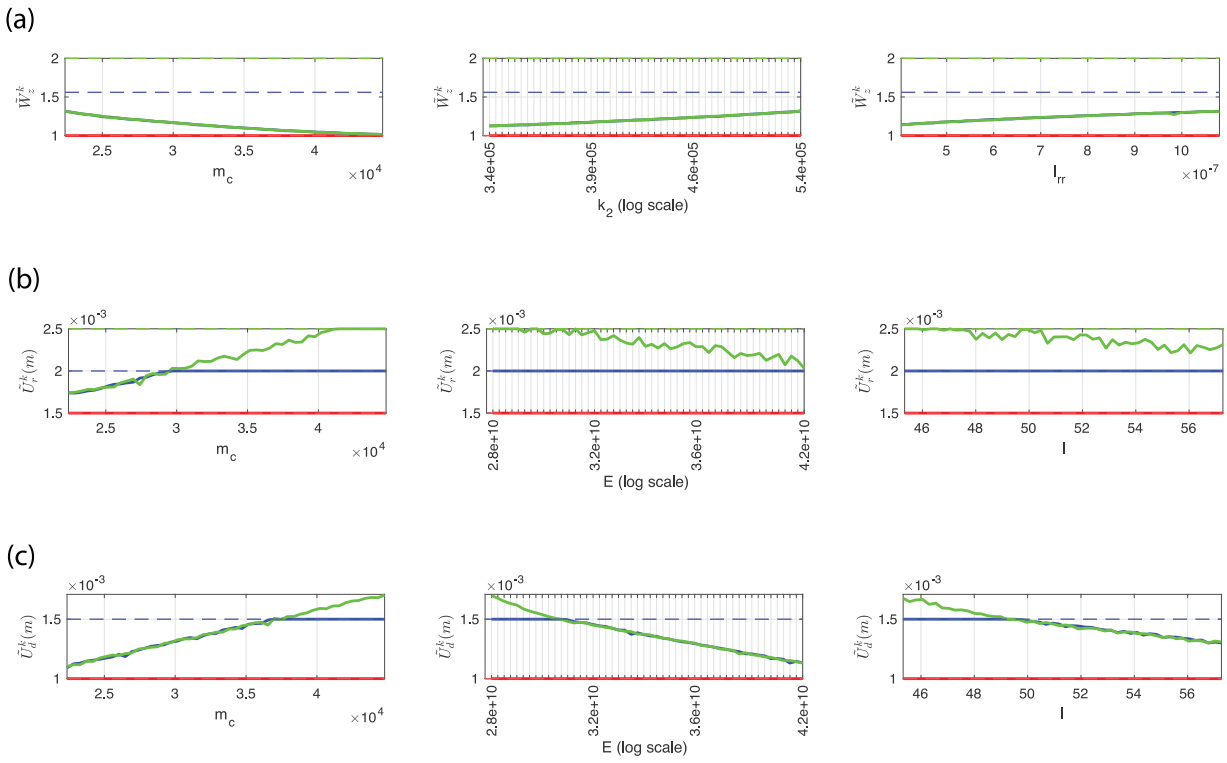

Fig. 8. Threshold-based response surface of (a) W_z, (b) U_r, and (c) U_d with regard to the most significant factors.

Fig. 8 shows the threshold-based response of the remaining criteria, focusing on the factors that contribute to substantial changes in these curves. In Fig. 8a, the limits for W_z are defined according to rating scales of the Sperling index and the response variability space from the ED (see Fig. 5a). $W_z < 1$ classifies the riding condition as 'Just noticeable'; $1 < W_z < 2$ means 'Clearly noticeable'. $W_z = 1.56$ defined for Scenario 2 refers to the nominal value of the limit in [49]. In Fig. 8b, the limits for U_r are defined referring to [49], where a range between 0.00102 and 0.0025 m is considered for the rail vertical deflection. Accordingly, the thresholds of $0.0015, 0.0020$, and 0.0025 m are determined for Scenarios 1, 2, and 3, respectively.

In Fig. 8c, the thresholds for Scenarios 1 and 2 are determined based on the response variability space obtained from the ED (see Fig. 5d). The threshold defined for Scenario 3 refers to EN 1990 [50], which specifies the maximum permissible vertical deflection for railway bridges based on factors such as the number of spans, span length, train speed, and bridge configuration. For the current bridge case, $U_d = 0.058$ m is derived and defined as the decision threshold for Scenario 3.

Firstly, it is observed from Fig. 8 that in Scenario 1 (the red lines), the response surfaces consistently align with the corresponding thresholds for all the criteria. This indicates that the design thresholds of these criteria can always be attained by a critical combination of the variables, where none of the variables play an impeding role.

