

Multi-Fidelity Design Framework Integrating Compositional Kernels to Facilitate Early-Stage Design Exploration of Complex Systems

Charisi, N.D.; Hopman, J.J.; Kana, A.A.

DOI

[10.1115/1.4065890](https://doi.org/10.1115/1.4065890)

Publication date

2025

Document Version

Final published version

Published in

Journal of Mechanical Design

Citation (APA)

Charisi, N. D., Hopman, J. J., & Kana, A. A. (2025). Multi-Fidelity Design Framework Integrating Compositional Kernels to Facilitate Early-Stage Design Exploration of Complex Systems. *Journal of Mechanical Design*, 147, Article 011701. <https://doi.org/10.1115/1.4065890>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



Multi-Fidelity Design Framework Integrating Compositional Kernels to Facilitate Early-Stage Design Exploration of Complex Systems

Nikoleta Dimitra Charisi¹

Maritime and Transport Technology Department,
Technical University of Delft,
2628 CA Delft, The Netherlands
e-mail: N.D.Charisi@tudelft.nl

Hans Hopman

Professor
Maritime and Transport Technology Department,
Technical University of Delft,
2628 CA Delft, The Netherlands
e-mail: J.J.Hopman@tudelft.nl

Austin A. Kana

Associate Professor
Maritime and Transport Technology Department,
Technical University of Delft,
2628 CA Delft, The Netherlands
e-mail: A.A.Kana@tudelft.nl

Early-stage design of complex systems is considered by many to be one of the most critical design phases because that is where many of the major decisions are made. The design process typically starts with low-fidelity tools, such as simplified models and reference data, but these prove insufficient for novel designs, necessitating the introduction of high-fidelity tools. This challenge can be tackled through the incorporation of multifidelity models. The application of multifidelity (MF) models in the context of design optimization problems represents a developing area of research. This study proposes incorporating compositional kernels into the autoregressive scheme (AR1) of multifidelity Gaussian processes, aiming to enhance the predictive accuracy and reduce uncertainty in design space estimation. The effectiveness of this method is assessed by applying it to five benchmark problems and a simplified design scenario of a cantilever beam. The results demonstrate significant improvement in the prediction accuracy and a reduction in the prediction uncertainty. Additionally, the article offers a critical reflection on scaling up the method and its applicability in early-stage design of complex engineering systems, providing insights into its practical implementation and potential benefits. [DOI: 10.1115/1.4065890]

Keywords: conceptual design, data-driven design, design space exploration, multi-fidelity Gaussian processes, compositional kernels

1 Introduction

Early-stage design of complex engineering systems is critical since most of the major design decisions are being made at this stage while engineers have minimal knowledge about the design. The importance of early design stages has been recognized in various engineering fields, including ship [1] and aircraft design [2]. More specifically, design decisions that dictate the vessel's overall configuration reduce design freedom and commit a significant portion of the overall cost. Therefore, engineers must conduct comprehensive design space exploration to identify design trends and trade-offs, leading to informed design decisions.

Throughout the design process, engineers employ a range of analysis methods with varying fidelities to evaluate different designs. At opposite ends of the spectrum, high-fidelity (HF) methods, often based on first principles, provide highly accurate results but entail significant computational expenses, while low-fidelity (LF) methods yield less accurate results while requiring

lower computational costs. Traditionally, engineers have used LF methods and tools to explore the broad multidimensional design space of complex systems [2]. These methods include simplified physical models, empirical and semi-empirical methods [3], and trends based on historical data [4]. Based on engineering experience, these methods are suitable for conventional designs for which we have a lot of gained knowledge.

However, nowadays, there is a drive toward innovative designs driven by the need to address challenges such as the integration of autonomous systems [5], enhancing safety [6], and promoting sustainability [7]. This necessitates the design of systems with capabilities that surpass our current state-of-the-art designs. Some examples of novel hull designs in naval architecture are the tumblehome wave-piercing hull [8] employed in the Zumwalt-class destroyer, the aluminum trimaran design [9] used for the USS Independence frigate, the AXE bow [10] used in various yacht designs, and the asymmetric hull configuration found in the Baltika icebreaker [11]. In aerospace, the flying-V [12] represents an innovative aircraft design that integrates the fuselage into the wing structure, aiming to enhance its performance.

Incorporating new features into systems introduces added uncertainty, primarily because engineers lack prior experience with these unique systems [13]. This introduced uncertainty needs to be quantified and taken into account in order to support decision-

¹Corresponding author.

Contributed by the Design Automation Committee of ASME for publication in the JOURNAL OF MECHANICAL DESIGN. Manuscript received January 30, 2024; final manuscript received June 13, 2024; published online July 23, 2024. Assoc. Editor: Douglas Allaire.

making [14]. Thus, in order to mitigate this heightened uncertainty, it becomes important to introduce HF analysis at an earlier stage in the design process. This is essential because relying solely on LF models and tools to examine the design space may result in designs that are not feasible or suboptimal when the underlying physics is not appropriately captured [15]. Additionally, in the case of ship design, engineers rely on established rules and guidelines prescribed by classification societies to guide their design practices. However, when dealing with novel ships featuring unconventional shapes and sizes, blindly following the class society formulations proves insufficient [16], as these guidelines are primarily based on conventional ship knowledge. A similar challenge is present across different design domains. Currently, advanced HF models are able to capture the underlying physics that dictate a design's performance, yet they require a significant amount of computational resources, human input, and extended lead times [17,18]. However, it is unrealistic to evaluate a sufficient number of designs using expensive HF methods in order to construct an approximation of the design space. Therefore, a significant challenge in the optimization of vessel design is the efficient increase of modeling accuracy required to capture the crucial physics that restricts or allows for a specific concept [19].

To address these issues, a promising approach is to construct multifidelity (MF) models, which are defined as models combining various fidelities [20]. According to Ref. [19], models are considered multifidelity when there is synergistic use of different mathematical descriptions (i.e., different physics typically represented by different governing equations, boundary conditions, or parametric attributions) in the analysis or design procedure. The goal of MF models is to achieve computational efficiency through the use of LF models while maintaining precision through the use of the HF model [21]. MF models are state-of-the-art for solving complex problems. For a comprehensive understanding of the various MF techniques, an extensive review of the various MF techniques can be found in Refs. [20,21]. MF models have demonstrated notable achievements across various engineering domains such as engineering design and analysis, and applied mathematics. They have been used to tackle diverse problems, encompassing solving partial differential equations [22], aiding in complex analyses like estimating the wave-induced vertical bending moment on ships [23,24], and supporting design applications [25].

Several state-of-the-art methods for design applications have been extended to an MF scheme. Examples of such methods include Monte Carlo (MC), Gaussian processes (GPs), and neural networks (NNs). For example, Ng and Willcox [26] proposed an MF estimator based on the control variate MC method. Diverse methods exist in the literature for constructing MF NN architectures. For instance, Meng and Karniadakis introduced a composite NN leveraging MF data. In their approach, the first NN approximates the LF data, while the second and third NNs model the correlation between the low- and high-fidelity data [27]. Examples of different approaches for building MF-NN architectures can be found in Refs. [28,29]. According to the findings reported in Guo et al. [29], the co-Kriging model demonstrated superior performance compared to the MF-NN architectures in the case of the Forrester function. Conversely, in the case of more complex functions such as the Jump Forrester function, the opposite trend was observed [29]. To estimate the uncertainty associated with the prediction, Meng et al. [30] introduced a MF Bayesian NN framework. Within this framework, uncertainty was estimated by predicting the uncertainty of learning the correlation between the low- and high-fidelity data. GPs have found extensive application in design problems. Their primary advantages lie in their ability to provide uncertainty quantification pertaining to predictions, which can be effectively used in the context of design optimization. Another advantage is their effectiveness in addressing problems characterized by small data regimes.

Despite the ongoing developments in MF models, there are still significant areas in design applications that remain unexplored. For example, for multidisciplinary design problems, Mainini et al.

[15] argue that there is no mathematical framework that is capable of determining (1) which design disciplines, (2) the degree of coupling for analysis tools, (3) the level of accuracy necessary to capture the crucial physics of a specific design, (4) where the data is best collected, and (5) how to make optimal design decisions with limited computational resources. Furthermore, Peherstorfer et al. [21] emphasize that for design frameworks, it is crucial to construct frameworks that do not solely focus on models but include additional information sources, so that decision-makers can effectively utilize a wide range of available information.

In design applications, a primary challenge lies in the necessity for a larger HF dataset to attain precise predictions, particularly for complex and higher-dimensional problems. Expanding the HF dataset poses difficulties since each data point is a product of computationally expensive analyses or, in some cases, physical experiments. Consequently, the acquisition of a substantial HF dataset is constrained by the limited computational budget available. To address this challenge, the present work proposes the integration of compositional kernels in a design framework based on the autoregressive scheme of MF GPs. The main idea is that a more accurate approximation of the design space can be attained by utilizing knowledge about the underlying structure of the design space revealed by the compositional kernels. The goal is to build a framework where fewer HF evaluations are necessary to create an accurate MF model, resulting in a reduction in computational cost. The proposed method has been applied to five benchmark problems, as well as to a cantilever beam design problem as a first step toward its application to early-stage design of complex engineering systems such as naval vessels.

Section 2 provides an overview of relevant work on the exploration of design space structure and uncertainty quantification for design applications. Section 3 provides the technical background of the proposed method. Section 4 examines the results from the conducted case studies. Finally, Sec. 5 outlines the conclusions derived from the research and offers recommendations for future research.

2 Relevant Works

The article addresses early-stage design exploration of complex systems by introducing an MF-GP-based method. In this section, the fundamental concepts of the topic are discussed individually, and relevant studies are presented. These concepts include exploration of design space's structure, uncertainty quantification in design exploration, and GPs.

2.1 Design Space: Exploration of Its Structure. Early-stage design of complex systems deals with multidimensional design spaces, and this forms a significant challenge. In practical applications, the design space often involves a large number of design variables. Therefore, the scalability of modeling methods becomes a crucial consideration. For instance, the design variables can range from a few dozen, as seen in the hydrostructural optimization of hydrofoils with a 17-dimensional design space [31], to thousands of variables in the case of an aerostructural optimization benchmark problem for commercial transport aircraft [32]. Another important challenge to address during this design stage is the quantification of uncertainty (UQ). As highlighted by Ref. [14], "UQ creates value to the extent that it holds the possibility of changing a decision that would otherwise be made differently."

The literature has examined numerous approaches aimed at facilitating design exploration. For example, Di Fiore et al. [33] proposed incorporating both information extracted from data and domain knowledge to facilitate the conceptual design of re-entry vehicles. Furthermore, Singh and Willcox [34] developed a framework grounded in Bayesian statistics and decision theory. This framework integrates information from different stages of a product's lifecycle to enhance decision-making in the design process. The method proposed in the present article is founded on the

premise that the design space under investigation possesses a specific structure, which can be uncovered and leveraged to enhance the efficiency and effectiveness of design exploration.

The idea of uncovering the patterns within the design space has been studied by Melati et al. [35]. The researchers proposed a machine learning-based approach rooted in pattern recognition to effectively map and characterize the multidimensional design space of nanophotonic components. Through pattern recognition techniques, the authors successfully unveiled relationships among an initial sparse set of optimized designs, thereby reducing the number of characterized variables.

