Taylor & Francis
Taylor & Francis Group

Engineering Optimization

ISSN: 0305-215X (Print) 1029-0273 (Online) Journal homepage: www.tandfonline.com/journals/geno20

Sampling-efficient design optimization for categorical configurations and application to turbine blade cooling structures

Yiming Zhang, Bingkun Guo, Xiaojian Liu, Jiaqi Luo, Shuyou Zhang, Weifei Hu & Nam-Ho Kim

To cite this article: Yiming Zhang, Bingkun Guo, Xiaojian Liu, Jiaqi Luo, Shuyou Zhang, Weifei Hu & Nam-Ho Kim (2025) Sampling-efficient design optimization for categorical configurations and application to turbine blade cooling structures, Engineering Optimization, 57:10, 2846-2868, DOI: 10.1080/0305215X.2024.2418969

To link to this article: https://doi.org/10.1080/0305215X.2024.2418969

ENGINEERING OPTIMIZATION
2025, VOL. 57, NO. 10, 2846–2868
https://doi.org/10.1080/0305215X.2024.2418969

Taylor & Francis
Taylor & Francis Group

Sampling-efficient design optimization for categorical configurations and application to turbine blade cooling structures

Yiming Zhang[a,b], Bingkun Guo[b], Xiaojian Liu[b], Jiaqi Luo[a,c], Shuyou Zhang[b], Weifei Hu[b] and Nam-Ho Kim[d]

[a]Huanjiang Laboratory, Zhejiang University, Zhuji, People's Republic of China; [b]School of Mechanical Engineering, Zhejiang University, Hangzhou, People's Republic of China; [c]School of Aeronautics and Astronautics, Zhejiang University, Hangzhou, People's Republic of China; [d]Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, USA

ABSTRACT

Optimizing structural designs, especially for complex systems like turbine blade cooling structures, requires efficient strategies for handling categorical configurations alongside computationally expensive simulations. This article presents a Bayesian optimization strategy tailored for integrated tasks involving categorical configurations and high-dimensional continuous design variables. A Gaussian process with Con-Cat kernel is proposed in order to merge response datasets from diverse configurations effectively, capturing inter-configuration and intra-configuration correlations seamlessly. Additionally, a supervised dimension–reduction scheme is developed based on subspace activation, utilizing a half-Cauchy distribution. Remarkably, the Con-Cat kernel represents a generalization of standard kernels, achieving equivalence in scenarios solely involving continuous variables. The subspace activation scheme enhances surrogate modelling performance without introducing extra model parameters, which is particularly beneficial for sparse datasets. Numerical evaluations, including three mathematical functions, a supported beam problem, and turbine blade cooling structure design optimization, demonstrate the superiority of the proposed Bayesian optimization strategy, exhibiting up to an 85% improvement over five alternative approaches, especially in scenarios with sparse data.

ARTICLE HISTORY
Received 28 March 2024
Accepted 15 October 2024

KEYWORDS
Structural design optimization; Bayesian optimization; high-dimensional optimization; categorical-variable kernel; surrogate modelling

1. Introduction

The integrated design of complex structures, incorporating both categorical and continuous variables, presents a formidable challenge (F. Wang, Zhang, and Zhou 2021; Kaveh, Rahmani, and Eslamlou 2022; Y. Zhang, Apley, and Chen 2020). Such designs require thoughtful consideration of various configurations and material selections, seamlessly integrated with continuous variables like material properties and geometric dimensions. Achieving this integration is crucial for constructing optimization frameworks that accurately reflect the intricacies of industrial design challenges.

CONTACT Jiaqi Luo jiaqil@zju.edu.cn

As the number of variables in a design space increases, the complexity of optimization tasks grows rapidly (Kirschner *et al.* 2019; Letham *et al.* 2020; Malu, Dasarathy, and Spanias 2021). This expansion creates a high-dimensional landscape that is difficult to navigate, posing a challenge to even the most sophisticated optimization algorithms. Finding optimal solutions within this vast space requires algorithms to be both computationally efficient and innovative to overcome the so-called 'curse of dimensionality'.

The scarcity of training data exacerbates the difficulty of constructing probabilistic models in structural design optimization problems. These models are crucial for making reliable predictions and quantifying uncertainties, yet they often suffer from limited accuracy and generalizability owing to the sparse nature of high-dimensional data. Developing precise and broadly applicable models is a pressing challenge, especially when the available data does not adequately represent the entire design space.

Furthermore, the computational expense of evaluating potential solutions cannot be overlooked. Objective functions relying on detailed numerical simulations are essential for capturing the intricate behaviour of complex systems. However, the high computational demands of such simulations can severely constrain the optimization process, particularly when multiple evaluations are necessary to explore the solution space effectively.

Addressing these multifaceted issues is not just an academic endeavour but a practical necessity driving innovation in structural design disciplines. This article aims to advance the field by exploring new optimization strategies that can navigate these challenges with greater efficiency and effectiveness, ultimately enhancing the structural design process.

Traditional methods for handling categorical variables (Daxberger *et al.* 2021; Deshwal, Belakaria, and Doppa 2021; Iyer *et al.* 2019; Luo *et al.* 2023; Nguyen *et al.* 2020; Qiu *et al.* 2024; Ru *et al.* 2020; Saves *et al.* 2022) can be categorized as three approaches. The first approach involves employing one-hot encoding, transforming each categorical variable into a binary vector. While preserving the distinctiveness of categorical variables, this method may lead to computational inefficiencies owing to increased dimensionality, especially with a large number of categories. The second method entails directly replacing categorical variables with fixed numerical values, simplifying computations but potentially introducing an ordinal relationship between categories and biases in model predictions. The third method constructs separate Gaussian Process (GP) models for each category within the dataset, allowing for personalized modelling and potentially more accurate predictions. However, this approach may escalate computational complexity and require more data for effective modelling, especially for less frequent categories.

These traditional approaches each present their own set of advantages and limitations in handling categorical variables (Daxberger *et al.* 2021; Deshwal, Belakaria, and Doppa 2021; Iyer *et al.* 2019; Nguyen *et al.* 2020; Qiu *et al.* 2024; Ru *et al.* 2020; Saves *et al.* 2022). Exploring the nuances of these traditional approaches is essential for determining the most suitable approach based on the dataset's specific characteristics and the modelling task's objectives. Additionally, further research may aim to devise hybrid or novel techniques that effectively mitigate drawbacks while capitalizing on existing methodologies' advantages.

This article focuses on adaptive sampling based on Bayesian Optimization (BO), a robust, sample-efficient technique for structural design optimization with computationally expensive simulations (Eriksson *et al.* 2019; Foumani *et al.* 2023; Luo *et al.* 2023; Qiu *et al.* 2024; Sabater *et al.* 2021; Shahriari *et al.* 2015; Sheikh *et al.* 2022; Snoek, Larochelle, and Adams 2012; X. Wang *et al.* 2023; Yang, Kissas, and Perdikaris 2022). However, its applicability has been restricted to low-dimensional and fully continuous problems owing to computational and statistical challenges in high-dimensional settings and hybrid spaces, including variable representation and optimization, search space complexity, algorithm adaptability and the trade-off between sampling efficiency and evaluation cost (Cavazzuti 2013; Helton and Davis 2003; Kirschner *et al.* 2019; Malu, Dasarathy, and Spanias 2021; Rodríguez *et al.* 2018; Ru *et al.* 2020; Sabater *et al.* 2021; Shields and Zhang 2016).

A novel adaptive sampling strategy leveraging Bayesian optimization is proposed to handle categorical configurations and high-dimensional continuous variables integratively. The key contributions include the following.

- Introduction of a Gaussian process with a novel Con-Cat kernel for fusing datasets from different structural configurations, capturing inter-configuration and intra-configuration correlations simultaneously.
- Development of a supervised dimension–reduction strategy using the half-Cauchy distribution, designed as the Gaussian process's prior, and activating critical design variables through maximum a-posteriori estimation.
- Proposal of a Bayesian optimization strategy for integrated design optimization of different configurations and high-dimensional variables, enhanced with a multi-start gradient descent algorithm for thorough optimization of the acquisition function. The strategy's superiority is validated for turbine blade cooling structure design with four configurations.

The outline of the article is as follows. Section 2 introduces the basics of BO. Section 3 elaborates on the proposed strategy with theoretical details. Section 4 describes numerical evaluations using three mathematical functions and the design optimization of a simply supported beam. Section 5 applies the proposed strategy for the design optimization of a turbine blade cooling structure. Conclusions and perspectives are finally summarized in Section 6.

2. Basics of Bayesian optimization

This section provides an overview of the fundamental concepts of a typical BO framework. BO for structural design involves a systematic process for efficiently exploring the design space and identifying optimal solutions while minimizing computational costs. It begins with the formulation of the optimization task $min_{\mathbf{x}\in\Omega} f(\mathbf{x})$ where f is the structural response to optimize with and $\mathbf{\Omega}$ is the design space.

In the initial stage, an exploration of the design space is conducted using an initial set of designs generated through techniques like Latin hypercube sampling or random sampling. These designs are then evaluated using simulation runs, which quantify the performance of each design based on criteria such as stress distribution, displacement, weight, or cost.

To navigate the design space efficiently, Bayesian optimization relies on surrogate modelling. A surrogate model, often based on GP regression, approximates the objective function. This model captures the relationship between the design variables and the objective function, allowing for efficient exploration. It provides not only predictions of the objective function but also estimates of uncertainty. GP is defined by a mean function $\mu(x)$ and a covariance function $k(x, x')$. A common choice for the covariance function is the Radial Basis Function (RBF) kernel, which is defined as

$$k_{\mathrm{RBF}}(x, x') = \sigma_k^2 \exp\left(-\frac{||x - x'||^2}{2\ell^2}\right). \tag{1}$$

With the surrogate model in place, an acquisition function is used to guide the selection of the next design to evaluate. The acquisition function balances exploration (sampling from regions of high uncertainty) and exploitation (sampling from regions likely to yield optimal designs). Expected Improvement (EI) is a common acquisition function, defined as

$$\mathrm{EI}(\mathbf{x}) = \mathbb{E}\left[\max(f(\mathbf{x}^*) - f(\mathbf{x}), 0)\right], \tag{2}$$

where $\mathbf{x}^*$ is the current best design, and the expectation is taken under the GP posterior distribution of f at $\mathbf{x}$.

When extending BO to handle constraints, the optimization task is reformulated as

$$\min_{\mathbf{x}\in\Omega} f(\mathbf{x}) \quad \text{subject to } c_j(\mathbf{x}) \leq 0,\ j = 1, \ldots, J. \tag{3}$$

Each constraint function $c_j(\mathbf{x})$ in Equation (3) is modelled with its individual GP, yielding posterior means $\mu_j(\mathbf{x})$ and variances $\sigma_j^2(\mathbf{x})$. A feasibility probability $\mathrm{PF}(\mathbf{x})$ is developed to quantify the satisfaction of constraints as

$$\mathrm{PF}(\mathbf{x}) = \prod_{j=1}^{J} \Phi\left(-\frac{\mu_j(\mathbf{x})}{\sigma_j(\mathbf{x})}\right), \tag{4}$$

where Φ is the cumulative distribution function of the standard normal distribution. The EI acquisition with constraints (EI_C) is then defined as

$$\mathrm{EI}_C(\mathbf{x}) = \mathrm{EI}(\mathbf{x}) \cdot \mathrm{PF}(\mathbf{x}), \tag{5}$$

which is adopted to guide the selection of valuable simulation runs.