In Scenarios 2 and 3, deviations occur in the considered criteria. As shown in Fig. 8a, the response surfaces in Scenario 2 (the blue solid lines) coincide with those simulated in Scenario 3 (the green solid lines), and the maximum response values consistently fall below the threshold for Scenario 2, $W_z = 1.56$. This implies that varying any factor within its range will not alter the system reliability state, as there are sufficient margins between the actual extreme model response and the given threshold.

However, Fig. 8b demonstrates the significant influence of m_c in determining the system reliability states ($U_r = 0.0020$ m and $U_r = 0.0025$ m). This observation aligns with the sensitivity index shown in Fig. 6, where m_c alone accounts for about 65.13% variability of the maximum U_r. Besides, in Scenario 3, despite slight fluctuations caused by numerical simulations, it is observed that the response curves of factors E and I intersect with the threshold $U_r = 0.0025$ m, indicating their relevance in reaching the corresponding reliability state.

In terms of the criterion U_d, the dominant factors indicated in Fig. 8c align with those shown in Fig. 8b. Specifically, in Scenario 2, varying the values in the range $m_c < 37\,000$ kg, $E > 3.12 \times 10^{10}$ N/m^2, or $I > 50$ m^4 can guarantee the maximum U_d below this threshold. In Scenario 3, there are always sufficient margins between the maximum U_d and given threshold $U_d = 0.058$ m, suggesting that all the factors can be considered as non-critical.

6. Conclusions

Different types of methods have been developed for sensitivity analysis (SA), where the Sobol method is possibly the most prevalent form of global sensitivity method in engineering applications. However, this method focuses on the average behavior of the systems, which may not be sufficient for safety-critical structures where often limit states and the corresponding extreme values of the response are of particular concern.

This paper focuses on the extreme response that a structure can potentially experience. The 'extreme' can be interpreted as either the maximum (or minimum) [20] or the response near a limit state (i.e., the threshold-based response). When a threshold is large enough, the threshold-based response implies the maximum value of the response, and vice versa. Therefore, the threshold-based response can be considered a generalized extreme response. For this, an efficient method is proposed to evaluate the sensitivity of the extreme model response to input factors, which extends the work [20] by incorporating the limit state design considerations in the formulation of extreme problems, i.e., the so-called threshold-based sensitivity method.

Further, since the sensitivity method is optimization-based, which requires iteratively maximizing (or minimizing) the model to search for the extreme model response, the computational cost involved in SA may become unaffordable in dealing with models that are computationally intensive. To circumvent this issue, the random model output