In the context of GPs, the covariance matrix via the kernel function can be used to define patterns in the design space. The efficacy of employing an appropriate kernel function for Bayesian optimization was demonstrated by Moss et al. [36]. In their work, the authors introduced a Bayesian optimization method for raw strings that seamlessly incorporates a string kernel, showcasing the power and effectiveness of this approach. In addition, the study conducted by Palar et al. [37] explores the potential of composite kernel learning and model selection in enhancing the accuracy of bi-fidelity GPs.

While the main focus of the current work aligns with the research paper of Palar et al. [37], the approach of defining the kernel functions differs. The approach of Palar et al. [37] is to build the compositional kernels as a weighted sum of basis kernels, and the weights are treated as hyperparameters. In contrast, the present study proposes an optimization routine where the kernel functions of the different fidelities are sequentially built. Thus, the kernel function for the i th fidelity is built based on the kernel function of the lower fidelity model $i - 1$. The proposed method can be extended to an s th fidelity problem setup.

2.2 Uncertainty Quantification in Design Exploration. UQ is a widely discussed topic within the realm of design applications [38]. Defining uncertainty poses challenges since it relates to a lack of knowledge. Uncertainty is commonly categorized into two types: aleatory and epistemic [39]. Aleatory uncertainty pertains to natural randomness that is inherent in the system such as the outcomes of tossing dice and drawing cards, and thus, it cannot be reduced or eliminated. On the other hand, epistemic uncertainty arises from a lack of knowledge or information. This type of uncertainty can be reduced as one gains a better understanding of the problem through further investigation and learning. However, there are arguments asserting that uncertainty can only be epistemic. As described in Ref. [40], uncertainty is an inherent aspect of the human brain and is not an inherent property of the external world. Probability, serving as a measure of uncertainty, reflects an individual's state of mind rather than an absolute state of affairs. In this article, we follow the common categorization to aleatory and epistemic uncertainty.

A typical example of aleatory uncertainty in this context of hydrodynamic analysis is the probabilistic formulation of ocean waves. Due to the inherent randomness and variability of wave conditions, a probabilistic approach is essential to adequately address the associated uncertainties in the hydrodynamic analysis. On the other hand, Mavris et al. [2] discuss several examples of epistemic uncertainties that are pertinent to early-stage design, including the treatment of assumptions, ambiguous requirements, the fidelity of different codes, economic uncertainties, and technological risks. Furthermore, according to Collette [41], the key epistemic uncertainties for ship structures involve operational profiles and behavior, model uncertainty, and the influence of the human engineer. Hence, within the domain of early-stage design for complex engineering systems, engineers confront both aleatory and epistemic uncertainties. In certain instances, as highlighted by Ref. [41] in the context of ship structures, numerous studies tend to simplify or overlook epistemic uncertainties.

UQ can be used to improve early-stage design exploration of novel and complex systems. First, UQ can be used to make more informed and optimal design decisions [38,42]. Second, when dealing with innovative concepts, there is an inherent introduction

of additional uncertainty into the design exploration problem. The additional uncertainty arises due to the inherent lack of knowledge concerning the performance of such systems. A real-world example in industries like automotive and aerospace involves the use of prototyping to acquire additional insights into the performance of new engineering systems. It is noteworthy to emphasize that, particularly in the domain of ship design, the large physical dimensions, the complexity, and the fact that ships are not being built in large series render the construction of full-scale prototypes unfeasible [1]. Therefore, it becomes important to consider and account for this uncertainty in order to effectively navigate the design space and make reliable design decisions.

A viable mathematical framework for addressing epistemic uncertainty lies within the family of Bayesian methods. Bayesian methods possess an inherent capacity to quantify uncertainty arising from our limited knowledge. This is because Bayesian statistics quantify the degree of belief in the truth of a particular proposition, rather than the frequency of event occurrence. For a more comprehensive discussion on the distinction between frequentist and Bayesian statistics, readers are referred to Ref. [39].

2.3 Gaussian Processes. The present article focuses on GPs, a subset of Bayesian methods. In general, a GP is a collection of random variables such that any subset of these variables is jointly Gaussian [43]. The MF schemes of GPs incorporate data obtained from various fidelities. One of the schemes is the linear autoregressive scheme (AR1) proposed by Kennedy and O'Hagan [44]. In addition to this, two nonlinear schemes have been introduced: the nonlinear autoregressive Gaussian process (NARGP) proposed by Perdikaris et al. [45] and the deep Gaussian processes (deep GPs) proposed by Damianou and Lawrence [46]. GPs and MF-GPs have demonstrated their effectiveness in addressing design optimization problems across different engineering fields such as aircraft design [47], ship design [48], and materials design [49]. Their popularity widely comes from the fact that they are well suited for small data regimes [50], which aligns them to the problem of early-stage exploration of novel systems where there is inherently limited data available. In contrast to AR1, which assumes a linear dependency between the fidelities, NARGP, and deep GPs are capable of capturing more complex nonlinear dependencies among the fidelities. A general trend, based on aerospace-related engineering problems as shown in [51], is that linear schemes exhibit superior performance in scenarios with limited data compared to nonlinear schemes, which require a larger amount of data for effective training. The present research targets engineering problems in the small data regime; thus, the AR1 scheme was selected as the basis for this work.

While design optimization has seen successful applications, ongoing research efforts are necessary to fully harness the potential of GPs for early-stage exploration purposes. This study investigates the concept of exploring the structure of the design space with the aim of leveraging this information to enhance prediction accuracy and reduce uncertainty. Such an approach holds potential for design applications in domains such as ship and aircraft design, where HF analyses, including computational fluid dynamics computations and experiments, can be significantly (computationally) expensive. Consequently, reducing the number of HF simulations would yield benefits for tackling such problems. The article builds upon the work of Charisi et al. [52] by expanding the model beyond the bi-fidelity case. This is achieved by developing compositional kernels for each fidelity level sequentially, starting from lower to higher fidelity. Furthermore, extensive analysis of the model's performance is conducted across multiple case studies involving analytical benchmark problems with varying attributes, enhancing our understanding of its applicability to diverse design problems.

3 Methods

This section contains the technical details of the proposed method in Sec. 3.1, which is composed of two primary components: (1) the

MF-GPs and (2) the compositional kernels. The mathematical formulation of GPs and MF-GPs is provided in Sec. 3.2, while the optimization process for the compositional kernels is described in Sec. 3.3.

3.1 Proposed Method. The authors propose integrating compositional kernels to the linear autoregressive scheme (AR1) to facilitate design exploration. The integration of the compositional kernels aims to capture the shape of the underlying HF design space with the goal of making improved predictions with less HF analysis data.

The core concept revolves around seeking the optimal compositional kernel, comprising kernels that effectively capture distinct characteristics of the design space, such as linear or periodic patterns. Constructing these compositional kernels involves solving a discrete optimization problem. The optimization of the compositional kernel is guided by the analysis data utilized as the training set for the MF-GP.

Let us assume that the design problem involves models with fidelities ranging from 1 to s , where fidelity 1 represents the lowest model and fidelity s represents the highest fidelity model. The first step is to build a compositional kernel for the data of the lowest fidelity (fidelity 1) based on a single fidelity GP model (technical details in Sec. 3.2). For each fidelity i ranging from 2 to s , a compositional kernel is built based on the bi-fidelity GP model (technical details in Sec. 3.2) using fidelity i data and fidelity $i-1$ data. The process is summarized in Algorithm 1 and Fig. 1.

Algorithm 1: Compositional Kernel optimization for MF-GPs

```

input:  $[X_i^{(j)}, Y_i^{(j)}], V_f, k;$  /* training data:  $i \in [1, N_f], j \in [0, s]$ ,
vector of fidelities, number of basis kernels */
output:  $V_k^{opt};$  /* vector of optimal compositional
kernels */
1  $S \leftarrow \{k_1, k_2, \dots, k_n\};$  /* where  $k_i$  are the basis kernel
functions */
2  $S_o \leftarrow \{\text{addition, multiplication}\};$ 
3 for  $l=1$  to  $k$  do
4    $V_1 \leftarrow \text{AllCombinations}(S, l);$ 
5    $V_2 \leftarrow \text{AllCombinations}(S_o, l-1);$ 
6   for  $i=1$  to  $\text{length}(V_1)$  do
7     for  $j=1$  to  $\text{length}(V_2)$  do
8       Apply the operations described in  $V_2j$  to functions in  $V_1i$ 
to build  $k_{comp_{ij}};$ 
9        $V_k \leftarrow V_k \cup k_{comp_{ij}}$ 
10    end
11  end
12 end
13 for  $f$  to  $V_f$  do
14   if  $f = \min(V_f)$  then
15     for  $k_{comp_{ij}}$  in  $V_k$  do
16       Build a SF GP model using  $[X_{i=1, \dots, N_f}^{(f)}, Y_{i=1, \dots, N_f}^{(f)}];$ 
17       Calculate  $BIC$  from Eq. (19);
18     end
19   end
20   else
21     for  $k_{comp_{ij}}$  in  $V_k$  do
22       Build a MF-GP model using
 $[X_{i=1, \dots, N_f}^{(f)}, Y_{i=1, \dots, N_f}^{(f)}] \cup [X_{i=1, \dots, N_{f-1}}^{(f-1)}, Y_{i=1, \dots, N_{f-1}}^{(f-1)}];$ 
23       Calculate  $BIC$  from Eq. (19);
24     end
25   end
26   Find the optimal kernel  $k_{comp}^{opt}$  with the minimum  $BIC$  value;
27    $V_k^{opt} \leftarrow V_k^{opt} \cup \{k_{comp}^{opt}\};$ 
28 end

```

3.2 Gaussian Processes: From the Single Fidelity to the Multifidelity Scheme. The mathematical formulation for the GPs is taken from Ref. [43]. A GP is defined as “a collection of random variables, any finite number of which have a joint Gaussian distribution, and it is fully characterized by its mean and covariance

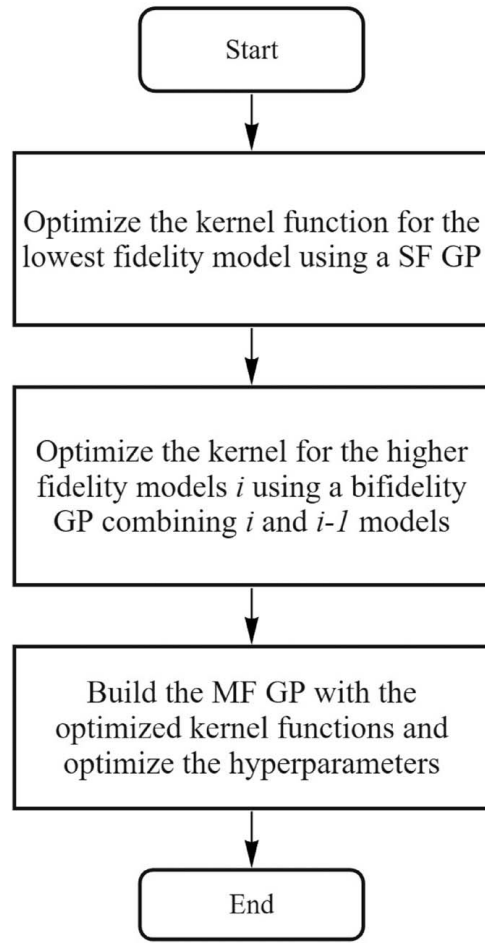


Fig. 1 Flowchart of the proposed method

function” [43]. GPs are used to build approximations of real-world processes $f(x)$, which can be fully defined by a mean $\mu(x)$ and a covariance function $k(x, x')$ according to Eqs. (1)–(3):