The selected designs are evaluated using the simulation runs, and the results are used to update the surrogate model. Bayesian optimization iterates until a termination criterion is met, such as reaching a predefined number of iterations or achieving a satisfactory level of convergence.

3. The proposed Bayesian optimization for categorical configurations

This section discusses the proposed Con-Cat kernel for fusing datasets from different configurations, the subspace activation scheme using the half-Cauchy distribution and the overall framework of Bayesian optimization.

3.1. The Con-Cat kernel for non-stationary surrogate modelling

The GP has been well-accepted for surrogate modelling of sparse datasets. The correlation between design variables is modelled through the kernel function of the GP. This article proposes the Con-Cat kernel, a novel GP kernel designed to merge continuous (x_{con}) and categorical (x_{cat}) variables adeptly into a cohesive framework. This kernel is crucial for approximating the target response values $f(x_h)$, where $x_h = [x_{\mathrm{con}}, x_{\mathrm{cat}}]$ represents the mixed input space.

The inputs to the Con-Cat kernel, denoted as x_h and x_h', are are defined as

$$\begin{aligned} x_h &= [x_{\mathrm{con}}, x_{\mathrm{cat}}] = [x_{\mathrm{con}}, i], \quad i = 1, 2, \ldots, N \\ x_h' &= \left[x_{\mathrm{con}}', x_{\mathrm{cat}}'\right] = \left[x_{\mathrm{con}}', j\right], \quad j = 1, 2, \ldots, N, \end{aligned} \tag{6}$$

where i represents x_h belonging to the ith category, j represents x_h' belonging to the jth category and N denotes the total number of categories.

The Con-Cat kernel is articulated through two primary components, where $k_1(x_h, x_h')$ emphasizes the inter-configuration correlation and $k_2(x_h, x_h')$ focuses on the intra-configuration correlation as

$$k_h\left(x_h, x_h'\right) = (1-\lambda)k_1\left(x_h, x_h'\right) + \lambda k_2\left(x_h, x_h'\right), \tag{7}$$

where the trade-off between $k_1(x_h, x_h')$ and $k_2(x_h, x_h')$ is controlled by a parameter $\lambda \in [0, 1]$, which is optimized jointly during GP training.

The first component $k_1(x_h, x_h')$ is engineered to compute the inter-category similarity while retaining the sensitivity to the continuous variables' nuances as

$$k_1\left(x_h, x_h'\right) = k_1\left([x_{\mathrm{con}}, i], \left[x_{\mathrm{con}}', j\right]\right) = k_{\mathrm{RBF}}\left(x_{\mathrm{con}}, x_{\mathrm{con}}'\right) \cdot m_{ij} \tag{8}$$

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1N} \\ \cdots & m_{22} & \cdots & m_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \cdots & \cdots & \cdots & m_{NN} \end{bmatrix}' \tag{9}$$

where $k_{\mathrm{RBF}}(x_{\mathrm{con}}, x'_{\mathrm{con}})$ is the RBF kernel, which captures the correlation between continuous variables x_{con} and $x'_{\mathrm{con}} \cdot (m_{ij})$ is the element of matrix $\mathbf{M}$ in the ith row and jth column, representing the correlation between the ith and jth categories, ranging between zero and one. $\mathbf{M}$ is a positive semi-definite matrix encapsulating the similarities across categories and enabling a unified view of the design space as the correlation m_{ij} are trainable weights obtained during GP training. A larger value of m_{ij} indicates that the ith and jth categories are more similar, and vice versa.

The second component $k_2(x_h, x'_h)$ aims to capture the intra-category similarities emphasizing the distinct variations within each categorical setting, defined as

$$k_2\left(x_h, x'_h\right) = k_2\left(\left[x_{\mathrm{con}}, i\right], \left[x'_{\mathrm{con}}, j\right]\right) = \sum_{q=1}^{N} k_{\mathrm{RBF}}^{(q)}\left(x_{\mathrm{con}}, x'_{\mathrm{con}}\right) \cdot \delta_{ij}^{(q)} \tag{10}$$

$$\delta_{ij}^{(q)} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{else,} \end{cases} \tag{11}$$

where q represents the categorical variable type and $\delta_{ij}^{(q)}$ is a switch function as defined in Equation (11). In evaluating the correlation between mixed variables x_h and x'_h, a correlation of zero is assigned if x_h and x'_h do not share the same category. This mechanism effectively sets up separate GPs for different categories, facilitating the adaptation to the non-stationary smoothness of the objective function under each category without inter-category information exchange.

3.2. Subspace activation using half-Cauchy distribution

Integrating priors into GP modelling significantly enhances the versatility of function representation. By carefully selecting priors, particularly through kernel function choices, function attributes such as smoothness and periodicity can be manipulated adeptly, as noted by Chen and Wang (2018). This flexibility allows GPs to handle a broad spectrum of data patterns adeptly, highlighting their strength as a non-parametric method. Additionally, priors act as a form of regularization in GP models, aiding in the prevention of overfitting to training data through strategic prior distribution selection, thereby boosting the model's generalization ability. This feature is especially vital in situations with sparse data or high noise levels, underlining the role of prior distributions in bolstering the robustness of GPs against overfitting.

In the realm of high-dimensional optimization, leveraging priors within GP modelling can identify low-effective-dimensional structures within high-dimensional functions. GP-based surrogate models are utilized, incorporating a subspace activation prior on kernel hyperparameters. This approach induces a sparse structure in the RBF kernel, efficiently pinpointing crucial low-dimensional subspaces via active dimensions. Specifically, the model is defined as

$$k_{\mathrm{RBF}}(x, x') = \sigma_k^2 \exp\left(-\frac{1}{2\ell^2}||x - x'||^2\right) = \sigma_k^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{D} \rho_i (x_i - x'_i)^2\right), \tag{12}$$

where ρ_i for $i = 1, \ldots, D$ are inverse squared length scales representing the smoothness and importance of the ith dimension and D represents the dimensionality of the input space.

Building on this, the subspace activation prior is applied over the kernel hyperparameters, notably the inverse squared length scales ρ_i, to discern the key low-dimensional subspaces indicated by active dimensions rapidly. The following prior distributions are developed as

$$\begin{aligned}
&\text{length scales}: \quad \rho_i \sim \mathcal{HC}(0,\alpha) \quad \text{for } i = 1,\ldots,D \\
&\text{kernel variance}: \quad \sigma_k^2 \sim \mathcal{LN}(0,10^2) \\
&\text{function values}: \quad \mathbf{f} \sim \mathcal{N}\left(\mathbf{0}, K_{\mathbf{xx}}^{\psi}\right) \\
&\text{observations}: \quad \mathbf{y} \sim \mathcal{N}\left(\mathbf{f}, \sigma^2 \mathbf{1}_N\right).
\end{aligned} \tag{13}$$

The half-Cauchy ($\mathcal{HC}$) and log-normal ($\mathcal{LN}$) distributions allow for parameter variability across a wide range, accommodating diverse data characteristics and averting overly deterministic parameter inferences in sparse-sample scenarios. The probability density functions of the half-Cauchy and log-normal distributions are given as

$$\begin{aligned}
&\text{Half}-\text{Cauchy}: \quad \text{PDF}_1(x;x_0,\sigma) = \frac{2}{\pi\sigma}\frac{1}{1+(x-x_0)^2/\sigma^2} \\
&\text{Log}-\text{Normal}: \quad \text{PDF}_2(x;\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-(\ln x-\mu)^2/2\sigma^2}.
\end{aligned} \tag{14}$$

The hyperparameters of the GP in Equations (12) and (13) are determined by employing Maximum A-Posteriori (MAP) estimation, which maximizes the posterior distribution of the hyperparameters given the observed data:

$$\hat{\psi}_{\text{MAP}} = \arg\max_{\psi} \left\{\log p(\mathbf{y}|\mathbf{X},\psi) + \log p(\psi)\right\}, \tag{15}$$

where ψ represents the hyperparameters of the GP, $\mathbf{y}$ is the vector of observed target values and $\mathbf{X}$ is the matrix of input features.

The inverse square length scales ρ_i in Equations (12) and (13) are influenced by the half-Cauchy prior in Equation (14) and Figure 1, leading them to tend towards zero. As a result, most dimensions are turned off. The half-Cauchy prior's distribution exhibits heavy-tailed properties. Consequently, ample evidence in the observed values y will shift the posterior of ρ_i towards higher values by the MAP estimation in Equation (15), effectively turning on the respective dimensions.

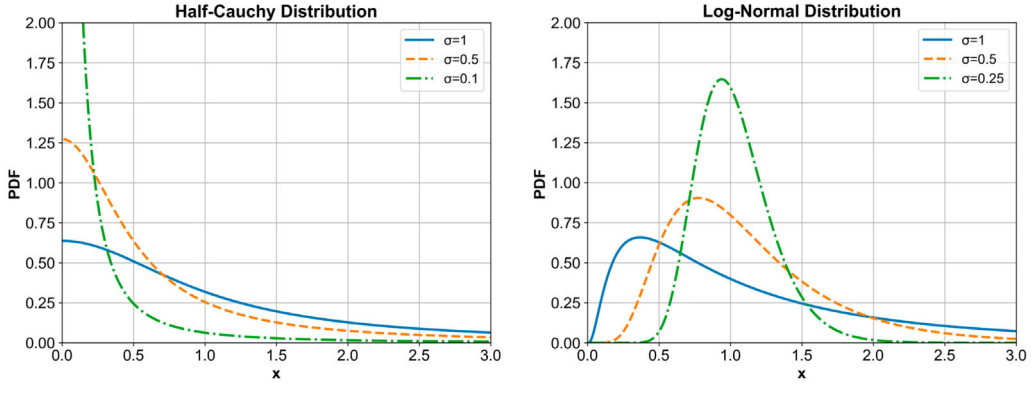

Figure 1. Probability density distribution functions of Half-Cauchy distribution (left) and Lognormal distribution (right) with different standard deviations. The location parameter x_0 of the Half-Cauchy distribution is set to zero, and the mean μ of the normal distribution obtained by the logarithmic transformation of the log-normal distribution is set to zero.

3.3. Overview of the proposed Bayesian optimization framework

This section delineates the overarching structure of the Bayesian optimization framework devised for structural optimization with categorical configurations. The framework, detailed in Algorithm 1, adopts surrogate modelling with optimization of the acquisition function to pinpoint the optimal solution within a constrained evaluation budget, adhering to specified constraints.