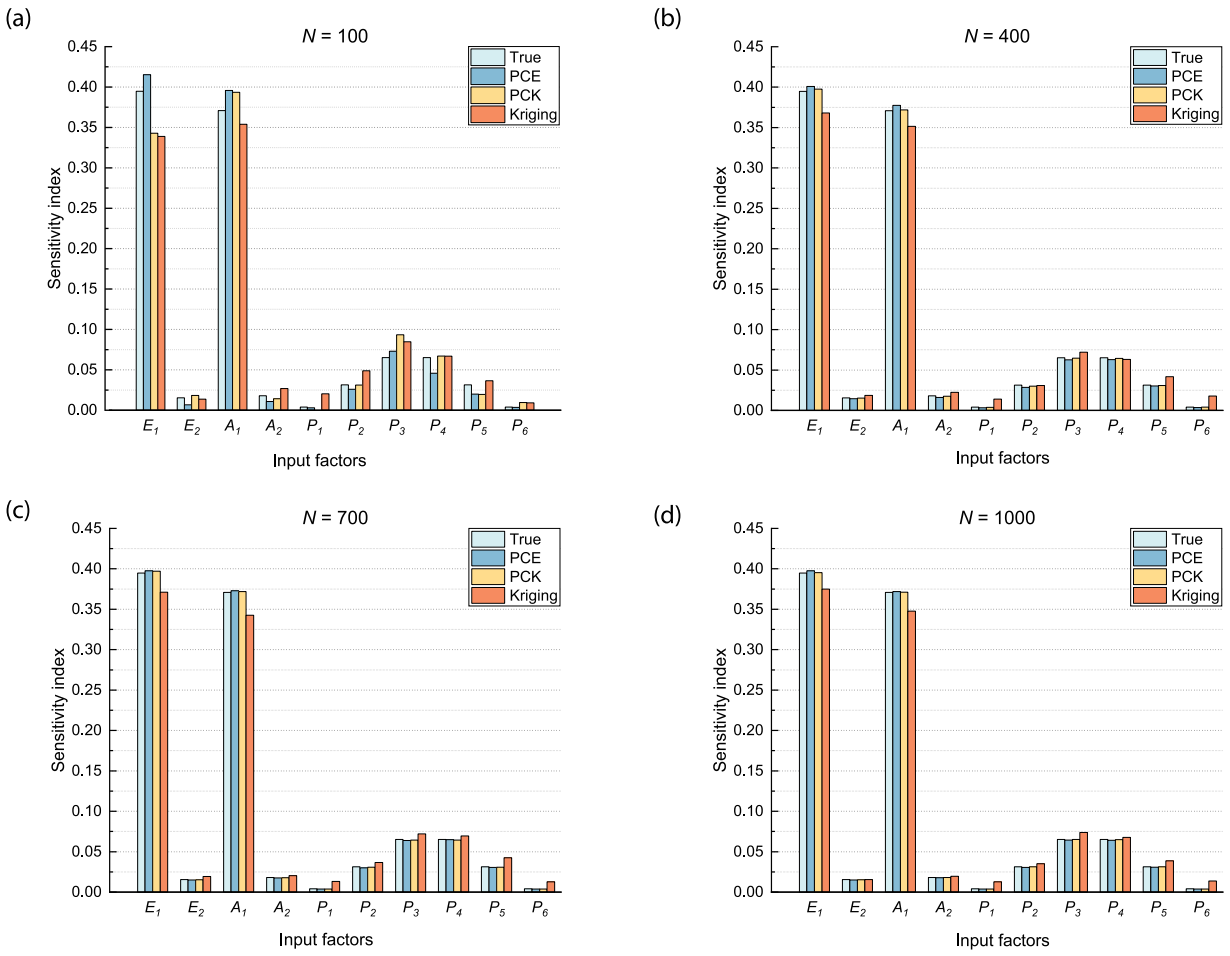

Fig. A.9. Importance measure of each input factor using true function evaluations (True), PCE, PC-Kriging (PCK), and Kriging metamodel.

is represented by polynomial chaos expansions (PCE), in which the original expensive model is replaced by an approximation that is faster to evaluate.

The case study screens the application on railway tracks, with a focus on the train–track–bridge dynamic interaction. The sensitivity of the maximum dynamic response is assessed while accounting for the existence of track irregularities and uncertainty in the factors of the train, track, and bridge. Four design criteria are defined to capture the overall level of vibration of the system. The result suggests the high relevance of loading magnitude to all the criteria considered. However, we have seen that the sensitivity between the input factors and each criterion can be very different. This indicates that SA should be performed with a specific purpose, rather than more generally on the model.

The threshold-based SA using PCE is applied in the train–track–bridge system to analyze the model sensitivity near the limit states, where the relevance of the input factors on reaching a given limit state is visualized. It helps analysts identify critical regions that lead to undesirable outcomes in the input design space under a given scenario. From the case study, it is highlighted that the chosen threshold has a substantial impact on the contribution of input factors to response variability. For this reason, the thresholds that determine the system limit states should be carefully defined for a proper SA.

The current SA method is generic and can be extended to accommodate different metamodel approaches. By using PCE, our focus is limited to functions that can be well-approximated by global smooth polynomials. Future work would be required to explore alternative approaches to handle functions with highly local non-smoothness or nonlinearity. For example, more involved versions of PCE, such as Stochastic Spectral Embedding, can be explored to fit the PCE in subdomains and capture local nonlinearities. Another extension of the method involves conducting a benchmark study to investigate the connection among function properties (dimensionality and complexity), metamodel types (and solvers), experimental design size, and sampling schemes.

CRediT authorship contribution statement

Yue Shang: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft. **Maria Nogal:** Conceptualization, Methodology, Software, Supervision, Writing – review & editing. **Rui Teixeira:** Conceptualization, Methodology, Writing – review & editing. **A.R. (Rogier) M. Wolfert:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

No potential conflict of interest was reported by the authors. It is noted that the manuscript is not being considered for publication, in whole or in part, in any other publication. If accepted, the authors agree that the article will not be published elsewhere in the same form, in English or in any other language, without the written consent of the publisher.

Data availability

Data will be made available on request.