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \quad (1)$$

$$m(x) = \mathbb{E}[f(x)] \quad (2)$$

$$k(x, x') = \mathbb{E}[f(x) - m(x)][f(x') - m(x')] \quad (3)$$

The available analysis or experimental data can be described according to Eq. (4):

$$y = f(x) + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma_n^2 I) \quad (4)$$

where f represents the function to be approximated and ε represents the error term. GPs belong to the family of Bayesian methods. For Bayesian methods, a critical element of the analysis is the prior distribution. The prior distribution encodes our prior knowledge or assumptions regarding the unknown function f . The prior distribution of the observed data X and the test data X' is determined according to Eq. (5):

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right) \quad (5)$$

where f_* are the function values evaluated at the test locations X_* . A common practice is to assign the prior a zero mean [43], since data can be normalized to have a zero mean, and a kernel function $K_{ij} = k(x_i, x_j; \theta)$. In Bayesian learning, the prior distribution is revised by incorporating the observed data, resulting in the

formation of the posterior distribution. Mathematically, the prior distribution is conditioned on the observed data to form the posterior distribution according to Eqs. (6)–(8):

$$f_*|X, X_*, y \sim \mathcal{N}(\mu_*, \sigma_*^2) \quad (6)$$

$$\mu_* = k_*^T [K + \sigma_n^2 I]^{-1} y \quad (7)$$

$$\sigma_*^2 = k_{**} - k_*^T [K + \sigma_n^2 I]^{-1} k_* \quad (8)$$

where $K = K(X, X)$, $k_{**} = k(x_*, x_*)$, and $k_* = k(x_*)$. There are various methods to optimize the kernel hyperparameters such as cross-validation and maximum likelihood estimation [53]. In the present work, the maximization of the marginal log-likelihood was applied. The marginal log-likelihood is defined according to Eq. (9).

$$\log p(y|x, \theta) = -\frac{1}{2} \log |K + \sigma_n^2 I| - \frac{1}{2} y^T [K + \sigma_n^2 I]^{-1} y - \frac{n}{2} \log 2\pi \quad (9)$$

The autoregressive scheme proposed by Kennedy and O'Hagan [44] assumes a linear dependency of the various fidelity models. It is assumed that there are s levels of code fidelity $(f_i(x))_{s=1, \dots, s}$ modeled by GPs $(F_i(x))_{s=1, \dots, s}$, where $X \in U \subset \mathbb{R}^d$. The code fidelity increases from 1 to s , thus f_s is the most accurate model. The mathematical formulation follows the description in Ref. [54]. The model is based on the Markov property, described in Eq. (10), which states that given the nearest point $F_{t-1}(x)$, we can learn no more for $F_t(x)$ from any other $F_{t-1}(x')$ for $x \neq x'$. This assumption leads to the autoregressive model.

$$\rho_{t-1}(x) = \frac{\text{cov}(F_t(x), F_{t-1}(x))}{\text{var}(F_{t-1}(x))}, \forall x \neq x' \quad (10)$$

The submodels are connected according to Eq. (11). The higher fidelity function connects to the lower fidelity function via a scaling function ρ_t (Eq. (13)) and an additive function δ_t (Eq. (14)). The scaling function ρ_t determines the scale factor and the correlation degree between two successive levels of code. The function δ_t is a Gaussian process independent of $F_{t-1}(x)$ (Eq. (12)). The lowest fidelity function F_1 is described by Eq. (15).

$$F_t(x) = \rho_{t-1}(x)F_{t-1}(x) + \delta_t(x) \quad (11)$$

$$F_{t-1}(x) \perp \delta_t(x) \quad (12)$$

$$\rho_{t-1} = g_{t-1}^T(x)\beta_{\rho_{t-1}} \quad (13)$$

$$\delta_t(x) \sim \mathcal{GP}(\mu_t^T(X)\beta_t, \sigma_t^2 r_t(x, x')) \quad (14)$$

$$F_1(x) \sim \mathcal{GP}(\mu_1^T(X)\beta_1, \sigma_1^2 r_1(x, x')) \quad (15)$$

where $g_{t-1}(x)$ is a vector of q_{t-1} regression functions, $r_t(x, x')$ is a correlation function, $\mu_t(x)$ is a vector of p_t regression functions, β_t is a p_t -dimensional vector, β_{t-1} is a q_{t-1} -dimensional vector, and σ_t^2 is a positive real number. The trend parameters are denoted as $\beta = (\beta_1^T, \dots, \beta_s^T)^T$, the adjustment parameters are represented as $\beta_\rho = (\beta_{\rho_1}^T, \dots, \beta_{\rho_s}^T)^T$, and the variance parameters are expressed as $\sigma = (\sigma_1, \dots, \sigma_s)$. The predictive model of the highest fidelity response f_s is calculated according to Eqs. (16)–(18).

$$[F_s(x)|F^{(s)} = f^{(s)}, \beta, \beta_\rho, \sigma^2] \sim \mathcal{N}(m_{F_s}(x), s_{F_s}^2(x)) \quad (16)$$

$$m_{F_s}(x) = h^{(s)}(x)^T \beta + t_s(x)^T (V^{(s)})^{-1} (f^{(s)} - H^{(s)} \beta) \quad (17)$$

$$s_{F_s}(x) = v_{F_s}^2(x) - t_s(x)^T (V^{(s)})^{-1} t_s(x) \quad (18)$$

where $V^{(s)}$ represents the covariance matrix of $F^{(s)}$, $t_s(x)$ denotes the vector of covariances between $F_s(x)$ and $F^{(s)}$, $H^{(s)} \beta$ stands for the mean of $F^{(s)}$, $h^{(s)}(x)^T \beta$ is the mean of $F_s(x)$, and $v_{F_s}^2(x)$ expresses the variance of $F_s(x)$. For further details regarding the mathematical formulations, the reader is referred to the original papers [44,54]. The method of optimizing the hyperparameters is similar to the one explained for single fidelity GPs by maximizing the marginal log-likelihood (Eq. (9)).

3.3 Compositional Kernels. The kernel, a measure of similarity between data points [43], incorporates the prior beliefs and knowledge about the function f . Kernel validity demands symmetry and positive semi-definiteness. Previous studies have produced basis functions for constructing valid covariance matrices, such as the periodic kernel for modeling repeating functions [55].

Duvenaud et al. [56] introduced compositional kernels, which are formed by combining a limited number of basis kernels through addition or multiplication. The idea of the method was to decompose the function to be learned into interpretable components.

For constructing the compositional kernels, a set of basis kernel functions was determined. The set included the exponential and squared exponential kernels, the linear kernel, the Brownian kernel, the white noise kernel, the Matérn 3/2 and Matérn 5/2 kernels, the constant kernel, and the periodic kernel. When selecting a limited number of basis functions to compose the compositional kernel, an exhaustive search method was employed. More, specifically, vector **V1** contains all possible combinations of k basis kernels out of nine possible functions. Vector **V2** contains all possible combinations of operations (addition, multiplication) for the k basis kernels. Each combination of operations described in **V2** is applied to every element of **V1**. This process yields a final vector **V3**. Each element of **V3** is assessed based on the Bayesian information criterion BIC as proposed in the original paper [56]. BIC is defined according to Eq. (19):

$$BIC = k_{\text{hyp}} \ln n - 2 \ln L \quad (19)$$

where n is the number of training data, k_{hyp} is the number of hyperparameters, and L is the maximized likelihood value. BIC consists of two components, a penalty term based on the number of model parameters and a term based on the likelihood function. The benefit of using BIC over maximizing the marginal log-likelihood lies in its consideration of the kernel function's complexity. By favoring functions with fewer hyperparameters, BIC helps prevent overfitting. The main idea of building the compositional kernel is shown in Algorithm 1.

4 Case Studies

This section presents the case studies and the results. The case studies considered encompass a range of analytical benchmark problems proposed by Ref. [15], as well as an engineering problem involving a cantilever beam. The case studies involving the analytical functions are the following:

- the Forrester function (used for conceptualization)
- the Jump Forrester function
- the ND Rosenbrock function
- the Heterogeneous function
- the 2D shifted-rotated Rastrigin function

Following a notation similar to that of the previous sections, the models are numbered as follows: the HF model is labeled as s , while the LF models are sequentially numbered from 1 to $s - 1$, where 1 represents the lowest fidelity among the LF models. The PYTHON packages used included GPy [57] and Emukit [58]. Latin hypercube sampling technique was employed for the selection of analysis points, and a total of 20 different design of experiments (DoEs) were used to calculate statistics pertaining to prediction errors. Throughout the case studies, the reference model refers to the

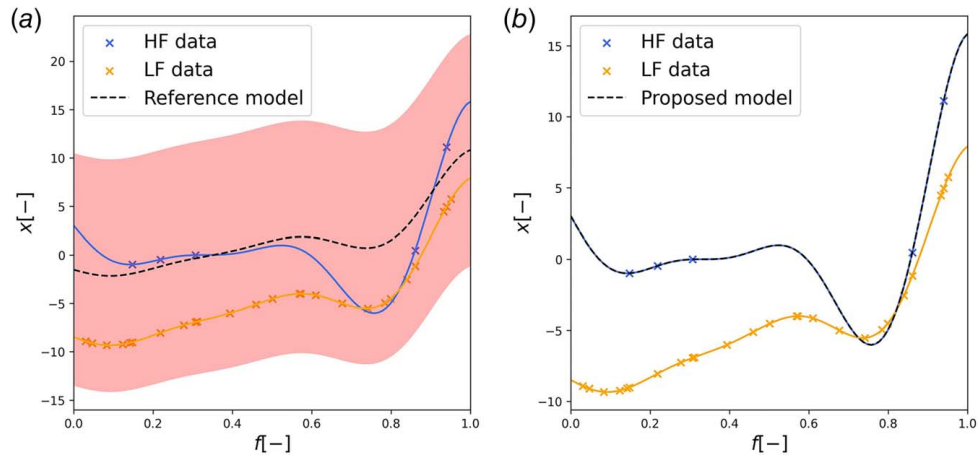


Fig. 2 Forrester function using 5 HF and 25 LF data: (a) reference model: RBF kernel and (b) proposed model: optimized kernel

AR1 model with the squared exponential kernel, whereas the proposed model refers to the AR1 model with compositional kernels. The prediction error was evaluated using two measures, namely, R^2 and the normalized root-mean-square error (RMSE), which are expressed as follows according to Eqs. (20) and (21):

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^S (y_i - \bar{y})^2} \quad (20)$$

$$\text{RMSE} = \frac{1}{y_{\max} - y_{\min}} \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (21)$$

where y_i refers to the observed value of each data point i , \hat{y}_i refers to the predicted value of each data point i , \bar{y} refers to the mean of the observed values for all the data points, and N refers to the total number of the samples. Both errors are used to give a more thorough understanding of the quality of the predictions. More specifically, R^2 measures the proportion of variability in a dependent variable, which can be captured by using the independent variable [59]. In the context of linear models, this measure provides a good intuitive understanding as its value ranges from 0 to 1 [60]. For nonlinear models, such as GPs, R^2 is defined, but it is not confined to the range (0, 1) [61]. A negative value would suggest that the model's performance is poorer than the average of the predicted values. The metric R^2 alone is inadequate for fully assessing the specific models' performance; therefore, it was used in conjunction with RMSE. The latter metric assesses the accuracy of the model in terms of the residual error.