Algorithm 1 The proposed adaptive sampling for handling categorical configurations

Input: Objective function f_{obj}; constraint functions g_j for $j = 1, 2, \ldots, J$; total evaluation budget $T > m$; and initial query set $x_{1:m}$ from the DoE $\mathcal{D}_t$ using the LHS method and evaluations of objective and constraint function $f_{\text{obj}}(x_{1:m}), g_j(x_{1:m})$ for $j = 1, 2, \ldots, J$.

Output: Solution of the problem $min_{x\in\Omega'} f(x)$ over design space $\Omega'$of mixed variables with constraints.

1 **while** *the stopping criterion is not satisfied* **do**

2 Build the surrogate model of the objective function to obtain the mean $\hat{f}$ and standard deviation prediction s^f at all given points $\{x\}$ from the DoE $\mathcal{D}_t$. Build the surrogate models for every constraint to obtain the mean $\hat{g}_j$ and standard deviation prediction s_j^g at all given points $\{x\}$ for $j = 1, 2, \ldots, J$ from the DoE $\mathcal{D}_t$. Construct the acquisition function with EI_C from the surrogate models; maximize the acquisition function with EI_C using multi-restart gradient-based optimizer to select the next sample point x_t. Add $x_t, f(x_t), g_j(x_t)$ for $j = 1, 2, \ldots, J$ to the DoE $\mathcal{D}_{t+1}$; increment t.

3 Identify the optimal solution x^* from $\mathcal{D}_t$ that has the lowest objective function value $f(x^*)$ while satisfying all constraints. **Return** x^* and $f(x^*)$.

Initiating with a Design of Experiments (DoE) $\mathcal{D}_t$ through Latin Hypercube Sampling (LHS), the strategy establishes an initial query set $x_{h,1:m}$ within the mixed space Ω' that serves as the basis for subsequent evaluations of the objective function $f_{\text{obj}}(x_{1:m})$ and J constraint functions g_j for $j = 1, 2, \ldots, J$.

GP models, leveraging the novel Con-Cat kernel introduced in Section 3.1, are iteratively refined from the DoE $\mathcal{D}_t$. These models are adept at predicting the mean and standard deviation of the aforementioned functions at new points. In the Con-Cat kernel, the subspace activation prior is applied to the RBF kernels, dictating that all hyperparameters $\boldsymbol{\rho}$ are optimized. This setup allows the closed-form computation of the observed data's marginal likelihood.

The strategy iteratively updates the DoE $\mathcal{D}_t$ by incorporating new samples, each iteration drawing closer to the optimal solution that minimizes the objective function while satisfying all constraints. The process leverages the surrogate models to construct and optimize the acquisition function, systematically identifying the next sampling point, thereby refining the search within the design space.

In addressing the challenge of maximising the acquisition function within highly non-stationary spaces, this framework adopts a multi-point restart gradient-based optimizer for searching the optimum acquisition function values. Compared with heuristic algorithms such as genetic algorithms, the multi-point restart gradient-based algorithm can more accurately identify directions of improvement, even in complex, multi-dimensional spaces characterized by non-stationarity. The efficient search of acquisition function values could further enhance the convergence speed of the design optimization framework.

The BO framework proposed here is for categorical and continuous high-dimensional spaces by enhancing mixed data representation, facilitating efficient high-dimensional optimization and ensuring thorough optimization of the acquisition function. This strategy not only adapts to complex

correlations within mixed data but also dynamically adjusts to new data without compromising the model's capability of capturing the essential features of the objective function, thereby streamlining the optimization process in engineering applications.

4. Numerical evaluations

This section presents numerical validation of the proposed BO strategy based on six mathematical functions and one finite element simulation. The efficacy of the proposed approach is demonstrated through comparison with seven benchmark optimization strategies: StandardBO (Snoek, Larochelle, and Adams 2012), Sparse Axis-Aligned Subspace BO (SAASBO) (Eriksson and Jankowiak 2021), Additive Latent Embedding BO (ALEBO) (Letham *et al.* 2020), Multilayer Perceptron Gaussian Process BO (MPGPBO) (Wilson *et al.* 2016), Latin Hypercube Sampling BO (LHSBO) (Shields and Zhang 2016), Continuous and Categorical BO (CoCaBO) (Ru *et al.* 2020) and Multi-armed Bandit BO (BanditBO) (Nguyen *et al.* 2020).

To address the challenge of categorical variables, One-hot encoding was employed for StandardBO, SAASBO, ALEBO, MPGPBO and LHSBO, enabling them to handle categorical variables effectively. One-hot encoding transforms each categorical variable into a binary vector, where each category is represented by a unique binary vector. This allows the optimization algorithms to process categorical variables as numerical inputs, facilitating their integration into the optimization process.

4.1. Evaluation setup

The selection of subsequent data points was based on maximizing the expected improvement with constraints (EI_C). The optimization process was repeated with 10 sets of initial samples for a comprehensive evaluation. Each optimization strategy was allocated the same number of iterations (*i.e.* 100), with the initial points set to 10 for each category.

Two metrics were adopted for assessing convergence performance: the Mean Error (ME) for convergence speed and the Median Deviation (MD) for algorithm robustness. A lower ME indicates faster convergence, while a lower MD suggests higher robustness, essential for minimization problems.

To compare the performance of the proposed method against benchmarks, the Promotion Rate (PR) metric was introduced. This metric combines measures of convergence speed and robustness relative to the theoretical optimization potential:

$$\mathrm{PR} = \left(\frac{0.5 \cdot (\mathrm{ME}_2 - \mathrm{ME}_1) + 0.5 \cdot (\mathrm{MD}_2 - \mathrm{MD}_1)}{\mathrm{UB} - \mathrm{LB}} \right) \times 100\%, \tag{16}$$

where ME_1 and MD_1 represent the mean and median of the final convergence results of the proposed scheme, while ME_2 and MD_2 represent other baselines for comparison. For mathematical functions, the global minimum served as the theoretical Lower Bound (LB) for optimization efficacy. In engineering applications, direct computation of the global minimum is often impractical; therefore, the best optimization result across all methods was used as LB, and the average initial value was used as the theoretical upper bound (UB). UB − LB quantifies the optimal optimization range achievable by all methods.

The weights of convergence speed and robustness on the impact of PR were set to 0.5 each in Equation (16). A higher PR value indicates superior performance of the proposed scheme, demonstrating its effectiveness in achieving faster convergence and greater robustness compared to the baseline methods.

4.2. Mathematical functions

4.2.1. Mathematical function 1

Function 1 explores a high-dimensional space with continuous and categorical variables, where the first three continuous variables and one categorical dimension are pivotal owing to their significant impact on the function's output, while the remaining 16 dimensions are deemed less significant and can be considered as noise. This setup tests the proposed method's ability to discern and prioritize influential dimensions for categorical–continuous design variables.

$$f_1\left(x_{\text{con}}, x_{\text{cat}}\right) = \begin{cases} -\sum_{i=1}^{3} x_i^2, & x_{\text{cat}} = A \\ \sum_{i=1}^{3} 1.2 * \sin\left(2\pi x_i\right), & x_{\text{cat}} = B \\ \sum_{i=1}^{3} \left\{0.1 * \sin\left(2\pi x_i\right) - 0.25 * pdf\left(x_i, 0.5, 0.05\right)\right\}, & x_{\text{cat}} = C \\ \sum_{i=1}^{3} \left\{0.1 * \sin\left(2\pi x_i\right) - 1.0\right\}, & x_{\text{cat}} = D \end{cases} \tag{17}$$

$$pdf(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{18}$$

$$x_{\text{cat}} = x_0 \in [A, B, C, D], \quad x_{\text{con}} = [x_1, x_2, \ldots, x_{19}] \in [0, 1]^d. \tag{19}$$

4.2.2. Mathematical function 2

Function 2, a variant of the Branin function adjusted for high-dimensional optimization with categorical choices, tests the method's performance in handling complex landscapes. The inclusion of different categorical configurations (A, B, C, D) introduces varying difficulty levels and explores the algorithm's flexibility. The first two continuous variables and one categorical dimension are considered important owing to their significant impact on the function's output, while the remaining 47 dimensions are deemed less significant and can be considered as noise.

$$f_2\left(x_{\text{con}}, x_{\text{cat}}\right) = \begin{cases} \text{Branin}\left(15 * x_1 - 5, 15 * x_2\right)/200 - 4.0, & x_{\text{cat}} = A \\ \text{Branin}\left(10 - 15 * x_1, 15 - 15 * x_2\right)/100 - 4.0, & x_{\text{cat}} = B \\ \text{Branin}\left(15 * x_1 - 5_1, 15 - 15 * x_2\right)/10 - 5.0, & x_{\text{cat}} = C \\ \text{Branin}\left(10 - 15 * x_1, 15 * x_2\right) - 6.0, & x_{\text{cat}} = D \end{cases} \tag{20}$$

$$\begin{aligned} &\text{Branin}(x_1, x_2) = a\left(x_2 - bx_1^2 + cx_1 - r\right)^2 + s(1-t)\cos\left(x_1\right) + s \\ &a = 1,\ b = 5.1/\left(4\pi^2\right),\ c = 5/\pi,\ r = 6,\ s = 10,\ t = 1/(8\pi) \end{aligned} \tag{21}$$

$$x_{\text{cat}} = x_0 \in [A, B, C, D], \quad x_{\text{con}} = [x_1, x_2, \ldots, x_{49}] \in [0, 1]^d. \tag{22}$$

4.2.3. Mathematical function 3

Function 3, inspired by the Hartmann-6 function, extends the challenge to a higher-dimensional space with continuous variables. It further complicates the optimization task with multiple peaks and valleys, evaluating the algorithm's precision in locating global minima amidst numerous local optima. The first six continuous variables and one categorical dimension are pivotal owing to their significant impact on the function's output, while the remaining 93 dimensions are deemed less significant and

can be considered as noise.

$$f_3\left(x_{\text{con}}, x_{\text{cat}}\right) = \begin{cases} -\frac{1}{1.94}\left[2.58 + \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}\left(x_j - P_{ij}\right)^2\right)\right], & x_{\text{cat}} = A \\ -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}\left(x_j - P_{ij}\right)^2\right) + \sum_{j=1}^{6} x_j, & x_{\text{cat}} = B \\ -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}\left(x_j - P_{ij}\right)^2\right) * \left(\sum_{j=1}^{6} x_j - 2\right), & x_{\text{cat}} = C \\ -\frac{1}{1.94}\left[2.58 + \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}\left(x_j - P_{ij}\right)^2\right)\right], & x_{\text{cat}} = D \end{cases} \tag{23}$$

$$\alpha = (1.0, 1.2, 3.0, 3.2)^{\mathrm{T}}$$

$$\mathbf{A} = \begin{pmatrix} 10 & 3 & 17 & 3.50 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{pmatrix} \tag{24}$$

$$\mathbf{P} = 10^{-4} \begin{pmatrix} 1312 & 1696 & 5569 & 124 & 8283 & 5886 \\ 2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\ 2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\ 4047 & 8828 & 8732 & 5743 & 1091 & 381 \end{pmatrix}$$

$$x_{\text{cat}} = x_0 \in [A, B, C, D], \quad x_{\text{con}} = [x_1, x_2, \ldots, x_{99}] \in [0, 1]^d. \tag{25}$$