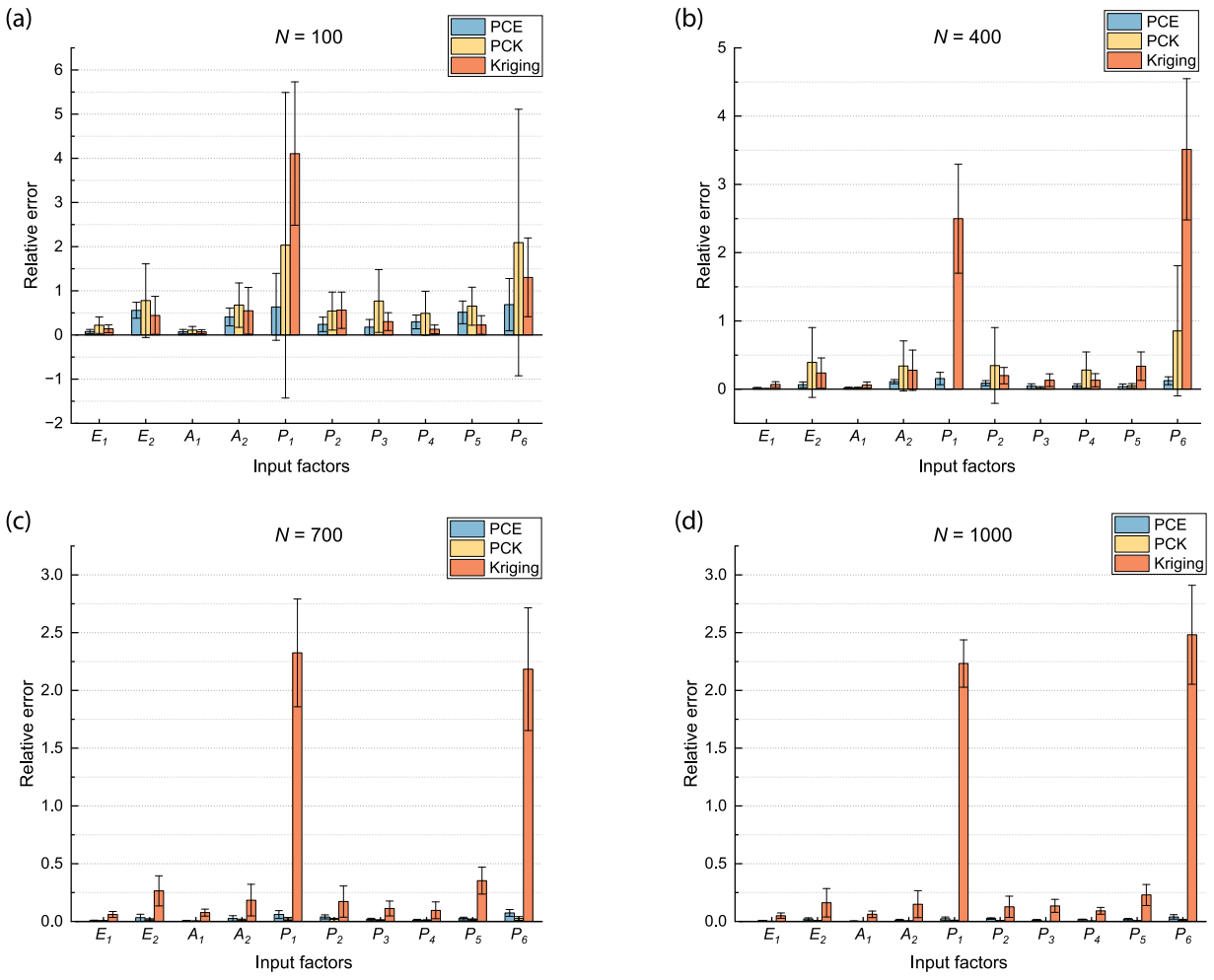

Fig. A.10. Accuracy of the importance measure of each input factor using PCE, PC-Kriging (PCK), and Kriging metamodel. The bar heights show the mean and error bars represent ±1 standard deviation over 20 trials.

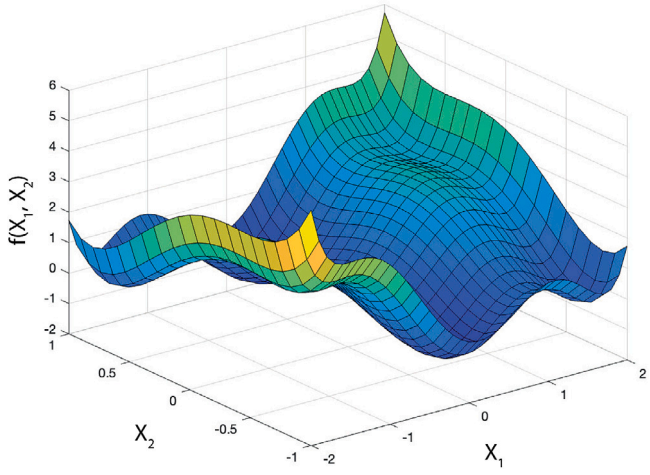

Fig. A.11. Visualization of the six-hump camel function.