4.1 Baseline Example: The Forrester Function. The primary objective of this particular case study, as opposed to the others, is to facilitate comprehension and visualization of the concept of exploring the shape of the design space. For this investigation, we adopt the Forrester function as described in Eqs. (22) and (23) to represent the design space.

$$f_1(x) = 0.5f_2(x) + 10(x - 0.5) - 5 \quad (22)$$

$$f_2(x) = (6x - 2)^2 \sin(12x - 4) \quad (23)$$

This case discusses and demonstrates the benefit of compositional kernels in early design exploration. For this analysis, 5 HF and 25 LF analysis data were used. Figure 2 shows the LF approximation represented by the blue line, the HF approximation depicted in orange, and the prediction illustrated by the black dashed line. As shown in Fig. 2(a), the prediction using the AR1 scheme with the squared exponential kernel and the given observational data does not effectively model the HF function. On the other hand, as shown in Fig. 2(b), the prediction using the AR1 scheme and the given observational data give an accurate prediction of the HF function, which represents the design space. In this instance, the kernel for the LF data was represented as a product of a linear kernel and a white noise kernel, while the squared exponential kernel was employed for the HF data. This specific case visually demonstrates that the additional information provided by the compositional kernel about the structure of the HF function can improve the performance of the framework. Thus, it is possible to make more accurate predictions with less HF data, thereby reduce the required computational cost. It is clear that the proposed approach incurs additional computational cost for developing the compositional kernels; however, the computational time and costs of running the

Table 1 Error measures calculated for the Jump Forrester function varying number of HF points

DoE	GP HF R^2 (std)	GP HF RMSE (std)	Ref. model R^2 (std)	Ref. model RMSE (std)	Prop. model R^2 (std)	Prop. model RMSE (std)
(5,25)	0.1894 (0.3832)	0.2409 (0.0569)	0.6953 (0.4458)	0.1297 (0.0787)	0.7435 (0.2294)	0.1280 (0.0547)
(8,25)	0.5451 (0.2742)	0.1772 (0.0544)	0.8704 (0.0545)	0.0971 (0.0191)	0.8554 (0.0968)	0.0996 (0.0316)
(10,25)	0.5052 (0.3329)	0.1805 (0.0695)	0.8085 (0.2703)	0.1058 (0.0574)	0.8732 (0.1296)	0.0892 (0.0403)
(15,25)	0.7322 (0.2826)	0.1235 (0.0706)	0.9095 (0.0490)	0.0798 (0.0217)	0.9384 (0.0682)	0.0626 (0.0271)

Table 2 Assessment of the various models for the Jump Forrester function

DoE	Improvement ref. model compared to the GP HF (%)	Improvement prop. model compared to the GP HF (%)	Improvement prop. model compared to the ref. model (%)
(5,25)	46	47	1
(8,25)	45	44	-3
(10,25)	41	51	15
(15,25)	35	49	22

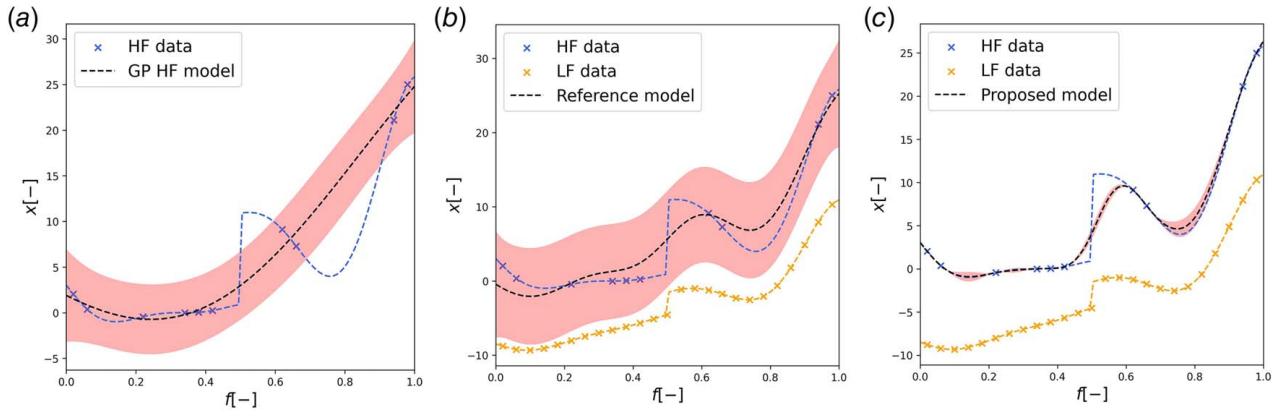


Fig. 3 Jump Forrester using 10 HF and 25 LF points: (a) GP HF model, (b) reference model, and (c) proposed model

Table 3 Error measures calculated for the ND Rosenbrock function varying number of HF points

Dimensions DoE	GP HF R^2 (std)	GP HF RMSE (std)	Ref. model R^2 (std)	Ref. model RMSE (std)	Prop. model R^2 (std)	Prop. model RMSE (std)
4	0.2143	0.1127	0.9685	0.0228	0.9955	0.0085
(45,140)	(0.2886)	(0.0243)	(0.0108)	(0.0037)	(0.0020)	(0.0018)
6	-0.0092	0.1449	0.8188	0.0594	0.9824	0.0187
(55,160)	(0.0368)	(0.0027)	(0.1065)	(0.0155)	(0.0078)	(0.0040)
8	0.0065	0.1247	0.6489	0.0725	0.9399	0.0301
(65,180)	(0.0404)	(0.0026)	(0.1421)	(0.0157)	(0.0315)	(0.0059)
10	-0.01198	0.1203	0.5530	0.0783	0.8574	0.0447
(75,180)	(0.0153)	(0.0009)	(0.1844)	(0.0163)	(0.0495)	(0.0067)
15	-0.00224	0.1232	0.1173	0.1143	0.7037	0.0660
(100,250)	(0.0186)	(0.0012)	(0.2470)	(0.0176)	(0.1341)	(0.0115)
20	0.0032	0.1402	-0.1913	0.1523	0.6381	0.0838
(125,300)	(0.0216)	(0.0015)	(0.1690)	(0.0122)	(0.0751)	(0.0083)

Table 4 Assessment of the various models for the ND Rosenbrock function

Dimensions	DoE	Improvement ref. model compared to the GP HF (%)	Improvement prop. model compared to the GP HF (%)	Improvement prop. model compared to the ref. model (%)
4	(45,140)	80	92	63
6	(55,160)	59	87	68
8	(65,180)	42	76	58
10	(75,200)	35	63	43
15	(100,250)	7	46	42
20	(125,300)	-8	40	45

MF analysis are less than obtaining additional HF computational or physical experimental data.

aimed at introducing discontinuities. The Jump Forrester is described by Eqs. (24) and (25).

4.2 Addressing Discontinuities: The Jump Forrester Function. The Jump Forrester is a variation of the Forrester

$$f_1(x) = \begin{cases} 0.5f_2(x) + 10(x - 0.5) - 5, & 0 \leq x < 0.5 \\ 0.5f_2(x) + 10(x - 0.5) - 2, & 0.5 \leq x \leq 1 \end{cases} \quad (24)$$

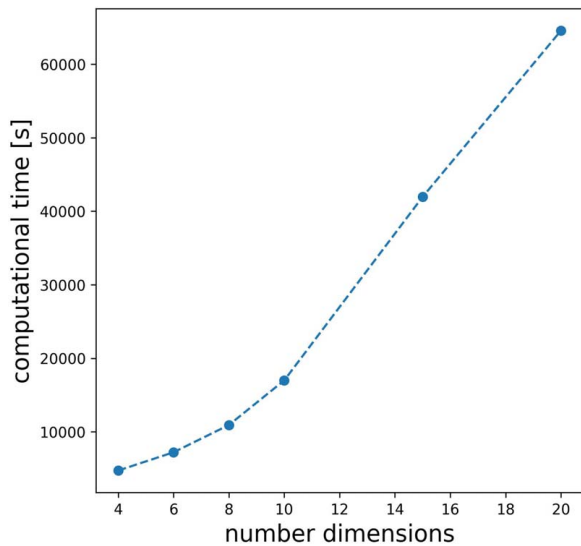


Fig. 4 Computational cost comparison across various dimensions of the ND Rosenbrock function

$$f_2(x) = \begin{cases} (6x - 2)^2 \sin(12x - 4), & 0 \leq x < 0.5 \\ (6x - 2)^2 \sin(12x - 4) + 10, & 0.5 \leq x \leq 1 \end{cases} \quad (25)$$

In this particular bi-fidelity case study, we used a total of 25 LF points while varying the number of HF points in the range of 5–15. The Latin hypercube sampling method was employed to generate 20 different datasets to determine the statistics of the error measures. The results are presented in Tables 1 and 2.

The results demonstrate that both, the proposed and reference models, exhibit similar performance in the cases of five and eight HF points. However, it outperforms the reference model when using 10 and 15 HF points. The improvement ranges from 1 to 22% depending on the number of HF points, but it is negative (–3%) in the case of eight HF points. This suggests that while the proposed model shows the potential for significant advancements in scenarios with limited data, where both the SF model and the reference model struggle to accurately represent the underlying design space, the amount of HF data needs to be adequate to properly capture the function’s structure. The most substantial improvement is observed when using 15 HF points. A representative case is illustrated in Fig. 3. The design space includes 10 HF and 25 LF data points. The SF model in Fig. 3(a) fails to capture the function. The reference model (Fig. 3(b)) is better but not entirely accurate. In contrast, the proposed model (Fig. 3(c)) demonstrates superior accuracy in predicting the function. In this case, the kernel function for the LF data was a multiplication of a linear and a Brownian kernel, while for the HF data, the Matérn 5/2 kernel was chosen. Based on these findings, it can be concluded that neither model

adequately captures the discontinuity; however, the proposed model exhibits enhanced capability in capturing the function within the remaining domain.

4.3 Scalability: The ND Rosenbrock Function. As previously mentioned, in practical applications, the design space often involves a large number of design variables. Therefore, the scalability of modeling methods becomes a crucial consideration. To illustrate the performance of the proposed model in such design spaces, the ND Rosenbrock function was employed as a representative test case. The Rosenbrock function was evaluated in various dimensions, ranging from 4D to 20D. Equations (26) and (27) were used to describe the Rosenbrock function.

$$f_1(x) = \frac{f_2(x) - 4 - \sum_{i=1}^D 0.5x_i}{\sum_{i=1}^D 0.5x_i} \quad (26)$$

$$f_2(x) = \sum_{i=1}^{D-1} 100(X_{i+1} - x_i^2)^2 + (1 - x_i)^2 \quad (27)$$

where $x_i \in [-2, 2]$.

In this specific bi-fidelity case study, we systematically increased the volume of HF and LF data in alignment with the number of dimensions. The quantity of HF data ranged from 45 to 125, while LF data varied from 140 to 300 data points. Importantly, this approach has no impact on the relative performance of the models, as all three models were trained with identical data volumes. To assess the statistics of error measures, LHS was employed to generate 20 distinct datasets. The resulting outcomes are presented in Tables 3 and 4. In the case of all three models, the results reveal a decline in their performance with the increasing dimensionality. This outcome aligns with our expectations, as higher dimensionality brings about greater complexity, necessitating a larger volume of training data to achieve the same level of accuracy. Despite augmenting the training data as the problem scaled, the extent of this increase did not compensate for the heightened complexity. The proposed model outperforms the reference model in all the examined cases and showed a significant improvement. It is noteworthy that in the context of 20 dimensions, the reference model proves ineffective in predicting the function ($R^2 = -0.1913$), while the proposed model maintains a satisfactory level of accuracy ($R^2 = 0.6381$).