4.2.4. Mathematical function 4

Function 4 is a variant of the Ackley function, tailored for high-dimensional optimization with categorical choices. This function evaluates the method's performance in navigating complex landscapes. The inclusion of different categorical configurations (A, B, C, D) introduces varying levels of difficulty, thereby testing the algorithm's adaptability. All variables, including the continuous and categorical dimensions, are considered crucial and effective due to their significant impact on the function's output.

$$f_4\left(x_{\text{con}}, x_{\text{cat}}\right) = \begin{cases} \text{Ackley}\left(x_{\text{con}}\right), & x_{\text{cat}} = A \\ \text{Ackley}\left(x_{\text{con}} + 1\right) + 1, & x_{\text{cat}} = B \\ \text{Ackley}\left(x_{\text{con}} + 2\right) + 2, & x_{\text{cat}} = C \\ \text{Ackley}\left(x_{\text{con}} + 3\right) + 3, & x_{\text{cat}} = D \end{cases} \tag{26}$$

$$\text{Ackley}(x) = -a \exp\left(-b\sqrt{\frac{1}{d}\sum_{i=1}^{d} x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^{d} \cos\left(cx_i\right)\right) + a + e$$

$$a = 20,\ b = 0.2,\ c = 2\pi,\ d = 49 \tag{27}$$

$$x_{\text{cat}} = x_0 \in [A, B, C, D], \quad x_{\text{con}} = [x_1, x_2, \ldots, x_{49}] \in [-5, 5]^d. \tag{28}$$

4.2.5. Mathematical function 5

Function 5 is a modified version of the Rosenbrock function, adapted for high-dimensional optimization with categorical choices. This function assesses the method's capability in handling intricate landscapes. The presence of different categorical configurations (A, B, C, D) introduces varying degrees of complexity, challenging the algorithm's flexibility. All variables, including the continuous

and categorical dimensions, are deemed essential and effective owing to their substantial influence on the function's output.

$$f_5\left(x_{\text{con}}, x_{\text{cat}}\right) = \begin{cases} \frac{1}{10^7}\,\text{Rosenbrock}\left(x_{\text{con}}\right), & x_{\text{cat}} = A \\ \frac{1}{10^7}\left(\text{Rosenbrock}\left(x_{\text{con}}+1\right)+1\right), & x_{\text{cat}} = B \\ \frac{1}{10^7}\left(\text{Rosenbrock}\left(x_{\text{con}}+2\right)+2\right), & x_{\text{cat}} = C \\ \frac{1}{10^7}\left(\text{Rosenbrock}\left(x_{\text{con}}+3\right)+3\right), & x_{\text{cat}} = D \end{cases} \tag{29}$$

$$\text{Rosenbrock}(x) = \sum_{i=1}^{d}\left[100\left(x_{i+1}-x_i^2\right)^2 + \left(1-x_i\right)^2\right]$$

$$d = 48 \tag{30}$$

$$x_{\text{cat}} = x_0 \in [A, B, C, D], \quad x_{\text{con}} = [x_1, x_2, \ldots, x_{49}] \in [-5, 10]^d. \tag{31}$$

4.2.6. Mathematical function 6

Function 6 is a modified Rastrigin function, designed for high-dimensional optimization with categorical choices. This function tests the method's effectiveness in managing complex landscapes. The inclusion of various categorical configurations (A, B, C, D) presents different levels of difficulty, exploring the algorithm's versatility. All variables, including the continuous and categorical dimensions, are considered pivotal and effective owing to their significant impact on the function's output.

$$f_6\left(x_{\text{con}}, x_{\text{cat}}\right) = \begin{cases} \frac{1}{10^3}\,\text{Rastrigin}\left(x_{\text{con}}\right), & x_{\text{cat}} = A \\ \frac{1}{10^3}\left(\text{Rastrigin}\left(x_{\text{con}}+1\right)+1\right), & x_{\text{cat}} = B \\ \frac{1}{10^3}\left(\text{Rastrigin}\left(x_{\text{con}}+2\right)+2\right), & x_{\text{cat}} = C \\ \frac{1}{10^3}\left(\text{Rastrigin}\left(x_{\text{con}}+3\right)+3\right), & x_{\text{cat}} = D \end{cases} \tag{32}$$

$$\text{Rastrigin}(x) = Ad + \sum_{i=1}^{d}\left[x_i^2 - A\cos\left(2\pi x_i\right)\right]$$

$$A = 10,\ d = 49 \tag{33}$$

$$x_{\text{cat}} = x_0 \in [A, B, C, D], \quad x_{\text{con}} = [x_1, x_2, \ldots, x_{49}] \in [-5, 5]^d. \tag{34}$$

4.2.7. Evaluation results of mathematical functions

The optimization outcomes across various benchmarks are detailed in Figures 2 and 3. The convergence performance metrics are outlined in Tables 1–6. For mathematical functions 1, 2 and 3 in Equations (17), (20) and (23), the ranges $[LB, UB]$ achieved by the proposed method are $[-6, -3.06]$, $[-5.60, -3.92]$ and $[-5.80, -0.245]$, respectively. For mathematical functions 4, 5 and 6 in Equations (26), (29) and (32), the ranges $[LB, UB]$ obtained by the proposed method are $[4.05, 10.13]$, $[0.00, 0.48]$ and $[0.36, 0.93]$, respectively. In Table 7, the proposed method consistently demonstrated higher prediction accuracy compared to other methods, as reflected in the normalized RMSE values. These results underscore the proposed strategy's superior performance, as evidenced by optimal mean and median outcomes across all tests. While ALEBO showed moderate effectiveness, StandardBO lagged significantly behind. The proposed approach notched improvements ranging from 7.49% to 85.08% over StandardBO, SAASBO, LHSBO, MPGPBO , ALEBO, CoCaBO and BanditBO across the board for these mathematical functions. Given StandardBO and BanditBO's underperformance and the high cost of experimentation in design optimization scenarios, they were subsequently removed from the baseline comparisons.

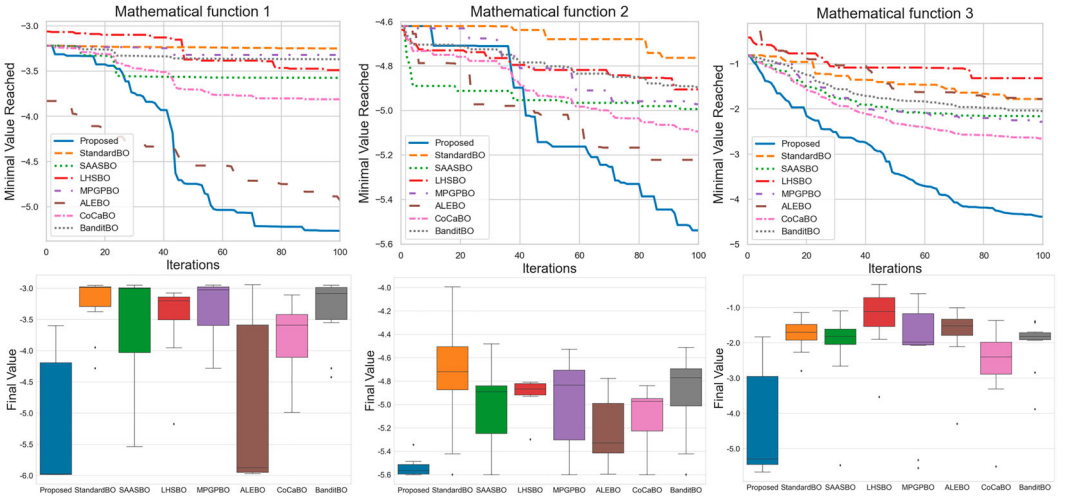

Figure 2. The optimization results of different baselines on three mathematical functions (1, 2 and 3) using a DoE of 40 points over 10 independent replications. For each baseline, the mean value of the best minimum found at a given iteration is depicted as a convergence curve in the **top row**. The distribution over the final approximate minimum after 100 iterations is encoded as a box plot in the **bottom row**, with horizontal bars corresponding to 5%, 50% and 95% quantiles. The proposed method is labelled 'Proposed'.

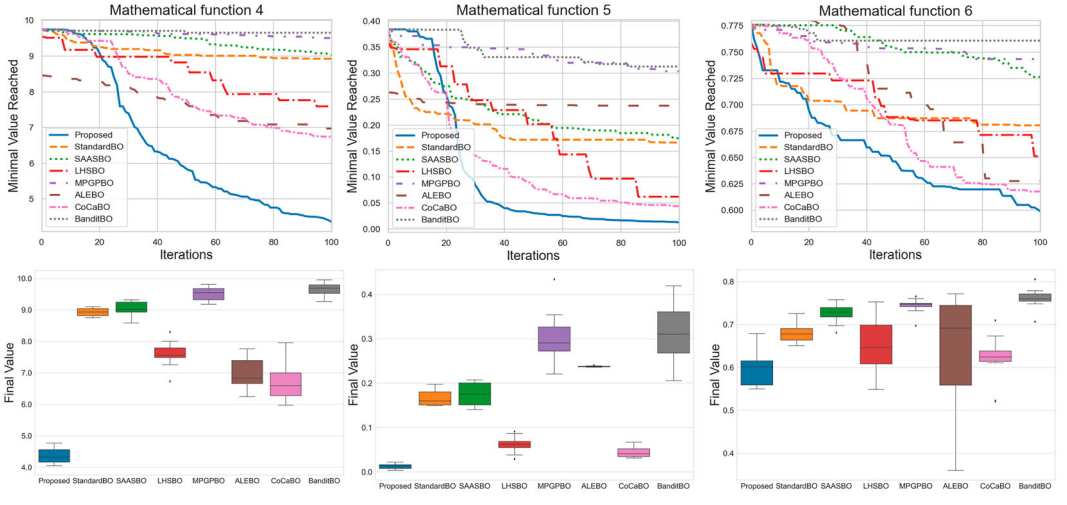

Figure 3. The optimization results of different baselines on three Mathematical functions (4, 5 and 6) using a DoE of 40 points over 10 independent replications. For each baseline, the mean value of the best minimum found at a given iteration is depicted as a convergence curve in **top row**. The distribution over the final approximate minimum after 100 iterations is encoded as a box plot in **bottom row**, with horizontal bars corresponding to 5%, 50% and 95% quantiles. The proposed method is labelled 'Proposed'.

Table 1. Performance metrics for evaluating Mathematical function 1.