Appendix A. Comparison of PCE with other metamodels

A.1. Results for the truss structure

The importance measures in the truss case are evaluated using different sample sizes (N) and metamodels (PCE, Kriging, and PC-Kriging), as shown in Figs. A.9 and A.10. In the alternative metamodels, an ordinary Kriging model with a Matérn 5/2 correlation function is applied. PC-Kriging considers a sequential formulation [46] where the set of polynomials and Kriging are determined sequentially. The LAR algorithm (Section 2.2) is applied to select the optimal set of polynomials, varying the maximum degree p from 3 to 15. This set of polynomials is then used as the trend of a Kriging model with a Gaussian correlation function, which is further calibrated using maximum likelihood. For

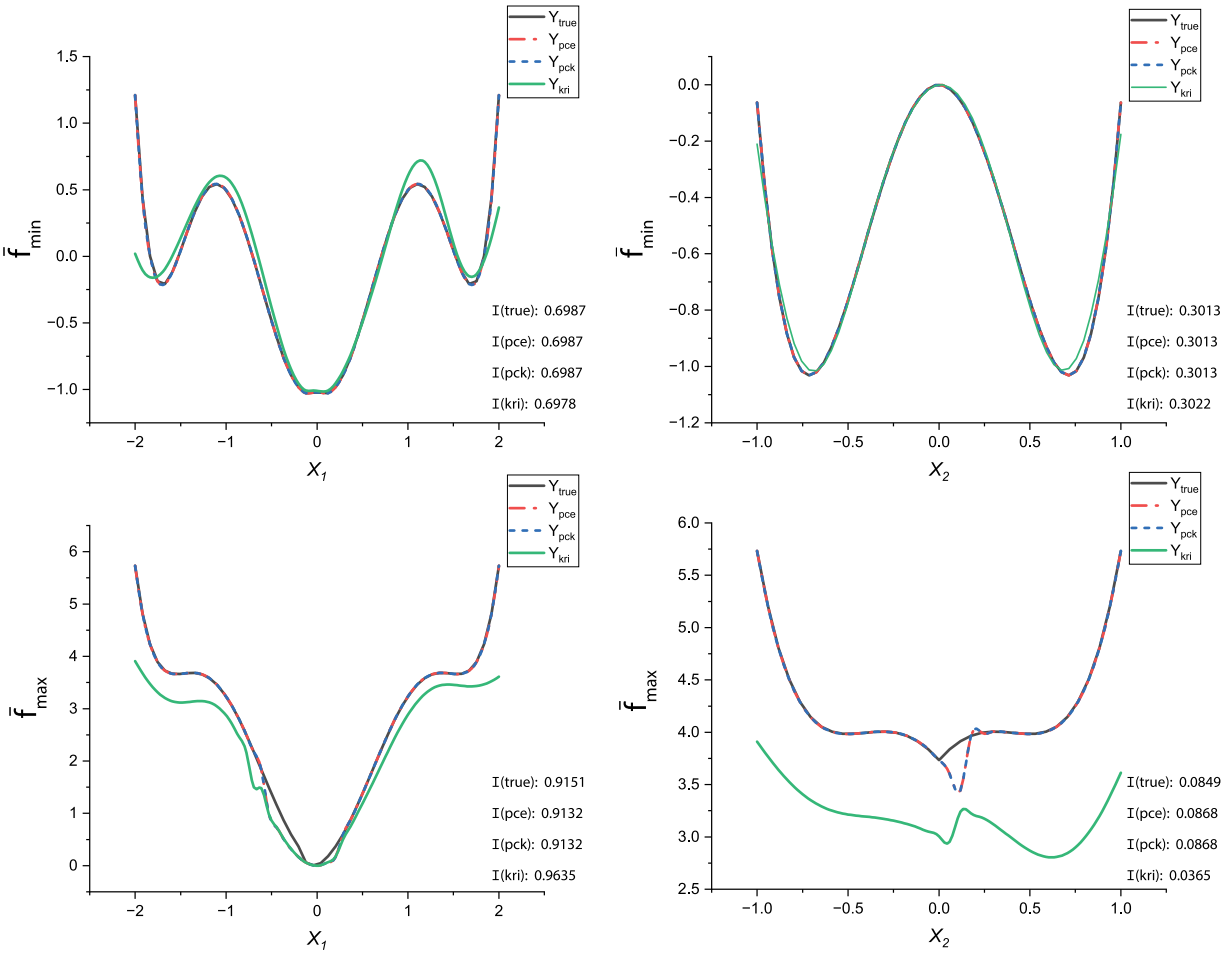

Fig. A.12. Output extreme (minimum $\bar{f}_{\min}$ and maximum $\bar{f}_{\max}$) surfaces of the six-hump camel function for each factor.

sensitivity evaluation, the model training is repeated five times for each approach, and the one with the lowest ϵ_{gen} is chosen.