One of the well-known challenges to address when using Gaussian processes is that the computational complexity of training a GP model is known to be $O(N^3)$, where N represents the number of data points [62]. This cubic complexity poses challenges when dealing with large datasets or high-dimensional design spaces. This significantly impacts the procedure of constructing compositional kernels, making it computationally expensive for high-dimensional input spaces. Figure 4 displays the escalating computational costs plotted against the dimensions of the function. To provide an indication of this increase, the average time of building the compositional kernels for the 4D Rosenbrock function is 4728 s, which

Table 5 Error measures calculated for the heterogeneous function varying the number of HF points

DoE	GP HF R^2 (std)	GP HF RMSE (std)	Ref. model R^2 (std)	Ref. model RMSE (std)	Prop. model R^2 (std)	Prop. model RMSE (std)
(5,25)	-0.3939 (1.6100)	0.4476 (0.2197)	0.1387 (1.5754)	0.3081 (0.2423)	0.6144 (0.7309)	0.2115 (0.1550)
(8,25)	0.5476 (0.3719)	0.2599 (0.1145)	0.9114 (0.0785)	0.1189 (0.0409)	0.9569 (0.0022)	0.0876 (0.0022)
(10,25)	0.6941 (0.2431)	0.2160 (0.0888)	0.9127 (0.0639)	0.1193 (0.0367)	0.9576 (0.0017)	0.0869 (0.0018)
(15,25)	0.8813 (0.1401)	0.1323 (0.0606)	0.9066 (0.0291)	0.1271 (0.0220)	0.9576 (0.0021)	0.0869 (0.0022)

Table 6 Assessment of the various models for the heterogeneous function

DoE	Improvement ref. model compared to the GP HF (%)	Improvement prop. model compared to the GP HF (%)	Improvement prop. model compared to the ref. model (%)
(5,25)	31	53	31
(8,25)	54	66	26
(10,25)	45	60	27
(15,25)	4	34	32

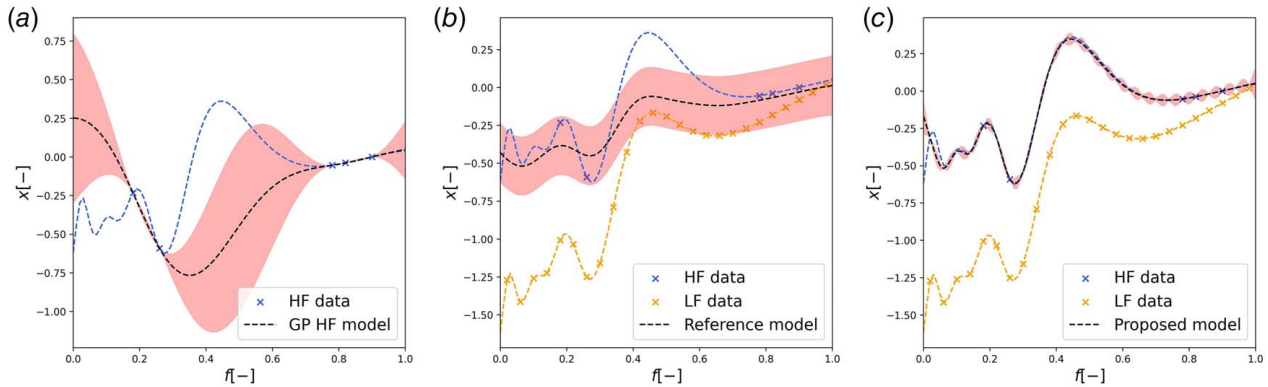


Fig. 5 Heterogeneous function using 5 HF and 25 LF points: (a) GP HF model, (b) reference model, and (c) proposed model

was 13 times lower than the time required for the 20D Rosenbrock function, which was 64,557 s.

4.4 Discovering Complex Patterns: The Heterogeneous Function. Complex design spaces are often characterized by intricate structures. To evaluate the performance of models in solving such design spaces, a commonly employed analytical function is the heterogeneous function, known for its localized and multimodal behavior [15]. The 1D heterogeneous function is described by Eqs. (28) and (29).

$$f_1(x) = (f_2(x) - 1.0 + x)/(1.0 + 0.25x) \quad (28)$$

$$f_2(x) = \sin 30(x - 0.9)^4 \cos 2(x - 0.9) + (x - 0.9)/2 \quad (29)$$

where $0 \leq x \leq 1$. In the analysis, similar to the Jump Forrester function, the number of HF points varied from 5 to 15, while the number of LF points remained constant at 25. The results are summarized in Tables 5 and 6.

Notably, the MF approach demonstrates a significant advantage over the SF approach, particularly across the range of tested HF points. Moreover, the proposed model exhibits improved prediction accuracy across all the tested DoEs. The improvement in the predictions of the proposed model compared to the reference model ranges from 26% to 32%. Insights into the performance of the models can be gained from Fig. 5, which specifically focuses on the case where a dataset of 5 HF points and 25 LF points is shown. In Figs. 5(a) and 5(b), it is evident that both the SF model and the reference model struggle to accurately predict the shape of the function. The proposed model employed a kernel function comprising the multiplication of the linear and Brownian kernels for the LF data, while a squared exponential kernel was used for the HF data. Notably, the proposed model achieves a more precise representation of the function throughout the entire domain. However, one drawback of the method is that the uncertainty bounds are reduced even in the area close to $x = 0$, where the model fails to capture the structure of the function.

4.5 Noisy Observations: The 2D Shifted-Rotated Rastrigin Function. In this case study, the 2D shifted-rotated Rastrigin function was employed. This function is characterized by multimodal

behavior. In practical applications, the analysis data and experimental data used for design optimization often contain noise. Therefore, it is important to investigate the performance of the model while dealing with noisy training data. To investigate this, a noise term e_{data} was added to the 2D shifted-rotated Rastrigin function, taken from Ref. [15]. Thus, for this analysis, Eqs. (30) and (31) are used. The function can be visualized in Fig. 6.

$$f_1(z, \phi_i) = f_2(z) + e_r(z, \phi_i) + e_{data} \quad (30)$$

where the resolution error e_r is defined according to Eq. (34).

$$f_2(z) = \sum_{i=1}^{D=2} (z_i^2 + 1 - \cos(10\pi z_i)) \quad (31)$$

where

$$z = R(\theta)(x - x^*) \quad (32)$$

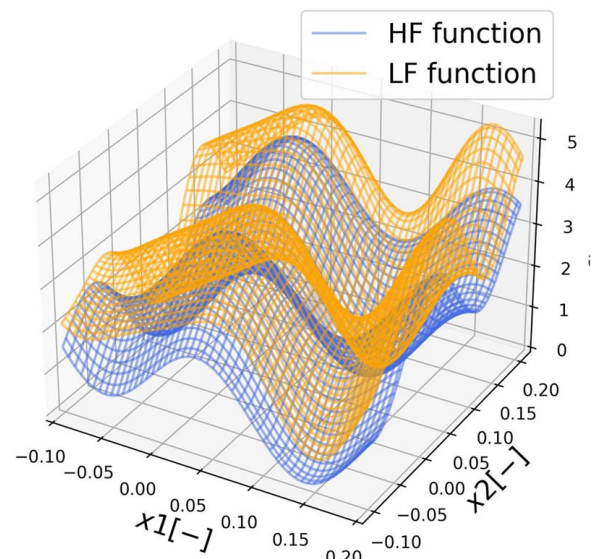


Fig. 6 Visualization of the Rastrigin function

Table 7 Error measures calculated for the 2D Rastrigin function varying number of HF points (bifidelity case)

DoE	GP HF R^2 (std)	GP HF RMSE (std)	Ref model R^2 (std)	Ref model RMSE (std)	Prop model R^2 (std)	Prop model RMSE (std)
(5,100)	-0.5539 (0.7471)	0.3023 (0.0628)	0.1910 (0.4308)	0.2114 (0.0704)	0.7925 (0.0604)	0.1118 (0.0153)
(10,100)	-0.2398 (0.3482)	0.2735 (0.0355)	0.2130 (0.5096)	0.2039 (0.0820)	0.8311 (0.0371)	0.1013 (0.0102)
(15,100)	-0.0627 (0.1489)	0.2547 (0.0186)	0.2727 (0.3860)	0.2002 (0.0671)	0.8543 (0.0248)	0.0942 (0.0082)
(20,100)	0.1169 (0.3703)	0.2260 (0.0558)	0.4944 (0.4170)	0.1586 (0.0761)	0.8692 (0.0599)	0.0878 (0.0179)
(25,100)	0.0332 (0.2743)	0.2384 (0.0497)	0.3945 (0.4485)	0.1531 (0.0847)	0.9028 (0.0325)	0.0762 (0.0127)
(30,100)	0.1018 (0.3339)	0.2267 (0.0610)	0.5224 (0.4562)	0.1489 (0.0845)	0.9458 (0.0323)	0.0559 (0.0142)

$$R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (33)$$

where $x_i \in [-0.1, 0.2]$ for $i = 1, \dots, D$, R is the rotation matrix, $\theta = 0.2$, and x^* is the location of the global optimum at $[0.1, \dots, 0.1]^T$.

$$e_r(z, \phi_i) = \sum_{i=1}^{D=2} \alpha(\phi) \cos^2(w(\phi)z_i + \beta\phi + \pi) \quad (34)$$

with $\alpha(\phi) = \Theta(\phi)$, $w(\phi) = 10\pi\Theta$, $\beta(\phi) = 0.5\pi\Theta(\phi)$, and $\Theta(\phi) = 1 - 0.0001\phi$. For the present case study, we chose $\phi = 2500$.

The outcomes are displayed in Tables 7 and 8. It is evident that the GP HF model falls short in capturing the underlying function. In contrast, both the proposed and the reference model demonstrated substantial enhancements. The reference model shows improvements ranging from 21% to 34%, while the proposed model delivers enhancements between 61% and 75%. In terms of statistical error analysis, we generated RMSE plots for the three models over 20 iterations to assess error convergence. As illustrated in Figs. 7(a)–7(c), it is clear that the error values reach a plateau around iteration 16. The case study was extended to a three-fidelity scenario, and the outcomes are presented in Tables 9 and 10. These results exhibit analogous trends to the bifidelity case, with the distinction that both the proposed and reference models show greater improvements in predictions, ranging from 32% to 63% and from 72% to 79%, respectively. Emphasizing the impact of incorporating additional models, it is worth noting that this results in escalating

Table 8 Assessment of the various models for the 2D Rastrigin function (bifidelity case)

DoE	Improvement ref. model compared to the GP HF (%)	Improvement prop. model compared to the GP HF (%)	Improvement prop. model compared to the ref. model (%)
(5,100)	30	63	47
(10,100)	25	62	50
(15,100)	21	63	52
(20,100)	30	61	45
(25,100)	27	68	55
(30,100)	34	75	62

computational costs. Figure 8 illustrates the computational cost for both the bifidelity and trifidelity scenarios.