	Proposed method	Standard BO	SAASBO	LHSBO	MPGPBO	ALEBO	CoCaBO	BanditBO
Mean result	**−5.26**	−3.25	−3.57	−3.49	−3.32	−4.93	−3.81	−3.37
Median result	**−5.97**	−2.99	−3.00	−3.20	−3.02	−5.87	−3.59	−3.08
PR		85.08%	79.34%	77.40%	83.24%	7.49%	65.26%	81.45%

Note: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.

Table 2. Performance metrics for evaluating Mathematical function 2.

	Proposed method	Standard BO	SAASBO	LHSBO	MPGPBO	ALEBO	CoCaBO	BanditBO
Mean result	**−5.54**	−4.76	−4.99	−4.90	−4.97	−5.22	−5.09	−4.89
Median result	**−5.57**	−4.72	−4.89	−4.87	−4.83	−5.33	−4.97	−4.77
PR		48.30%	36.34%	39.69%	38.69%	16.51%	30.95%	42.85%

Note: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.

Table 3. Performance metrics for evaluating Mathematical function 3.

	Proposed method	Standard BO	SAASBO	LHSBO	MPGPBO	ALEBO	CoCaBO	BanditBO
Mean result	**−4.39**	−1.78	−2.16	−1.32	−2.28	−1.78	−2.66	−2.04
Median result	**−5.29**	−1.70	−1.82	−1.11	−1.98	−1.52	−2.40	−1.82
PR		55.77%	51.28%	65.24%	48.76%	57.42%	41.58%	52.35%

Note: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.

Table 4. Performance metrics for evaluating Mathematical function 4.

	Proposed method	Standard BO	SAASBO	LHSBO	MPGPBO	ALEBO	CoCaBO	BanditBO
Mean result	**4.36**	8.93	9.04	7.59	9.51	6.97	6.74	9.65
Median result	**4.32**	8.93	9.01	7.56	9.55	6.84	6.60	9.69
PR		75.44%	77.00%	53.16%	85.28%	42.14%	38.26%	87.65%

Note: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.

Table 5. Performance metrics for evaluating Mathematical function 5.

	Proposed method	Standard BO	SAASBO	LHSBO	MPGPBO	ALEBO	CoCaBO	BanditBO
Mean result	**0.013**	0.167	0.175	0.062	0.304	0.237	0.044	0.313
Median result	**0.014**	0.160	0.175	0.063	0.291	0.237	0.041	0.310
PR		31.26%	33.74%	10.24%	59.16%	46.66%	6.10%	62.15%

Note: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.

Table 6. Performance metrics for evaluating Mathematical function 6.

	Proposed method	Standard BO	SAASBO	LHSBO	MPGPBO	ALEBO	CoCaBO	BanditBO
Mean result	**0.599**	0.681	0.726	0.651	0.743	0.628	0.618	0.761
Median result	**0.601**	0.678	0.729	0.647	0.747	0.692	0.624	0.760
PR		13.87%	22.38%	8.56%	25.47%	10.44%	3.66%	28.08%

Note: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.

4.3. Simply supported beam design optimization

This study extends the proposed Bayesian Optimization strategy to the simulation-based design problem of a simply supported beam as shown in Figure 4. This beam, characterized by fixed supports at both ends, spans 10 meters and bears a uniformly distributed load of 10,000 N on its top surface. The

Table 7. Comparison of normalized RMSE for different Gaussian process models on six mathematical functions based on 200 training samples and 100 testing samples.

	Proposed method	GP–MLE	GP–Category	MPGP
Mathematical function 1	**0.32**	0.63	0.57	0.74
Mathematical function 2	**0.43**	0.52	0.53	0.70
Mathematical function 3	**0.24**	0.90	0.36	1.11
Mathematical function 4	**0.21**	0.81	0.45	1.05
Mathematical function 5	**0.26**	0.84	0.42	1.68
Mathematical function 6	**0.19**	1.23	0.65	1.83

Notes: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.
We compare: (Proposed method) Gaussian Process using our proposed Con-Cat kernel and employing MAP estimation; (GP-MLE) Standard Gaussian Process fitted using Maximum Likelihood Estimation (MLE) with one-hot encoding for categorical variables; (GP-Category) Separate Gaussian Process models for each category, using MLE; (MPGP) Multilayer Perceptron Gaussian Process with one-hot encoding and MLE.

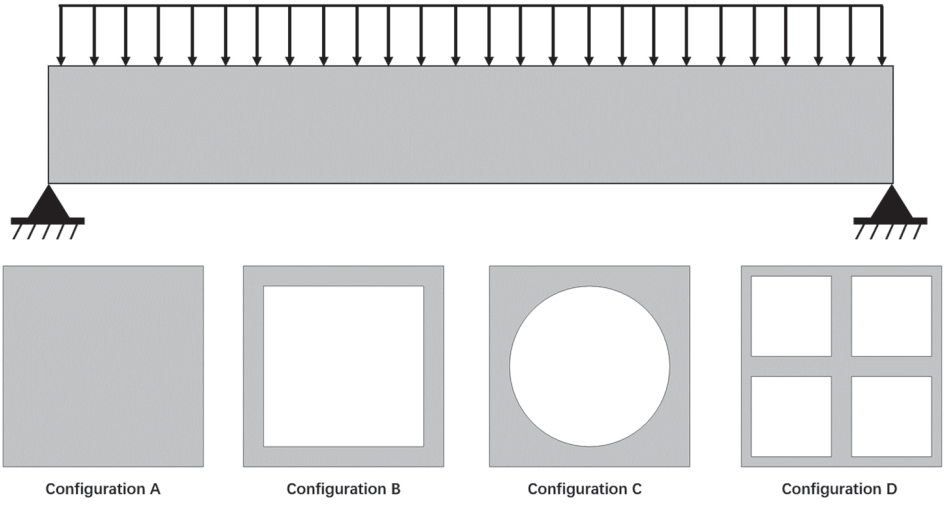

Figure 4. Constraints, loads and four different cross-sectional shapes of the simply supported beam. The upper part of the figure shows the constraints and loads of the supported beam, while the lower part displays the four different cross-sectional shapes of the supported beam.

design challenge involves adjusting shape parameters–namely, the side length of the square cross-section, the cross-section shape, and the material removal ratio–to minimize maximum deformation in Figure 5 while adhering to a mass constraint of 30 tons.

The design space encompasses three key variables as shown in Table 8: the side length of the cross-section, the material removal ratio, and the cross-section shape, the latter being a categorical variable with four configurations: solid, square-removed, circle-removed, and cross. The objective is to determine the optimal combination of these parameters to achieve the lowest possible deformation under given constraints and loads, as illustrated in the optimization model equation and detailed in the design space table .

The proposed approach considers the varied impact of the cross-section's side length and material removal ratio across different shapes, aligning with the Gaussian process kernel's capability to handle categorical variables. The optimization framework aims to find the minimal deformation configuration, constrained by a maximum mass limit, using a mathematical model that integrates these design

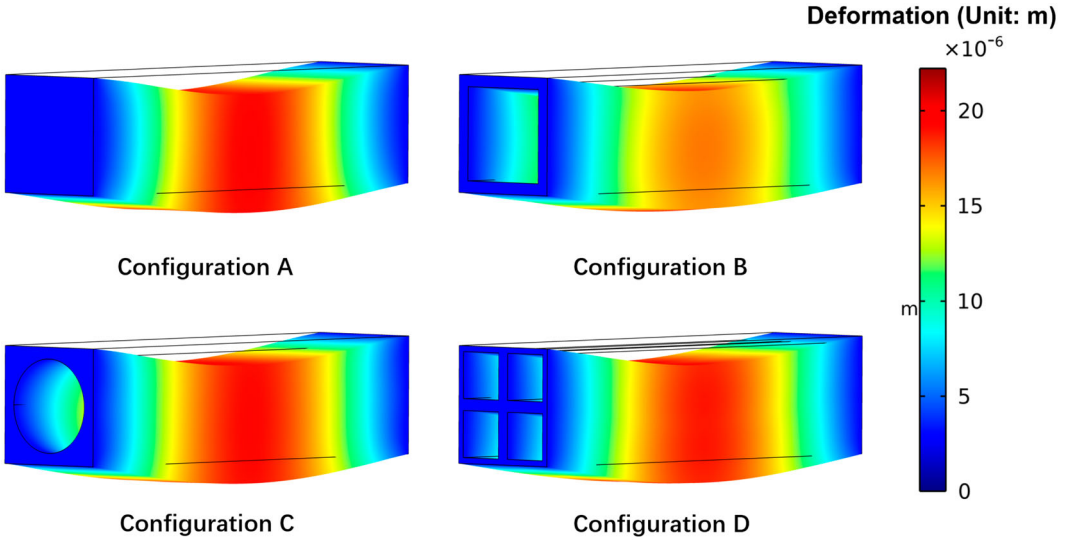

Figure 5. Deformations of simply supported beams with different cross-sectional shapes.

Table 8. Design space for the simply supported beam.

Variable	Unit	Range	Meaning of variable
BM_P1	mm	[1.0,2.0]	Side length of the outer profile
BM_P2		[0.2,0.8]	Material removal ratio
BM_P3		{*A*, *B*, *C*, *D*}	Shape configuration of the cross section

considerations as follows:

$$\text{Find} \begin{cases} \min Y = F(X) \\ X = (x_1, x_2, x_3) \end{cases}$$
$$\text{s.t.} \begin{cases} x_{ij} \le x_i \le x_{im} \ \ (i = 1, 2) \\ x_3 \in (v1, v2, v3, v4) \\ \text{mass} = M(X) \le 30t. \end{cases} \tag{35}$$

The range [LB, UB] for the simply supported beam design is established between [34.64, 39.93]. Detailed optimization outcomes from various benchmarks are depicted in Figure 6, and performance metrics are consolidated in Table 9. Notably, ALEBO's performance was hindered by its less precise handling of boundary conditions, leading to a significant number of samples falling outside the feasible domain. This characteristic adversely affected its performance, particularly in constrained optimization scenarios. In contrast, the proposed strategy excelled, achieving the best mean and median results, underscoring its effectiveness. While SAASBO showed moderate performance, both ALEBO and LHSBO lagged considerably. The proposed approach marked a pronounced improvement, showcasing enhancements ranging from 7.05% to 128.45% over the methodologies of SAASBO, LHSBO, MPGPBO , ALEBO and CoCaBO in the optimization of the simply supported beam design.