The comparison of the sample size and metamodels is presented in Fig. A.9. It can be observed that Kriging shows the largest deviation from the true values across all sample cases. PCE and PC-Kriging demonstrate comparable performance, with PCE showing better results especially when the sample size is small. Note that for $N = 100$, both Kriging and PC-Kriging provide incorrect rankings for the first and second important factors (E_1 and A_1, respectively). However, as the sample size increases, the estimated measures gradually converge towards the true values and align with the true rankings.

The analysis is replicated to account for the uncertainties in experimental designs. Fig. A.10 shows the relative errors of the sensitivity measures over 20 trials. For small sample sizes ($N = 100$ or 400), PCE demonstrates the best performance in terms of both mean and variation of the errors, while PC-Kriging shows the largest variation of the errors when $N = 100$, especially for factors P_1 and P_6 that have a negligible effect on the model output. However, as the sample size N increases ($N = 700$ or 1000), PC-Kriging shows a slight improvement over PCE, as indicated by the lower mean and variation of the errors. This difference can be attributed to the fact that PC-Kriging is more prone to overfitting when dealing with small ED sizes [46]. On the other hand, Kriging shows considerable bias in the mean of the errors across all the sample sizes, and no clear improvement in its performance is observed as the sample size increases.

A.2. Results for the TTB system

The performance of PCE is compared with Kriging and PC-Kriging approaches in the TTB system. The results are presented in Table A.5. The training of Kriging and PC-Kriging follows the same methodology as described in Appendix A.1. Ten independent runs are carried out for each metamodel, and the one with the lowest relative generalization error ϵ_{gen} is chosen for further comparison. Results from Table A.5 indicate that PCE outperforms the others in terms of ϵ_{LOO} and function evaluation time. PCK performs slightly better than PCE (with the same order of accuracy) for the relative generalization error ϵ_{gen} of two response indicators (W_z and U_d). Furthermore, PCE demonstrates significant computational efficiency compared to the other approaches, which is advantageous for conducting sensitivity analyses of large-scale engineering systems.

A.3. Additional verification using an analytical function

In this section, we illustrate the concept of output extreme surfaces of the $d-1$ dimensional space (Section 2.1.1) and further verify the use of PCE for the extreme-based sensitivity method through an analytical function. The six-hump camel function is chosen because it has a global nonlinear behavior, as depicted in Fig. A.11, and involves only two input factors, allowing for visualization. This function is defined as

$$f(X_1, X_2) = (4 - 2.1X_1^2 + \frac{X_1^4}{3})X_1^2 + X_1X_2 + (-4 + 4X_2^2)X_2^2, \tag{A.1}$$

with X_1 defined in the interval $[-2, 2]$ and X_2 defined in $[-1, 1]$.

Fig. A.12 depicts the output extreme surfaces ($\bar{f}_{\min}$ and $\bar{f}_{\max}$) of the six-hump camel function when fixing each factor at a time. The estimated importance measures are also displayed for each factor. A sample size $N = 100$ is considered given the limited dimension of the problem. The extreme surfaces evaluated from the true function

Table A.5
Train–track–bridge system: comparing the performance of metamodels to approximate Sperling index W_z, vertical rail displacement U_r (unit: m), vertical sleeper acceleration A_s (unit: m/s^2), and vertical deck deflection U_d (unit: m). Minimum values are highlighted for each index (ϵ_{LOO}, ϵ_{gen} and average function evaluation time).

Response	Metamodel	Minimum ϵ_{LOO}	Minimum ϵ_{gen}	Average function evaluation time (s)
W_z	PCE	$\mathbf{7.13 \times 10^{-6}}$	2.92×10^{-5}	**0.0044**
	Kriging	5.00×10^{-4}	5.09×10^{-4}	0.1904
	PC-Kriging	9.17×10^{-6}	$\mathbf{2.73 \times 10^{-5}}$	0.2168
U_r	PCE	$\mathbf{5.93 \times 10^{-3}}$	$\mathbf{6.98 \times 10^{-3}}$	**0.0039**
	Kriging	1.77×10^{-2}	1.42×10^{-2}	0.2037
	PC-Kriging	7.18×10^{-3}	9.17×10^{-3}	0.2141
A_s	PCE	$\mathbf{1.74 \times 10^{-3}}$	$\mathbf{3.06 \times 10^{-3}}$	**0.0044**
	Kriging	5.62×10^{-3}	5.92×10^{-3}	0.1804
	PC-Kriging	2.18×10^{-3}	3.38×10^{-3}	0.2003
U_d	PCE	$\mathbf{1.44 \times 10^{-5}}$	4.51×10^{-5}	**0.0074**
	Kriging	1.55×10^{-3}	4.22×10^{-4}	0.1967
	PC-Kriging	1.59×10^{-5}	$\mathbf{3.77 \times 10^{-5}}$	0.2513