4.6 Simplified Design Problem: The Cantilever Beam. The proposed framework was tested on a structural design problem involving a cantilever beam. This particular problem was chosen because it serves as a simplified representation of real-world, complex engineering problem, such as estimating lifetime loads on intricate structures like aircraft or ships. The formulation of the problem was taken from Ref. [51] and modified.

The cantilever beam is shown in Fig. 9(a). The square-section beam is fixed to the wall on one end, while a concentrated load is applied to the opposite end. In addition, there is a hole on the side that is anchored to the wall. The aim is the calculation of the

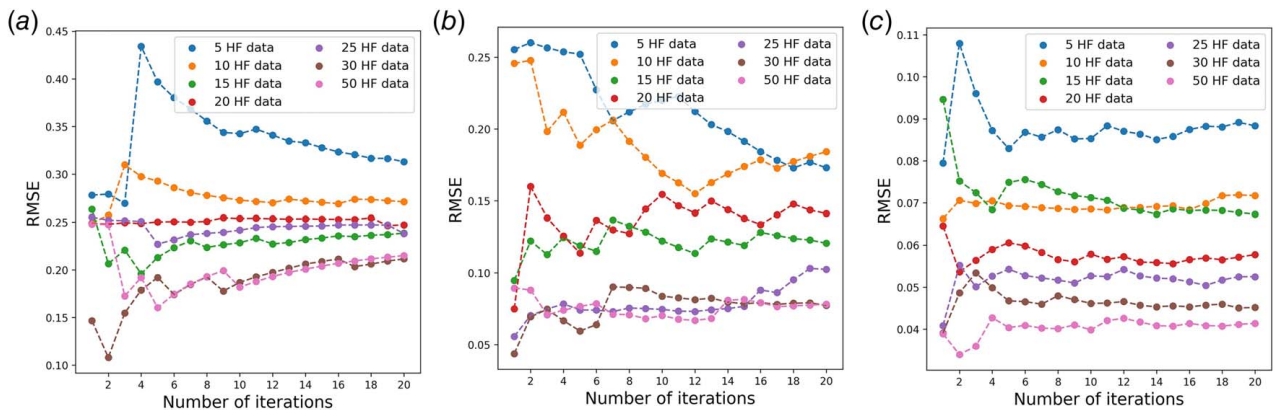


Fig. 7 Convergence of the RMSE based on the Rastrigin bifidelity case study: (a) GP HF model, (b) reference model, and (c) proposed model

Table 9 Assessment of the various models for the 2D Rastrigin function (three fidelity case)

DoE	GP HF R^2 (std)	GP HF RMSE (std)	Ref model R^2 (std)	Ref model RMSE (std)	Prop model R^2 (std)	Prop model RMSE (std)
(5,50,100)	-0.9417 (2.7959)	0.3131 (0.1454)	0.4089 (0.4451)	0.1730 (0.0797)	0.8663 (0.0622)	0.0884 (0.0200)
(10,50,100)	-0.2253 (0.4342)	0.2710 (0.0414)	0.3428 (0.4362)	0.1842 (0.0800)	0.9145 (0.0267)	0.0717 (0.0101)
(15,50,100)	0.0385 (0.3127)	0.2385 (0.0461)	0.6996 (0.3448)	0.1206 (0.0623)	0.9234 (0.0319)	0.0673 (0.0130)
(20,50,100)	-0.0154 (0.2139)	0.2472 (0.0348)	0.5554 (0.4671)	0.1412 (0.0856)	0.9440 (0.0189)	0.0577 (0.0102)
(25,50,100)	0.0394 (0.2688)	0.2381 (0.0474)	0.7627 (0.3339)	0.1025 (0.0636)	0.9527 (0.0217)	0.0525 (0.0121)
(30,50,100)	0.1826 (0.3981)	0.2115 (0.0736)	0.8700 (0.2054)	0.0773 (0.0446)	0.9655 (0.0137)	0.0452 (0.0089)

developed von Mises (VM) stress. The problem is set up as a bi-fidelity problem, where the LF method is the analytical estimation of the maximal von Mises (VM) stress and the HF method is the numerical estimation of the maximal VM stress. Furthermore, the problem was modeled as a two-dimensional (2D) problem, with the independent variables being the beam's length (L), and diameter (d). The problem domain was defined within the ranges of $L \in [2.0, 3.0]$ m, and $d \in [0.25, 0.4]$ m. The applied force was established as a constant value of 950 kN.

The equations can be found in Ref. [63]. To accurately calculate the Von Mises (VM) stress through analytical means, both the shear force and the bending moment, exerted along the beam's length when it is subjected to a transverse force, need to be accounted for. The maximal bending stress σ_b is calculated according to Eq. (35):

$$\sigma_b = -\frac{M_b(\frac{d}{2})}{I} \quad (35)$$

$$M_b = -FL \quad (36)$$

where M_b is the bending moment, I is the moment of inertia, F is the applied force, and L is the beam's length. For square cross sections, the moment of inertia I is calculated based on Eq. (37).

$$I = \frac{d^4}{12} \quad (37)$$

The calculation for the average shear force, τ , is determined according to Eq. (38):

$$\tau = \frac{FQ}{Id} \quad (38)$$

where Q is the first moment of area with respect to the neutral axis that lies above the point of interest.

$$Q = \int_A y dA = \frac{d^3}{2} \quad (39)$$

The maximal VM stress is calculated according to Eq. (40).

$$\sigma_{VM}^{max} = \sqrt{\sigma_b^2 + 3\tau^2} \quad (40)$$

In the numerical model, a hole was incorporated into the cantilever beam design. The dimensions of the hole were determined based on the beam's main dimensions according to Eqs. (41)–(43). The material properties of the beam were specified as follows: steel with a Young's modulus (E) of $2e + 11$ GPa and a Poisson's ratio (ν) of 0.30. The model was developed using ANSYS

Table 10 Assessment of the various models for the 2D Rastrigin function (three fidelity case)

DoE	Improvement ref. model compared to the GP HF (%)	Improvement prop. model compared to the GP HF (%)	Improvement prop. model compared to the ref. model (%)
(5,50,100)	44	72	49
(10,50,100)	32	74	61
(15,50,100)	49	72	44
(20,50,100)	43	77	49
(25,50,100)	57	78	49
(30,50,100)	63	79	42

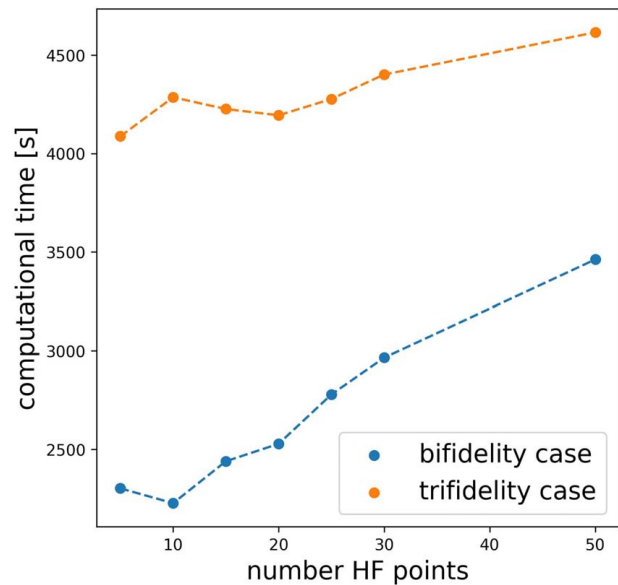


Fig. 8 Computational cost comparison based on the Rastrigin function case

software.

$$\alpha_i = \frac{L_i}{L_{max}} \quad (41)$$

$$L_{hole}^i = 0.3\alpha_i \cdot L_i \quad (42)$$

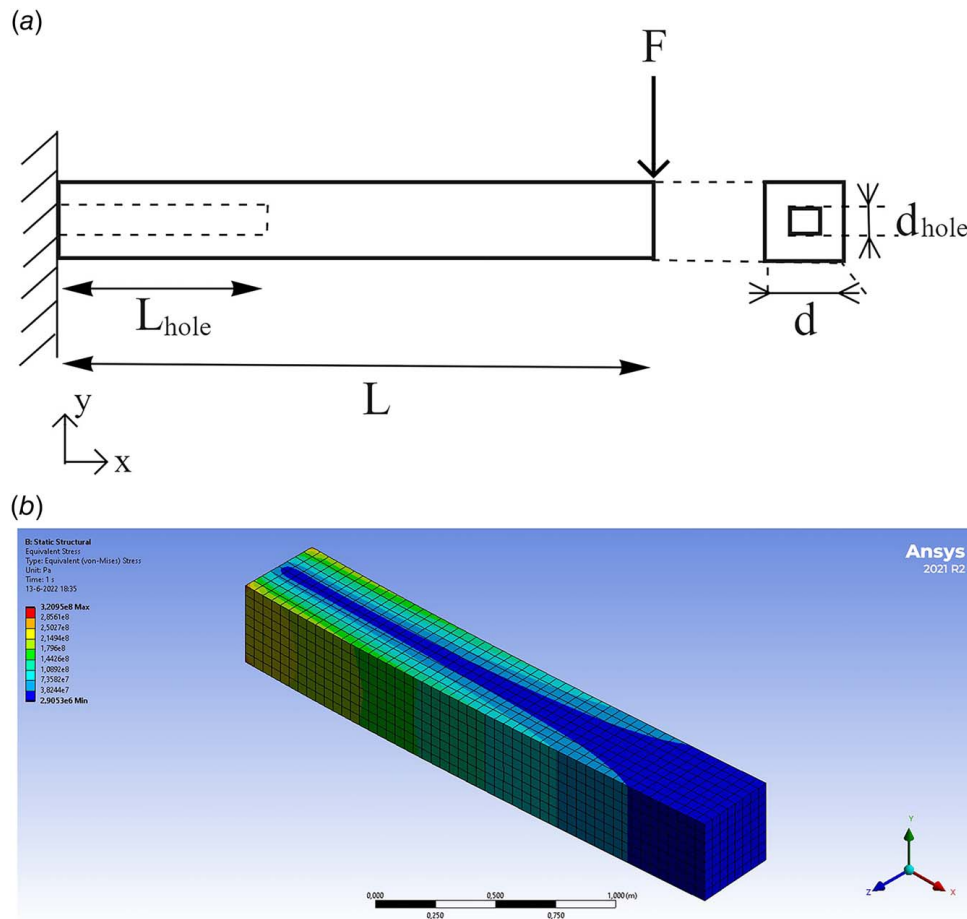


Fig. 9 Cantilever beam case study: (a) schematic representation and (b) equivalent von Mises stresses calculated by ANSYS

$$d_{\text{hole}}^i = 0.3\alpha_i \cdot d_i \quad (43)$$

The outcomes are presented in Tables 11 and 12. These results indicate that the reference model yields results that are on par with the SF model. This can primarily be attributed to the considerable disparity between the LF fidelity model and the HF model, which is attributed to the presence of the hole. In contrast, the predictions of the proposed model are closer to the HF design space. Based on Table 12, the improvement compared to the SF model ranged from 15% to 24%. An example of the problem can be visualized in Fig. 10.

5 Discussion

5.1 Discussion on the Presented Case Studies. In summary, the findings demonstrated that the incorporation of compositional kernels significantly enhanced the predictive capabilities of the AR1. Various analytical benchmark problems were simulated to thoroughly test the proposed model.