5. Design optimization for the turbine stator blade cooling structure

5.1. Turbine stator blade cooling challenges

In the high-stress environment of gas turbines, where operational conditions can exceed pressures of 40 bar and temperatures of 1000K, component durability is paramount. The turbine stator blade,

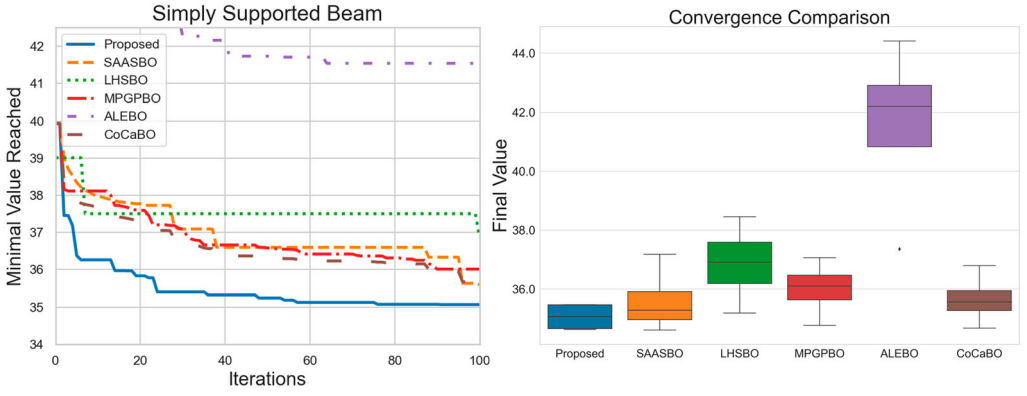

Figure 6. The optimization results of different baselines on the simply supported beam design optimization were obtained over five independent replications. For each baseline, the mean value of the best minimum found at a given iteration is depicted as a convergence curve (left). The distribution over the final approximate minimum after 100 iterations is encoded as a box plot (right), with horizontal bars corresponding to 5%, 50% and 95% quantiles. The proposed method is labelled 'Proposed'.

Table 9. Performance indicators of convergence results from different baselines of simply supported beam design optimization.

	Proposed method	SAASBO	LHSBO	MPGPBO	ALEBO	CoCaBO
Mean result	**35.07**	35.60	36.87	36.01	41.54	35.65
Median result	**35.08**	35.29	36.91	36.10	42.20	35.63
PR		7.05%	34.34%	18.63%	128.45%	10.72%

Note: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.

a critical component, must withstand intense heat, stress and aerodynamic loads. Cooling strategies, such as internal and film cooling, are essential to prevent blade degradation or melting, utilizing 'cooler' air bled from the compressor casing (G. Zhang *et al.* 2022; Nagaiah and Geiger 2014). This study focuses on optimizing the shape and configuration parameters of the turbine stator blade to enhance cooling efficiency.

Figure 7 demonstrates a typical structure of the turbine stator blade and finite element analysis result. The proposed approach addresses the design challenges by exploring four different cooling configurations (A, B, C and$D$) as depicted in Figure 8, each with unique features aimed at improving cooling efficiency. Configuration A, with a U-shaped cooling channel, serves as the baseline. Configurations B and D introduce additional cooling baffles to increase surface contact with the cooling medium, whereas Configurations C and D incorporate extra cooling holes to form a protective cooling film on the blade's surface. These designs aim to balance cooling performance with the structural integrity and weight constraints of the blade.

5.2. Design optimization with different configurations

In this article, the primary focus is on the performance of the optimization algorithm itself. All algorithms were subjected to the same experimental setup, as follows.

Given the complexity of the physical fields inside a gas turbine and considering the experimental costs, a simplified approach was employed in the simulation software. The material selection assumed that the blades and assembly parts were made of directionally solidified (DS) GTD111 nickel-based alloy with high tensile strength. The linear elastic model was set with a reference temperature of 310K and a Poisson's ratio of 0.33, comparable to other stainless steels.

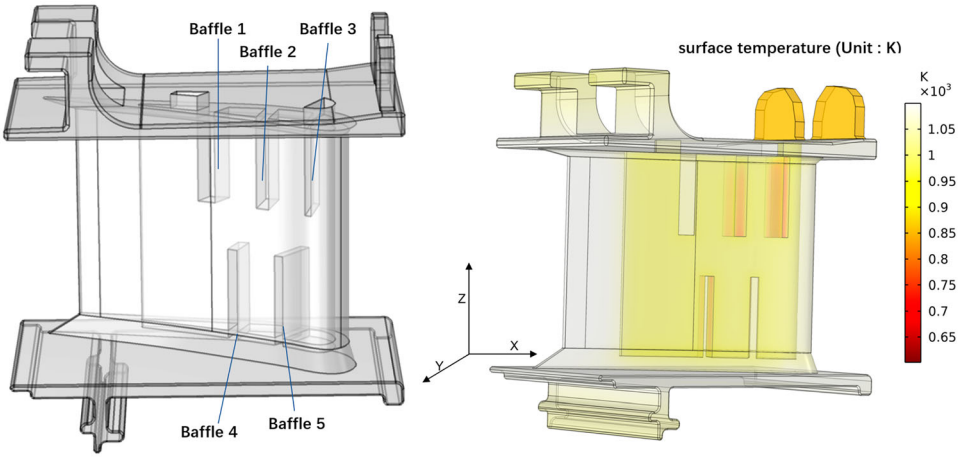

Figure 7. Structure of the turbine stator blade (left) and finite element analysis result for surface temperature (right). The colour scale on the right indicates the surface temperature in kelvins (K).

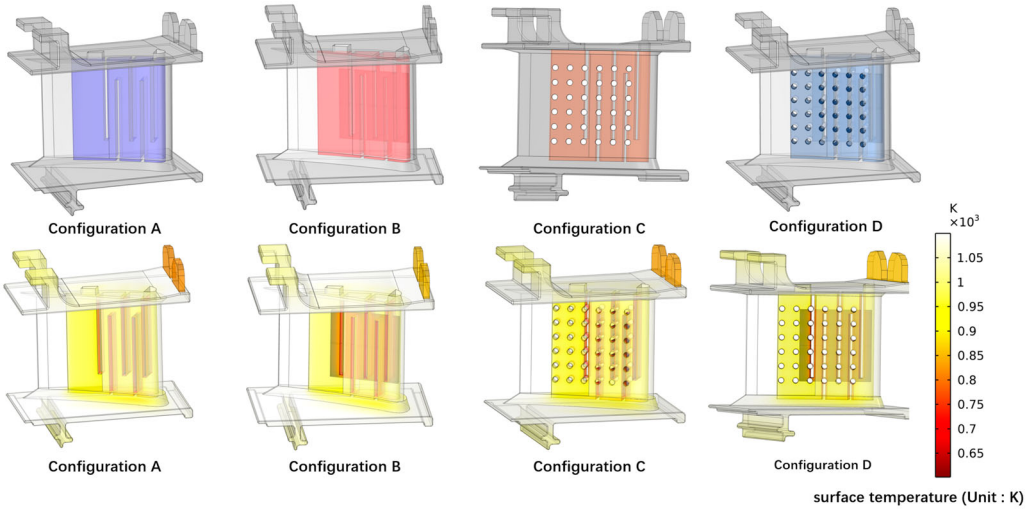

Figure 8. Structure of the turbine stator blade and finite element analysis result of surface temperature. The top row shows the four different structures of the turbine stator blade, while the bottom row displays the finite element analysis result for the surface temperature of the turbine stator blade. Configuration *A* is the original configuration. In configurations *B* and *D*, additional cooling baffles were set for the turbine blades, and in configurations *C* and *D*, additional cooling holes were set for the turbine blades. The colour scale indicates the surface temperature in kelvins (K).

The complex turbulent flows within the cooling ducts were not simulated. Instead, the average Nusselt number correlation (Bredberg 2002) was used to calculate the heat transfer coefficient. The cooling fluid was assumed to be air at a pressure of 30 bar and a temperature of 650 K. The heat flux on the stator blade surfaces was calculated using the heat transfer coefficient. The pressure and suction sides were approximated as two flat plates using the local heat transfer coefficient for external forced convection. The combustion gases were approximated as air at a pressure of 30 bar and a temperature of 1150 K, with a corresponding speed of sound of approximately 650 m/s.

The design variables for the turbine stator blade are shown in Table 10. The design space comprises 19 continuous variables influencing the serpentine cooling channel's dimensions and orientations, and one categorical variable determining the cooling pipe's configuration. The objective function

Table 10. Design space for the turbine stator blade.

Variable	Unit	Range	Meaning of variable
TS_P1	mm	[3, 6]	Width of 1st cooling baffle
TS_P2	mm	[7, 9]	Width of 2nd cooling baffle
TS_P3	mm	[10, 12]	Width of 3rd cooling baffle
TS_P4	mm	[4, 6]	Width of 4th cooling baffle
TS_P5	mm	[6, 8]	Width of 5th cooling baffle
TS_P6	mm	[10, 120]	Height of 1st cooling baffle
TS_P7	mm	[10, 120]	Height of 2nd cooling baffle
TS_P8	mm	[10, 120]	Height of 3rd cooling baffle
TS_P9	mm	[10, 120]	Height of 4th cooling baffle
TS_P10	mm	[10, 120]	Height of 5th cooling baffle
TS_P11	mm	[-30, 30]	Rotation angle of 1st baffle around the *z*-axis
TS_P12	mm	[-30, 30]	Rotation angle of 2nd baffle around the *z*-axis
TS_P13	mm	[-30, 30]	Rotation angle of 3rd baffle around the *z*-axis
TS_P14	mm	[-30, 30]	Rotation angle of 4th baffle around the *z*-axis
TS_P15	mm	[-30, 30]	Rotation angle of 5th baffle around the *z*-axis
TS_P16	mm	[4, 6]	Diameter of the cooling hole
TS_P17	mm	[50, 100]	Length of the additional cooling baffle
TS_P18	mm	[3, 6]	Width of the additional cooling baffle
TS_P19	mm	[20, 120]	Height of the additional cooling baffle
TS_P20		{*A*, *B*, *C*, *D*}	Shape configuration of the cooling pipe

aims to minimize the average surface temperature of the cooling channel, indicating enhanced cooling performance. The optimization model is constrained by the blade's maximum allowable mass, ensuring the design's feasibility for real-world application. In this study, enhancing the cooling of the turbine blade may necessitate the use of more expensive materials, thereby increasing costs. Additionally, increasing the blade's mass can affect the overall weight of the turbine, impacting the efficiency and fuel consumption of the gas turbine. The total mass of the blade was constrained not to exceed 10.80 kg in the experiments. Upon completion of the experiments, a Pareto front was obtained that included both the average temperature of the blade cooling channel and the total mass of the blade.

$$\text{Find} \begin{cases} \min Y = G(X) \\ X = (x_1, x_2, \ldots, x_{19}, x_{20}) \end{cases} \tag{36}$$

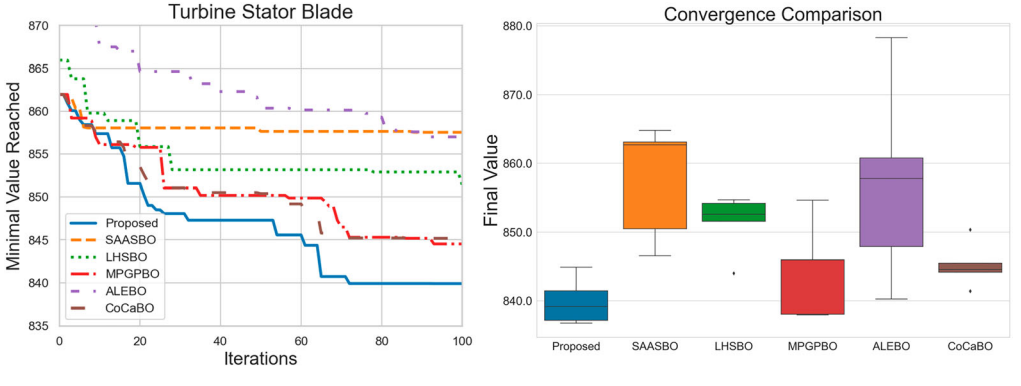

Figure 9. The optimization results for different baselines on the turbine stator blade design optimization using a DoE of 40 points over five independent replications. For each baseline, the mean value of the best minimum found at a given iteration is depicted as a convergence curve (left). The distribution over the final approximate minimum after 100 iterations is encoded as a box plot (right), with horizontal bars corresponding to the 5%, 50% and 95% quantiles. The proposed method is labelled 'Proposed'.