are shown to compare the accuracy of PCE, Kriging, and PC-Kriging. It can be observed that both PCE and PC-Kriging effectively capture the global shape of the function and accurately represent the minimal surfaces ($\bar{f}_{\min}$). However, they fail to identify the minimum point in the maximum surface ($\bar{f}_{\max}$) for X_2. Despite this, their estimated importance measures remain close to the true measures. In contrast, Kriging has limitations in approximating the global behavior of the function, resulting in a large deviation between the estimated and true importance measures, especially for the maximum surfaces ($\bar{f}_{\max}$). Note that the extreme surfaces are obtained by discretizing each factor at a time, with $n_i = 50$. Therefore, the surfaces may not appear smooth.

References

[1] Soboĺ IM. Sensitivity estimates for nonlinear mathematical models. Math Model Comput Exp 1993;1:407.
[2] Saltelli A, Tarantola S, Chan K-S. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 1999;41(1):39–56.
[3] Saltelli A, Chan K. Sensitivity analysis. Wiley series in probability and mathematical statistics, Chichester etc.: Wiley; 2000.
[4] Borgonovo E. A new uncertainty importance measure. Reliab Eng Syst Saf 2007;92(6):771–84.
[5] Soboĺ IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simulation 2001;55(1–3):271–80.
[6] Sudret B. Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst Saf 2008;93(7):964–79.
[7] Ehre M, Papaioannou I, Straub D. Global sensitivity analysis in high dimensions with PLS-PCE. Reliab Eng Syst Saf 2020;198:106861.
[8] Xiu D, Karniadakis GE. The Wiener–Askey polynomial chaos for stochastic differential equations. SIAM J Sci Comput 2002;24(2):619–44.
[9] Blatman G, Sudret B. Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. Reliab Eng Syst Saf 2010;95(11):1216–29.
[10] Thapa M, Missoum S. Uncertainty quantification and global sensitivity analysis of composite wind turbine blades. Reliab Eng Syst Saf 2022;222:108354.
[11] Mara TA, Becker WE. Polynomial chaos expansion for sensitivity analysis of model output with dependent inputs. Reliab Eng Syst Saf 2021;214:107795.
[12] Schöbi R, Sudret B. Global sensitivity analysis in the context of imprecise probabilities (p-boxes) using sparse polynomial chaos expansions. Reliab Eng Syst Saf 2019;187:129–41.
[13] Sudret B, Mai CV. Computing derivative-based global sensitivity measures using polynomial chaos expansions. Reliab Eng Syst Saf 2015;134:241–50.
[14] Papaioannou I, Breitung K, Straub D. Reliability sensitivity estimation with sequential importance sampling. Struct Saf 2018;75:24–34.
[15] Sarazin G, Morio J, Lagnoux A, Balesdent M, Brevault L. Reliability-oriented sensitivity analysis in presence of data-driven epistemic uncertainty. Reliab Eng Syst Saf 2021;215:107733.
[16] Papaioannou I, Straub D. Variance-based reliability sensitivity analysis and the FORM α-factors. Reliab Eng Syst Saf 2021;210:107496.
[17] Maume-Deschamps V, Niang I. Estimation of quantile oriented sensitivity indices. Statist Probab Lett 2018;134:122–7.
[18] Geraci G, Congedo PM, Abgrall R, Iaccarino G. High-order statistics in global sensitivity analysis: Decomposition and model reduction. Comput Methods Appl Mech Engrg 2016;301:80–115.
[19] Wong CY, Seshadri P, Parks G. Extremum sensitivity analysis with polynomial Monte Carlo filtering. Reliab Eng Syst Saf 2021;212:107609.
[20] Nogal M, Nogal A. Sensitivity method for extreme-based engineering problems. Reliab Eng Syst Saf 2021;216:107997.
[21] Rocha J, Henriques A, Calçada R. Probabilistic safety assessment of a short span high-speed railway bridge. Eng Struct 2014;71:99–111.
[22] Wan H-P, Ni Y-Q. An efficient approach for dynamic global sensitivity analysis of stochastic train-track-bridge system. Mech Syst Signal Process 2019;117:843–61.
[23] Saltelli A, Aleksankina K, Becker W, Fennell P, Ferretti F, Holst N, Li S, Wu Q. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environ Model Softw 2019;114:29–39.