The 1D Jump Forrester function, representing a discontinuous space, was treated as a bi-fidelity problem. The number of HF points ranged from 5 to 15, while LF points remained constant at 25. The proposed model yielded an improvement in predictions,

Table 11 Error measures calculated for the cantilever beam varying the number of HF points

DoE	GP HF R^2 (std)	GP HF RMSE (std)	Ref model R^2 (std)	Ref model RMSE (std)	Prop model R^2 (std)	Prop model RMSE (std)
(10,50)	-0.1286 (0.8933)	0.1892 (0.0748)	-0.2348 (1.0045)	0.1979 (0.0784)	0.0710 (1.0450)	0.1578 (0.0989)
(15,50)	0.3159 (0.4988)	0.1484 (0.0535)	0.3626 (0.4623)	0.1437 (0.0518)	0.5561 (0.5115)	0.1127 (0.0627)
(20,50)	0.4264 (0.3890)	0.1398 (0.0514)	0.4255 (0.3847)	0.1400 (0.0504)	0.5339 (0.4326)	0.1190 (0.0648)
(25,50)	0.6122 (0.1661)	0.1149 (0.0290)	0.5875 (0.1872)	0.1184 (0.0308)	0.6883 (0.2393)	0.0975 (0.0418)
(30,50)	0.6200 (0.1840)	0.1129 (0.0256)	0.5941 (0.1851)	0.1170 (0.0250)	0.6989 (0.1951)	0.0991 (0.0322)
(35,50)	0.6562 (0.1528)	0.1073 (0.0162)	0.6484 (0.1575)	0.1085 (0.0158)	0.7391 (0.1798)	0.0910 (0.0308)

Table 12 Assessment of the various models for the cantilever beam

DoE	Improvement ref. model compared to the GP HF (%)	Improvement prop. model compared to the GP HF (%)	Improvement prop. model compared to the ref. model (%)
(10,50)	-5	17	20
(15,50)	3	24	22
(20,50)	0	15	15
(25,50)	-3	15	17
(30,50)	-4	12	15
(35,50)	-1	15	16

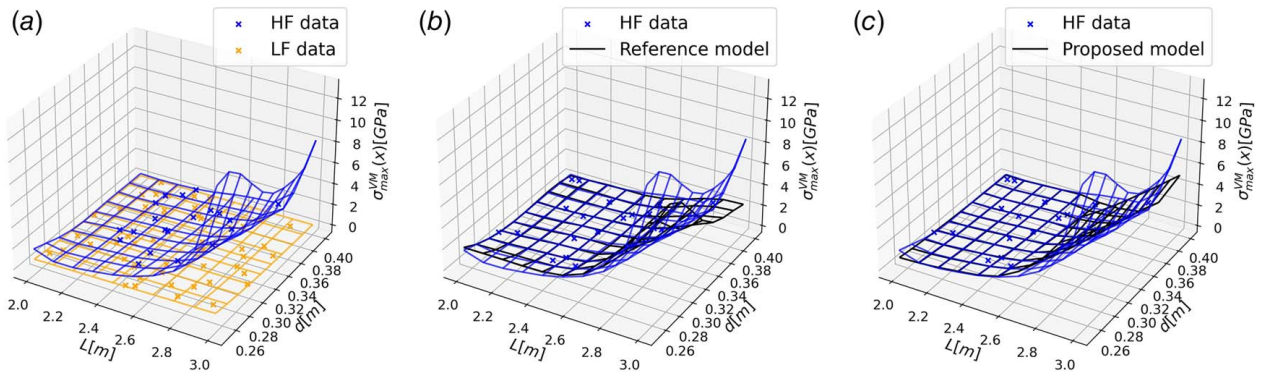


Fig. 10 2D structural problem of a cantilever beam using 20 HF data and 50 HF data: (a) design spaces for the cantilever beam, (b) reference model, and (c) proposed model

reaching up to 22% in the case of 15 HF points. Similarly, the 1D heterogeneous function case study followed a modeling approach akin to the 1D Jump Forrester scenario. The proposed model exhibited improvement, with predictions reaching up to 32% in the case of 15 HF points. The 2D shifted-rotated Rastrigin function, employed to assess multimodal behavior, was modeled both as a bi-fidelity and tri-fidelity problem. In the bi-fidelity scenario, the HF points ranged from 5 to 30, while LF points remained constant at 100. An improvement of 62% was observed with 30 HF points. In the tri-fidelity case, HF points ranged from 5 to 30, medium-fidelity points were held constant at 50, and LF points remained constant at 100. In this case, the improvement was measured 49% for the majority of the cases. For the cantilever beam problem, the reference model produced results comparable to the SF GP. However, the proposed model demonstrated improved results, showing enhancements ranging from 15% to 22%. Overall, these results hold promise for the application of the model in addressing complex design problems within multidimensional spaces.

5.2 Critical Reflection on Scaling-Up the Method to Address Early-Ship Design of Complex Vessels. The objective of the proposed method is to facilitate early-stage design exploration of complex vessels. While the presented case studies involved a lower complexity level, it is crucial to critically reflect on the scalability of the method. The authors believe that the method has undergone extensive testing in various analytical case studies, which simulate challenges sharing similarities with real design problems, such as the presence of discontinuities and complex patterns. Furthermore, the chosen case studies demonstrate good alignment with benchmark problems that hold wide acceptance within the research design community. An important consideration when applying the suggested method to high-dimensional realistic design problems is the increased computational costs associated with the development of the compositional kernels. The associated computational cost depends on the dimensionality of the problem, the size of the training set, and the number of analysis methods used in the MF model. The latter is not of interest because inherently these design problems deal with small data regimes. The

impact of problem dimensionality was explored in the ND Rosenbrock case study. Additionally, the effect of integrating additional fidelity models was examined in the case of the Rastrigin function. Overall, the integration of compositional kernels introduces a trade-off between the computational benefits arising from reduced training dataset sizes and the supplementary computational expenses stemming from kernel optimization. The determining factor in this trade-off is contingent upon the nature of the design problem. Particularly in scenarios characterized by design exploration tasks featuring KPIs that are expensive to evaluate, we assert that the integration of compositional kernels presents a promising avenue.

The expansion of the method to tackle high-dimensional problems will inevitably result in increased computational expenses, presenting a notable challenge. However, it is important to note that HF analysis techniques in ship design problems such as computational fluid dynamics (CFD) analysis and model tests can be significantly (computationally) expensive. Therefore, the authors believe that the application of the proposed method and the subsequent reduction in the required number of HF simulations will yield computational benefits for design space exploration problems.

5.3 Recommendations for Future Research. Several suggestions for future research efforts are worth considering. First, the method should be tested to real engineering engineering problems to ensure its capability to effectively capture the complex patterns inherent in real multidimensional design spaces. In the context of real-world applications, another significant consideration involves evaluating the enhancements in accuracy and the reduced uncertainty in predicting the design space. In addition, it is important to assess whether the computational benefits stemming from the reduced necessity for HF simulations are offset by the computational costs associated with optimizing the kernel function.

Acknowledgment

This publication is part of the project “Multi-fidelity Probabilistic Design Framework for Complex Marine Structures” (project number TWM.BL.019.007) of the research program “Topsector

Water & Maritime: The Blue Route,” which is (partly) financed by the Dutch Research Council (NWO). The authors thank DAMEN, the Netherlands Defence Materiel Organisation (DMO), and MARIN for their contribution to the research.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