Table 11. Performance indicators of convergence results from different baselines of turbine stator blade design optimization.

	Proposed method	SAASBO	LHSBO	MPGPBO	ALEBO	CoCaBO
Mean result	**839.90**	857.54	851.42	844.53	857.00	845.19
Median result	**839.17**	862.69	852.61	845.95	857.80	844.55
PR		70.56%	42.79%	19.56%	61.25%	18.29%

Note: The values for the Proposed method are in bold to highlight its superior performance. Specifically, the Proposed method shows the best results in mean and median metrics compared to the other methods.

$$\text{s.t.} \begin{cases} x_{ij} \le x_i \le x_{im} (i = 1, 2, \ldots, 19) \\ x_{20} \in (v1, v2, v3, v4) \\ \text{mass} = M(X) \le 10.80\,\text{kgs}. \end{cases}$$

The optimization problem for this study is defined by Equation (36). The optimization results for different baselines are shown in Figure 9 and the performance metrics of the convergence results are shown in Table 11. The proposed strategy exhibits the best performance as both its mean and median results are optimal. Notably, the proposed method achieves the lowest average temperature, signifying the highest cooling efficiency among tested configurations. The results, as summarized in the performance indicators table, highlight the proposed strategy's significant improvements over conventional methods such as SAASBO, LHSBO, MPGPBO , ALEBO and CoCaBO, with improvements ranging from 18.29% to 70.56%.

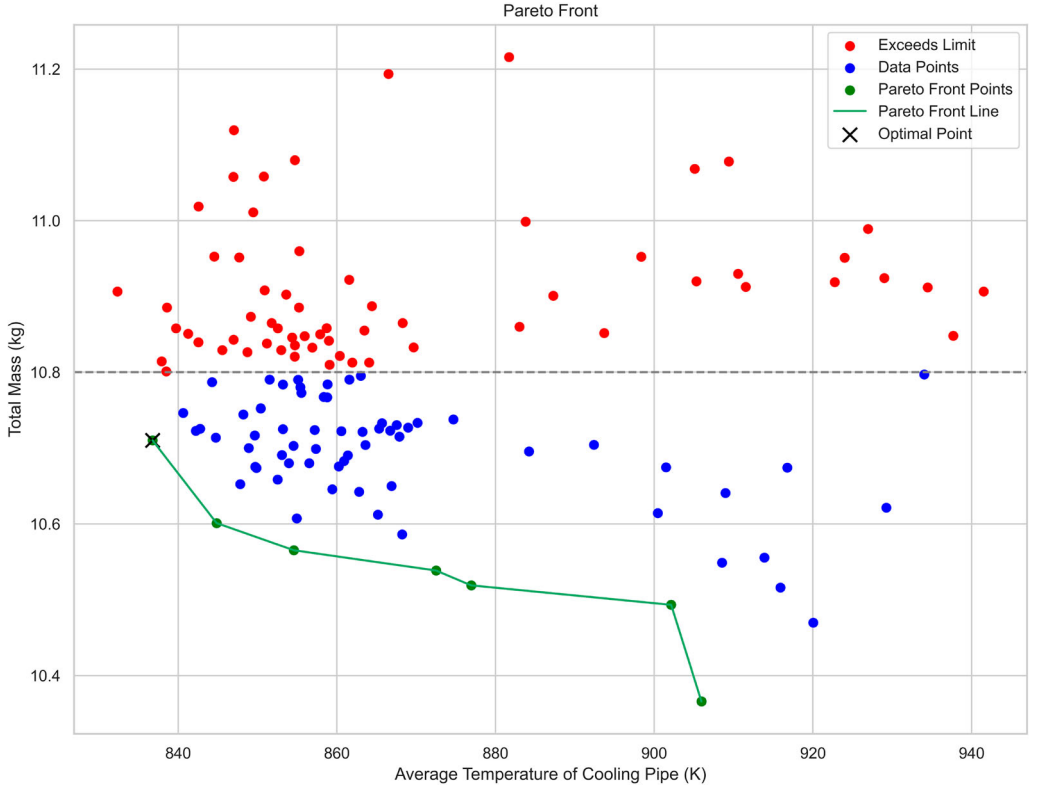

Figure 10. Pareto front obtained by the proposed method, showing the trade-off between the average temperature of the blade cooling channel and the total mass of the blade. The optimal solution, which meets the weight constraint and achieves the lowest average temperature, is indicated on the Pareto front.

The Pareto front obtained by the proposed method, which includes both the average temperature of the blade cooling channel and the total mass of the blade, is shown in Figure 10. The sample point with the lowest average temperature of the blade cooling channel that meets the weight constraint was selected as the optimal solution. The optimal solution configuration is identified as *D*, with an average temperature of 836.77K and a total mass of 10.71 kg.

Furthermore, an independent complete simulation of the optimal solution was conducted using the average Nusselt number correlation to calculate the heat transfer coefficient, as a substitute for complex turbulent models within the cooling ducts. Additionally, the predictive accuracy of the surrogate model used was verified. These results, as shown in Figure 11, confirm the validity of the proposed approach.

This engineering application of the proposed Bayesian optimization method to turbine stator blade cooling structure design exemplifies the method's potential in addressing complex, real-world engineering challenges. By systematically optimizing the cooling structure's design parameters, the proposed method not only enhances cooling efficiency but also ensures compliance with critical weight constraints, showcasing a significant advancement in turbine blade design optimization.

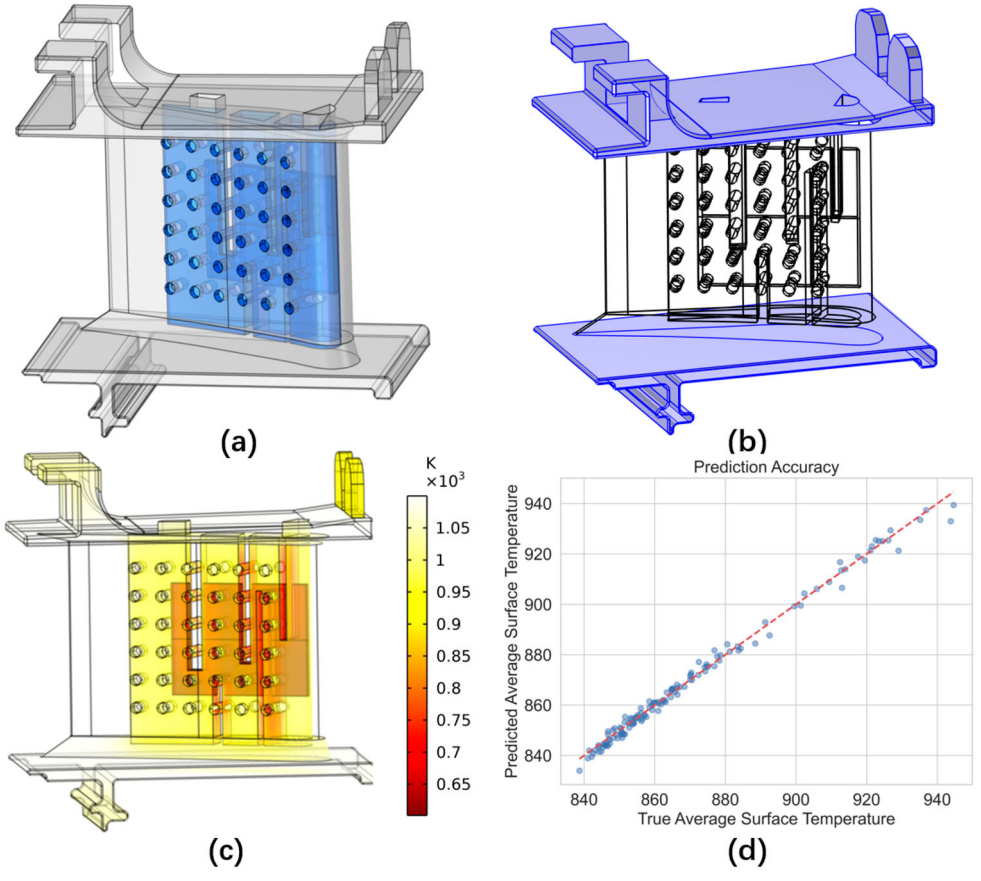

Figure 11. (a) **Optimal cooling channel configuration**. The layout of the cooling channel designed for maximum cooling efficiency, meeting the highest cooling efficiency and mass constraints, with the cooling channel highlighted. (b) **Heat exchange boundary**. The boundary where heat exchange occurs between the turbine stator blade and the cooling air, highlighted. (c) **Surface temperature distribution**. The surface temperature distribution of the turbine stator blade under the optimal cooling channel configuration, with an average surface temperature of the cooling channel at 836.77K. (d) **Prediction accuracy of the optimal surrogate model**. The prediction accuracy of the optimal surrogate model, GP with Con-Cat kernel, using 400 training data points to predict 100 test data points, with a predicted normalized RMSE of 0.021.

6. Conclusions

This article tackles the complex challenge of designing intricate structures, emphasizing the critical importance of choosing among various geometric configurations. A novel sampling-efficient optimization strategy is proposed that adeptly navigates the complexities of categorical configurations and high-dimensional continuous variables in a supervised manner. The proposed methodology leverages the Con-Cat kernel, which effectively bridges the gap between different configurations, enhancing the Gaussian process's capability for surrogate modelling. The proposed Con-Cat kernel represents a generalization of standard kernels, achieving equivalence in situations that involve only continuous variables. By employing half-Cauchy distributions as priors for Gaussian process modelling, critical design variables are efficiently pinpointed without the need for additional trainable parameters. This method does not add any extra trainable parameters, making it particularly beneficial for sparse datasets and ensuring robust and accurate modelling.