[24] Wang W, Zhang Y, Ouyang H. Modeling uncertainties of vehicle-track coupled dynamic systems. Mech Based Des Struct Mach 2021;49(7):947–68.
[25] Oregui M, Núñez A, Dollevoet R, Li Z. Sensitivity analysis of railpad parameters on vertical railway track dynamics. J Eng Mech 2017;143(5):04017011.
[26] Xu L, Zhai W, Gao J. Global sensitivity analysis for vehicle–track interactions: special attention on track irregularities. J Comput Nonlinear Dyn 2018;13(3).
[27] Liu X, Jiang L, Lai Z, Xiang P, Chen Y. Sensitivity and dynamic analysis of train-bridge coupled system with multiple random factors. Eng Struct 2020;221:111083.
[28] Xin L, Li X, Zhu Y, Liu M. Uncertainty and sensitivity analysis for train-ballasted track–bridge system. Veh Syst Dyn 2020;58(3):453–71.
[29] Blatman G, Sudret B. Adaptive sparse polynomial chaos expansion based on least angle regression. J Comput Phys 2011;230(6):2345–67.
[30] Ghanem RG, Spanos PD. Stochastic finite elements: a spectral approach. Courier Corporation; 2003.
[31] Sudret B. Uncertainty propagation and sensitivity analysis in mechanical models–Contributions to structural reliability and stochastic spectral methods (Ph.D. thesis), Clermont-Ferrand, France: Habilitationa diriger des recherches, Université Blaise Pascal; 2007.
[32] Blatman G. Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis (Ph.D. thesis), Clermont-Ferrand 2; 2009.
[33] Lee SH, Kwak BM. Response surface augmented moment method for efficient reliability analysis. Struct Saf 2006;28(3):261–72.
[34] Ferreira A, Fantuzzi N. MATLAB codes for finite element analysis: Solids and structures. Solid mechanics and its applications, Springer International Publishing; 2020.
[35] MATLAB. version 9.11.0 (R2021b). Natick, Massachusetts, United States: The MathWorks Inc.; 2021.
[36] Cantero D. TTB-2D: Train–Track–Bridge interaction simulation tool for Matlab. SoftwareX 2022;20:101253.
[37] Shen C, Dollevoet R, Li Z. Fast and robust identification of railway track stiffness from simple field measurement. Mech Syst Signal Process 2021;152:107431.
[38] Zhai W. Vehicle–track coupled dynamics: Theory and applications. Singapore: Springer Nature; 2019.
[39] Iwnick S. Manchester benchmarks for rail vehicle simulation. Veh Syst Dyn 1998;30(3–4):295–313.
[40] Zhai W, Wang K, Lin J. Modelling and experiment of railway ballast vibrations. J Sound Vib 2004;270(4–5):673–83.
[41] Xia H, Zhang N, De Roeck G. Dynamic analysis of high speed railway bridge under articulated trains. Comput Struct 2003;81(26–27):2467–78.
[42] Railway applications - Track - Track geometry quality Part 5: Geometric quality levels - Plain line, switches and crossings. Standard, Brussels: European Committee for Standardization; 2017.
[43] Deng C, Zhou J, Thompson D, Gong D, Sun W, Sun Y. Analysis of the consistency of the Sperling index for rail vehicles based on different algorithms. Veh Syst Dyn 2021;59(2):313–30.
[44] Marelli S, Sudret B. UQLab: A framework for uncertainty quantification in Matlab. In: Vulnerability, uncertainty, and risk: quantification, mitigation, and management. American Society of Civil Engineers; 2014, p. 2554–63.
[45] Teixeira R, Nogal M, O'Connor A. Adaptive approaches in metamodel-based reliability analysis: A review. Struct Saf 2021;89:102019.
[46] Schöbi R, Sudret B, Wiart J. Polynomial-chaos-based Kriging. Int J Uncertain Quantif 2015;5(2).
[47] Shang Y, Nogal M, Teixeira R, Wolfert ARM. Optimal design of rail level crossings and associated transition zones using adaptive surrogate-assisted optimization. Eng Struct 2023;282:115740.
[48] Milošević MD, Pålsson BA, Nissen A, Nielsen JC, Johansson H. Reconstruction of sleeper displacements from measured accelerations for model-based condition monitoring of railway crossing panels. Mech Syst Signal Process 2023;192:110225.
[49] Xu L, Zhai W. A new model for temporal–spatial stochastic analysis of vehicle–track coupled systems. Veh Syst Dyn 2017;55(3):427–48.
[50] Eurocode - Basis of structural and geotechnical design. Standard, Brussels: European Committee for Standardization; 2021.