- [1] Andrews, D., 2018, “The Sophistication of Early Stage Design for Complex Vessels,” *Inter. J. Maritime Eng.*, **160**(18), p. 12.
- [2] Mavris, D., DeLaurentis, D., Bandte, O., and Hale, M., 1998, “A Stochastic Approach to Multi-disciplinary Aircraft Analysis and Design,” *36th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, Jan. 12–15.
- [3] Papanikolaou, A., 2014, *Ship Design: Methodologies of Preliminary Design*, Springer, The Netherlands.
- [4] Horvath, B. L., and Wells, D. P., 2018, “Comparison of Aircraft Conceptual Design Weight Estimation Methods to the Flight Optimization System,” *2018 AIAA Aerospace Sciences Meeting*, Kissimmee, FL, Jan. 8–12.
- [5] Chaal, M., Ren, X., BahooToroody, A., Basnet, S., Bolbot, V., Banda, O. A., and Gelder, P. V., 2023, “Research on Risk, Safety, and Reliability of Autonomous Ships: A Bibliometric Review,” *Safety Sci.*, **167**, p. 106256.
- [6] van Essen, S., and Seyffert, H., 2022, “Finding Dangerous Waves – Towards an Efficient Method to Obtain Wave Impact Design Loads for Marine Structures,” *Proceedings of the ASME 2022 41st International Conference on Ocean, Offshore and Arctic Engineering*, Volume 1: Offshore Technology, Hamburg, Germany, June 5–10.
- [7] Geertsma, R. D., Negenborn, R. R., Visser, K., and Hopman, J. J., 2017, “Design and Control of Hybrid Power and Propulsion Systems for Smart Ships: A Review of Developments,” *Appl. Energy*, **194**, pp. 30–54.
- [8] Kim, B. S., Park, D. M., and Kim, Y., 2022, “Study on Nonlinear Heave and Pitch Motions of Conventional and Tumblehome Hulls in Head Seas,” *Ocean Eng.*, **247**, p. 110671.
- [9] Fuentes, D., Salas, M., Tampier, B. G., and Troncoso, C., 2015, “Structural Design and Optimisation of an Aluminium Trimaran,” *Analysis and Design of Marine Structures V*, G. Soares, and Sheno, eds., CRC Press, Boca Raton, FL, pp. 549–558.
- [10] Rijkens, A.A.K., and Mikelic, A., 2022, “The Hydrodynamic Comparison Between a Conventional and an Axe Bow frigate Hull,” *16th International Naval Engineering Conference and Exhibition incorporating the International Ship Control Systems Symposium*, Delft, Netherlands, Nov. 8–10.
- [11] Hovilainen, M., Romu, T., Heinonen, T., and Uuskallio, A., 2016, “Initial Operational Experience From the Oblique Icebreaker,” *All Days, OTC*, St. John’s, Newfoundland and Labrador, Canada, Oct. 24–26.
- [12] Oosterom, W., and Vos, R., 2022, “Conceptual Design of a Flying-V Aircraft Family,” *AIAA AVIATION 2022 Forum*, Chicago, IL & Virtual, June 27–July 1.
- [13] Charisi, N. D., Hopman, H., and Kana, A., 2022, “Early-Stage Design of Novel Vessels: How Can We Take a Step Forward?” *14th International Marine Design Conference*, Vancouver, Canada, June 26–30.
- [14] Bickel, J. E., and Bratvold, R. B., 2008, “From Uncertainty Quantification to Decision Making in the Oil and Gas Industry,” *Energy Exploration Exploit.*, **26**(5), pp. 311–325.
- [15] Mainini, L., Serani, A., Rumpfkeil, M. P., Minisci, E., Quagliarella, D., Pehlivan, H., Yildiz, S., Ficini, S., Pellegrini, R., Di Fiore, F., Bryson, D., Nikbay, M., Diez, M., and Beran, P., 2022, “Analytical Benchmark Problems for Multifidelity Optimization Methods,” *AVT-354 Multi-Fidelity Methods for Military Vehicle Design*, Varna, Bulgaria, Sept. 26–28.
- [16] Parunov, J., Guedes Soares, C., Hirdaris, S., Iijima, K., Wang, X., Brizzolara, S., Qiu, W., Mikulić, A., Wang, S., and Abdelwahab, H. S., 2022, “Benchmark Study of Global Linear Wave Loads on a Container Ship With Forward Speed,” *Marine Struct.*, **84**, p. 103162.
- [17] Decker, K., and Mavris, D., 2022, “Modeling Hypersonic Vehicle Performance and Operations Using a Multi-Fidelity Reduced Order Modeling Approach,” *AVT-354 Multi-Fidelity Methods for Military Vehicle Design*, Varna, Bulgaria, Sept. 26–28.
- [18] Papageorgiou, A., Tarkian, M., Amadori, K., and Ölvander, J., 2018, “Multidisciplinary Design Optimization of Aerial Vehicles: A Review of Recent Advancements,” *Inter. J. Aerosp. Eng.*, **2018**, p. 4258020.
- [19] Beran, P. S., Bryson, D. E., Thelen, A. S., Diez, M., and Serani, A., 2020, “Comparison of Multi-fidelity Approaches for Military Vehicle Design,” *AIAA Aviation 2020 Forum*, Virtual Event, June 15–19.
- [20] Fernández-Godino, M. G., Park, C., Kim, N.-H., and Haftka, R. T., 2016, “Review of Multi-fidelity Models,” *AIAA J.*, **57**(5), pp. 2039–2054.
- [21] Peherstorfer, B., Willcox, K., and Gunzburger, M., 2018, “Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization,” *SIAM Review*, **60**(3), pp. 550–591.
- [22] Raissi, M., Perdikaris, P., and Karniadakis, G. E., 2016, “Inferring Solutions of Differential Equations Using Noisy Multi-fidelity Data,” *J. Comput. Phys.*, **335**, pp. 736–746.
- [23] Drummen, I., Hageman, R. B., and Stambaugh, K., 2022, “Multifidelity Approach for Predicting Extreme Global Bending Load Effects,” *9th International Conference on Hydroelasticity in Marine Technology*, Rome, Italy, June 10–13.
- [24] Guth, S., Champenois, B., and Sapsis, T. P., 2022, “Application of Gaussian Process Multi-fidelity Optimal Sampling to Ship Structural Modeling,” *34th Symposium on Naval Hydrodynamics*, Washington, DC, June 26–July 1.
- [25] Ng, L. W. T., and Willcox, K. E., 2015, “Monte Carlo Information-Reuse Approach to Aircraft Conceptual Design Optimization Under Uncertainty,” *J. Aircraft*, **53**(2), pp. 1–12.
- [26] Ng, L. W., and Willcox, K. E., 2014, “Multifidelity Approaches for Optimization Under Uncertainty,” *Inter. J. Num. Methods Eng.*, **100**(10), pp. 746–772.
- [27] Meng, X., and Karniadakis, G. E., 2020, “A Composite Neural Network that Learns From Multi-fidelity Data: Application to Function Approximation and Inverse PDE Problems,” *J. Comput. Phys.*, **401**, p. 109020.
- [28] Pawar, S., San, O., Vedula, P., Rasheed, A., and Kvamsdal, T., 2022, “Multi-fidelity Information Fusion With Concatenated Neural Networks,” *Sci. Rep.*, **12**(1), p. 5900.
- [29] Guo, M., Manzoni, A., Amendt, M., Conti, P., and Hesthaven, J. S., 2022, “Multi-fidelity Regression Using Artificial Neural Networks: Efficient Approximation of Parameter-Dependent Output Quantities,” *Comput. Methods Appl. Mech. Eng.*, **389**, p. 114378.
- [30] Meng, X., Babae, H., and Karniadakis, G. E., 2021, “Multi-fidelity Bayesian Neural Networks: Algorithms and Applications,” *J. Comput. Phys.*, **438**, p. 110361.
- [31] Bonfiglio, L., Perdikaris, P., and Karniadakis, G., 2018, “A Probabilistic Framework for Multidisciplinary Design: Application to the Hydrostructural Optimization of Supercavitating Hydrofoils,” *Inter. J. Num. Methods Eng.*, **116**(4), pp. 246–269.
- [32] Brooks, T. R., Kenway, G. K., and Martins, J. R., 2017, “Undelected Common Research Model (uCRM): An Aerostructural Model for the Study of High Aspect Ratio Transport Aircraft Wings,” *35th AIAA Applied Aerodynamics Conference*, Denver, CO, June 5–9.
- [33] Di Fiore, F., Maggiore, P., and Mainini, L., 2021, “Multifidelity Domain-aware Learning for the Design of Re-entry Vehicles,” *Structural and Multidisciplinary Optimization*, **64**(23), pp. 1–19.
- [34] Singh, V., and Willcox, K. E., 2021, “Decision-Making Under Uncertainty for a Digital Thread-Enabled Design Process,” *ASME J. Mech. Des.*, **143**(9), p. 091707.
- [35] Melati, D., Grinberg, Y., Kamandar Dezfouli, M., Janz, S., Cheben, P., Schmid, J. H., Sánchez-Postigo, A., and Xu, D. X., 2019, “Mapping the Global Design Space of Nanophotonic Components Using Machine Learning Pattern Recognition,” *Nat. Commun.*, **10**(1), pp. 1–9.
- [36] Moss, H. B., Beck, D., Gonzalez, J., Leslie, D. S., and Rayson, P., 2020, “BOSS: Bayesian Optimization Over String Spaces,” *Adv. Neural Inform. Process. Syst.*, **2020**(12).
- [37] Palar, P. S., Zuhail, L. R., and Shimoyama, K., 2020, “Gaussian Process Surrogate Model With Composite Kernel Learning for Engineering Design,” *AIAA J.*, **58**(4), pp. 1864–1880.
- [38] Hulse, D., Hoyle, C., Tumer, I. Y., and Goebel, K., 2021, “How Uncertain Is Too Uncertain? Validity Tests for Early Resilient and Risk-Based Design Processes,” *ASME J. Mech. Des.*, **143**(1), p. 011702.
- [39] O’Hagan, T., 2004, “Dicing With the Unknown,” *Significance*, **1**(3), pp. 132–133.
- [40] North, D. W., 2017, *Decision Analytic and Bayesian Uncertainty Quantification for Decision Support*, Springer International Publishing, Cham, pp. 1361–1399.
- [41] Collette, M., 2017, *Uncertainty Approaches in Ship Structural Performance*, Springer International Publishing, Cham, pp. 1567–1588.
- [42] Aughenbaugh, J. M., and Paredis, C. J., 2006, “The Value of Using Imprecise Probabilities in Engineering Design,” *ASME J. Mech. Des.*, **128**(4), pp. 969–979.
- [43] Rasmussen, C. E., and Williams, C. K. I., 2006, *Gaussian Processes for Machine Learning*, Vol. 3176, The MIT Press, p. 63.
- [44] Kennedy, M. C., and O’Hagan, A., 2000, “Predicting the Output From a Complex Computer Code When Fast Approximations Are Available,” *Biometrika*, **87**(1), pp. 1–13.
- [45] Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N. D., and Karniadakis, G. E., 2017, “Nonlinear Information Fusion Algorithms for Data-Efficient Multi-fidelity Modelling,” *Proc. R. Soc. A: Math., Phys. Eng. Sci.*, **473**(2198), p. 20160751.
- [46] Damianou, A. C., and Lawrence, N. D., 2012, “Deep Gaussian Processes,” *J. Mach. Learn. Res.*, **31**, pp. 207–215.
- [47] Feldstein, A., Lazzara, D., Princen, N., and Willcox, K., 2020, “Multifidelity Data Fusion: Application to Blended-Wing-Body Multidisciplinary Analysis Under Uncertainty,” *AIAA J.*, **58**(2), pp. 889–906.
- [48] Scholcz, T., and Klinkenberg, J., 2022, “Hull-Shape Optimisation Using Adaptive Multi-Fidelity Kriging,” *AVT-354 Multi-Fidelity Methods for Military Vehicle Design*, Varna, Bulgaria, Sept. 26–28.
- [49] Bessa, M. A., Glowacki, P., and Houlder, M., 2019, “Bayesian Machine Learning in Metamaterial Design: Fragile Becomes Supercompressible,” *Adv. Mater.*, **31**(48), p. 1904845.

- [50] Nitzler, J., Biehler, J., Fehn, N., Koutsourelakis, P. S., and Wall, W. A., 2022, "A Generalized Probabilistic Learning Approach for Multi-fidelity Uncertainty Quantification in Complex Physical Simulations," *Comput. Methods. Appl. Mech. Eng.*, **400**, p. 115600.
- [51] Brevault, L., Balesdent, M., and Hebbal, A., 2020, "Overview of Gaussian Process Based Multi-fidelity Techniques With Variable Relationship Between Fidelities, Application to Aerospace Systems," *Aeros. Sci. Tech.*, **107**, p. 106339.
- [52] Charisi, N. D., Kana, A., and Hopman, H., 2022, "Compositional Kernels to Facilitate Multi-fidelity Design Analysis: Applications for Early-Stage Design," AVT-354 Multi-Fidelity Methods for Military Vehicle Design, Varna, Bulgaria, Sept. 26–28
- [53] Bachoc, F., 2013, "Cross Validation and Maximum Likelihood Estimations of Hyper-parameters of Gaussian Processes With Model Misspecification," *Comput. Stat. Data Anal.*, **66**, pp. 55–69.
- [54] Le Gratiet, L., and Garnier, J., 2014, "Recursive Co-Kriging Model for Design of Computer Experiments With Multiple Levels of Fidelity," *Inter. J. Uncertainty Quantifi.*, **4**(5), pp. 365–386.
- [55] Duvenaud, D. K., 2014, "Automatic Model Construction With Gaussian Processes," Ph.D. thesis, University of Cambridge.
- [56] Duvenaud, D., Lloyd, J. R., Grosse, R., Tenenbaum, J. B., and Ghahramani, Z., 2013, "Structure Discovery in Nonparametric Regression Through Compositional Kernel Search," Proceedings of the 30th International Conference on Machine Learning, PMLR 28(3), pp. 1166–1174.
- [57] GPy, 2012, "GPy—A Gaussian Process (GP) Framework in Python," <https://github.com/SheffieldML/GPy>
- [58] GitHub - EmuKit/emukit, "A Python-Based Toolbox of Various Methods in Decision Making, Uncertainty Quantification and Statistical Emulation: Multi-fidelity, Experimental Design, Bayesian Optimisation, Bayesian Quadrature, etc," <https://github.com/EmuKit/emukit>
- [59] James, G., Witten, D., Hastie, T., and Tibshirani, R., 2014, *An Introduction to Statistical Learning: with Applications in R*, Springer Publishing Company, Inc.
- [60] Spiess, A. N., and Neumeyer, N., 2010, "An Evaluation of R2 as an Inadequate Measure for Nonlinear Models in Pharmacological and Biochemical Research: A Monte Carlo Approach," *BMC Pharmacol.*, **10**(1), pp. 1–11.
- [61] Cameron, A. C., and Windmeijer, F. A., 1997, "An R-squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models," *J. Econ.*, **77**(2), pp. 329–342.
- [62] Liu, H., Ong, Y.-S., Shen, X., and Cai, J., 2020, "When Gaussian Process Meets Big Data: A Review of Scalable GPs," *IEEE Trans. Neural Netw. Learn. Syst.*, **31**(11), pp. 4405–4423.
- [63] Öchsner, A., 2021, *Classical Beam Theories of Structural Mechanics*, Springer.