Through rigorous numerical evaluations across three mathematical functions and a simulation-based design problem—the design optimization of a simply supported beam—the proposed strategy has proven superior, demonstrating improvements ranging from 7.49% to 85.08% over established methods such as StandardBO, SAASBO, LHSBO, MPGPBO , ALEBO, CoCaBO and BanditBO. Furthermore, when applied to the optimization of turbine blade cooling structures, the proposed method showcased significant enhancements in cooling efficiency, outperforming alternative strategies with improvements between 18.29% and 70.56%. These findings not only advance the theoretical understanding of Bayesian optimization in handling complex design spaces but also demonstrate the practical efficacy of the proposed method in real-world engineering design challenges.

Disclosure statement

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

Funding

This work has been supported by the National Key Research and Development Program of China [2022YFE0196400]; the National Natural Science Foundation of China [52305288]; the specialized research projects of Huanjiang Laboratory, the Zhejiang Provincial Natural Science Foundation of China [Grant No. LDT23E05013E05]; the 'Leading Goose' R&D Programme of Zhejiang [2023C01166]; the 'Pioneer' R&D Programme of Zhejiang [2023C01060].

Data availability statement

The data that support the findings of this study are available from the authors upon reasonable request.

ORCID


Weifei Hu http://orcid.org/0000-0002-1571-583X

References

Bredberg, Jonas. 2002. "Turbulence Modelling for Internal Cooling of Gas-Turbine Blades." PhD diss., Chalmers University of Technology Göteborg, Sweden. https://www.tfd.chalmers.se/lada/postscript_files/jonas_PhD.pdf.

Cavazzuti, Marco. 2013. "Design of Experiments." In *Optimization Methods: From Theory to Design—Scientific and Technological Aspects in Mechanics*, 13–42. Heidelberg, Germany: Springer. http://dx.doi.org/10.1007/978-3-642-31187-1.

Chen, Zexun, and Bo Wang. 2018. "How Priors of Initial Hyperparameters Affect Gaussian Process Regression Models." *Neurocomputing* 275:1702–1710. https://doi.org/10.1016/j.neucom.2017.10.028.

Daxberger, Erik, Anastasia Makarova, Matteo Turchetta, and Andreas Krause. 2021. "Mixed-Variable Bayesian Optimization." In *Proceedings of the 29th International Conference on Artificial Intelligence*, 2633–2639. Darmstadt, Germany: IJCAI. http://dx.doi.org/10.24963/ijcai.2020/365.

Deshwal, Aryan, Syrine Belakaria, and Janardhan Rao Doppa. 2021. "Bayesian Optimization over Hybrid Spaces." In *International Conference on Machine Learning*. PMLR: *Proceedings of Machine Learning Research* 139:2632–2643.

Eriksson, David, and Martin Jankowiak. 2021. "High-Dimensional Bayesian Optimization with Sparse Axis-Aligned Subspaces." In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*. PMLR: *Proceedings of Machine Learning Research* 161:493–503. https://proceedings.mlr.press/v161/eriksson21a.html.

Eriksson, David, Michael Pearce, Jacob Gardner, Ryan D. Turner, and Matthias Poloczek. 2019. "Scalable Global Optimization via Local Bayesian Optimization." In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 5496–5507. New York: Association for Computing Machinery (ACM). https://dl.acm.org/doi/10.5555/3454287.3454780.

Foumani, Zahra Zanjani, Mehdi Shishehbor, Amin Yousefpour, and Ramin Bostanabad. 2023. "Multi-Fidelity Cost-Aware Bayesian Optimization." *Computer Methods in Applied Mechanics and Engineering* 407:115937. https://doi.org/10.1016/j.cma.2023.115937.

Helton, Jon C., and Freddie Joe Davis. 2003. "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems." *Reliability Engineering & System Safety* 81 (1): 23–69. https://doi.org/10.1016/S0951-8320(03)00058-9.

Iyer, Akshay, Yichi Zhang, Aditya Prasad, Siyu Tao, Yixing Wang, Linda Schadler, L. Catherine Brinson, *et al.* 2019. "Data-Centric Mixed-Variable Bayesian Optimization for Materials Design." In *International Design Engineering Technical Conference & Computers and Information in Engineering*, Vol. 59186, V02AT03A066. American Society of Mechanical Engineers (ASME). https://doi.org/10.1115/DETC2019-98222.

Kaveh, A., P. Rahmani, and A. Dadras Eslamlou. 2022. "An Efficient Hybrid Approach Based on Harris Hawks Optimization and Imperialist Competitive Algorithm for Structural Optimization." *Engineering with Computers* 38 (S2): 1555–1583. https://doi.org/10.1007/s00366-020-01258-7.

Kirschner, Johannes, Mojmir Mutny, Nicole Hiller, Rasmus Ischebeck, and Andreas Krause. 2019. "Adaptive and Safe Bayesian Optimization in High Dimensions via One-Dimensional Subspaces." In *Proceedings of the 36th International Conference on Machine Learning*. PMLR: *Proceedings of Machine Learning Research* 97:3429–3438. http://proceedings.mlr.press/v97/kirschner19a/kirschner19a.pdf.

Letham, Ben, Roberto Calandra, Akshara Rai, and Eytan Bakshy. 2020. "Re-Examining Linear Embeddings for High-Dimensional Bayesian Optimization." *Advances in Neural Information Processing Systems* 33:1546–1558.

Luo, Jiaqi, Zhen Fu, Yiming Zhang, Wenhao Fu, and Jianjun Chen. 2023. "Aerodynamic Optimization of a Transonic Fan Rotor by Blade Sweeping Using Adaptive Gaussian Process." *Aerospace Science and Technology* 137:108255. https://doi.org/10.1016/j.ast.2023.108255.

Malu, Mohit, Gautam Dasarathy, and Andreas Spanias. 2021. "Bayesian Optimization in High-Dimensional Spaces: A Brief Survey." In *Proceedings of the 2021 12th IEEE International Conference on Information, Intelligence, Systems & Applications (IISA)*, 1–8. Piscataway, NJ: IEEE.

Nagaiah, Narasimha R., and Christopher D. Geiger. 2014. "Evolutionary Numerical Simulation Approach for Design Optimization of Gas Turbine Blade Cooling Channels." *International Journal for Simulation and Multidisciplinary Design Optimization* 5:A22. https://doi.org/10.1051/smdo/2014001.

Nguyen, Dang, Sunil Gupta, Santu Rana, Alistair Shilton, and Svetha Venkatesh. 2020. "Bayesian Optimization for Categorical and Category-Specific Continuous Inputs." In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34, 5256–5263. Reston, VA: American Institute of Aeronautics and Astronautics (AIAA).

Qiu, Na, Zhuoqun Yu, Depei Wang, Mingwei Xiao, Yiming Zhang, Nam H. Kim, and Jianguang Fang. 2024. "Bayesian Optimization of Origami Multi-Cell Tubes for Energy Absorption Considering Mixed Categorical–Continuous Variables." *Thin-Walled Structures* 199:111799. https://doi.org/10.1016/j.tws.2024.111799.

Rodríguez, Pau, Miguel A. Bautista, Jordi Gonzalez, and Sergio Escalera. 2018. "Beyond One-Hot Encoding: Lower Dimensional Target Embedding." *Image and Vision Computing* 75:21–31. https://doi.org/10.1016/j.imavis.2018.04.004.

Ru, Binxin, Ahsan Alvi, Vu Nguyen, Michael A. Osborne, and Stephen Roberts. 2020. "Bayesian Optimisation over Multiple Continuous and Categorical Inputs." In *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, PMLR: *Proceedings of Machine Learning Research* 119: 8276–8285. https://api.semanticscholar.org/CorpusID:195316915.

Sabater, Christian, Olivier Le Maitre, Pietro Marco Congedo, and Stefan Görtz. 2021. "A Bayesian Approach for Quantile Optimization Problems with High-Dimensional Uncertainty Sources." *Computer Methods in Applied Mechanics and Engineering* 376:113632. https://doi.org/10.1016/j.cma.2020.113632.

Saves, Paul, Nathalie Bartoli, Youssef Diouane, Thierry Lefebvre, Joseph Morlier, Christophe David, Eric Nguyen Van, and Sébastien Defoort. 2022. "Correction: Multidisciplinary Design Optimization with Mixed Categorical Variables for Aircraft Design." In *AIAA Scitech 2022 Forum*, 0082–c1. Reston, VA: American Institute of Aeronautics and Astronautics (AIAA).

Shahriari, Bobak, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. 2015. "Taking the Human out of the Loop: A Review of Bayesian Optimization." *Proceedings of the IEEE* 104 (1): 148–175. https://doi.org/10.1109/JPROC.2015.2494218.

Sheikh, Haris Moazam, Tess A. Callan, Kealan J. Hennessy, and Philip S. Marcus. 2022. "Optimization of the Shape of a Hydrokinetic Turbine's Draft Tube and Hub Assembly Using Design-by-Morphing with Bayesian Optimization." *Computer Methods in Applied Mechanics and Engineering* 401:115654. https://doi.org/10.1016/j.cma.2022.115654.

Shields, Michael D., and Jiaxin Zhang. 2016. "The Generalization of Latin Hypercube Sampling." *Reliability Engineering & System Safety* 148:96–108. https://doi.org/10.1016/j.ress.2015.12.002.

Snoek, Jasper, Hugo Larochelle, and Ryan P. Adams. 2012. "Practical Bayesian Optimization of Machine Learning Algorithms." In *Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS'12)*, Vol. 2, 2951–2959. New York: Association for Computing Machinery (ACM). https://dl.acm.org/doi/10.5555/2999325.2999464.

Wang, Xilu, Yaochu Jin, Sebastian Schmitt, and Markus Olhofer. 2023. "Recent Advances in Bayesian Optimization." *ACM Computing Surveys* 55 (13s): 1–36.

Wang, Feng, Heng Zhang, and Aimin Zhou. 2021. "A Particle Swarm Optimization Algorithm for Mixed-Variable Optimization Problems." *Swarm and Evolutionary Computation* 60:100808. https://doi.org/10.1016/j.swevo.2020.100808.

Wilson, Andrew Gordon, Zhiting Hu, Ruslan Salakhutdinov, and Eric P. Xing. 2016. "Deep Kernel Learning." In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*. PMLR: *Proceedings of Machine Learning Research* 51:370–378. http://proceedings.mlr.press/v51/wilson16.pdf.

Yang, Yibo, Georgios Kissas, and Paris Perdikaris. 2022. "Scalable Uncertainty Quantification for Deep Operator Networks Using Randomized Priors." *Computer Methods in Applied Mechanics and Engineering* 399:115399. https://doi.org/10.1016/j.cma.2022.115399.

Zhang, Yichi, Daniel W. Apley, and Wei Chen. 2020. "Bayesian Optimization for Materials Design with Mixed Quantitative and Qualitative Variables." *Scientific Reports* 10 (1): 4924. https://doi.org/10.1038/s41598-020-60652-9.

Zhang, Guohua, Zhu Rui, Xie Gongnan, Li Shulei, and Bengt Sunden. 2022. "Optimization of Cooling Structures in Gas Turbines: A Review." *Chinese Journal of Aeronautics* 35 (6): 18–46. https://doi.org/10.1016/j.cja.2021.08.029.