


Multiobjective optimization for flapping foil hydrodynamics with a multitask and multifidelity approach

Zhangyuan Wang^{ⓧ*} and Dehan Yuan^{ⓧ*}
Zhejiang University, Hangzhou 310027, China

Chenglong Wu
College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

Xu Chen
China Ship Scientific Research Center, Wuxi 214082, China

Ruipeng Li^{ⓧ,†}, Weicheng Cui^{ⓧ,‡} and Dixia Fan[§]
Key Laboratory of Coastal Environment and Resources of Zhejiang Province, School of Engineering, Westlake University, Hangzhou 310030, China

 (Received 25 June 2023; revised 21 November 2023; accepted 15 December 2023; published 23 January 2024)

We develop a multitask and multifidelity Gaussian process (MMGP) model to accurately predict and optimize the multiobjective performance of a flapping foil while minimizing the cost of high-fidelity data. Through a comparison of three kernels, we have selected and applied the spectral mixture kernel and validated the robustness and effectiveness of a multiacquisition function. To effectively incorporate data with varying levels of fidelity, we have adopted a linear prior formula-based multifidelity framework. Additionally, Bayesian optimization with a multiacquisition function is adopted by the MMGP model to enable multitask active learning. The results unequivocally demonstrate that the MMGP model serves as a highly capable and efficient framework for effectively addressing the multiobjective challenges associated with flapping foils.

DOI: [10.1103/PhysRevE.109.015103](https://doi.org/10.1103/PhysRevE.109.015103)

I. INTRODUCTION

Flapping their fins and wings, birds and fish have attained the remarkable capacity to generate lift and thrust [1]. Consequently, there is an increasing interest in investigating the formation and characteristics of unsteady flow and fluid forces surrounding flapping foils, with particular emphasis on determining the optimal motion parameters to achieve hydrodynamic objectives [2–6].

The motion of the flapping foil is effectively simplified through sinusoidal pitching and heaving motions [7]. The Strouhal number (St) is a crucial dimensionless parameter that has a significant influence [5]. Overall, the thrust increases with an increase of St , while the propulsion efficiency η demonstrates a distinct peak within a narrow St

range ($0.2 < St < 0.4$), reaching impressive levels of up to 80%. In addition, by optimizing trajectories, the flapping foil performance can be further enhanced [8]. As a result, the flapping foil demonstrates the capability to accomplish multiobjective tasks (e.g., balance between propulsive efficiency and thrust, creating thrust with a small lifting force analogous to swimmers, and creating large lifting forces similar to flyers with low speed) with multimodal parametric motions [9,10]. Detailed physics can be found in the following reviews, Refs. [11–14].

In the pursuit of multiobjective optimization for flapping foils, it is imperative to accurately predict the relationship between the foil's performance and the prescribed trajectories, which requires the use of high-fidelity simulations or experiments. Previous researchers have successfully harnessed the potential of the genetic algorithm (GA), leading to noteworthy advancements in both thrust and propulsion efficiency for the flapping foil [10,15]. For flapping foils, the high-dimensional problem of systematic exploration is intractable, which gives the GA an advantage in the preliminary stage. However, the GA is limited to optimization and cannot model and predict the system. It needs to be reoptimized whenever different optimization objectives are changed, resulting in most time being spent on data generation. On the other hand, machine learning and artificial intelligence methods provide innovative approaches for optimizing the geometry and controlling the motion of flapping foils. Nevertheless, these

*Also at School of Engineering, Westlake University, Hangzhou 310030, China.

†Also at Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou 310024, China, liruipeg@westlake.edu.cn

‡Also at Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou 310024, China.

§Also at Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou 310024, China, fandixia@westlake.edu.cn

methods heavily rely on time-varying data to effectively capture the evolving characteristics of the system. To address this challenge, the field of machine learning has introduced multifidelity modeling, which employs a probabilistic regression framework to seamlessly integrate information from diverse sources with varying levels of fidelity. This approach enables the synthesis of valuable insights from different data sources, enhancing the accuracy and robustness of the optimization and control processes for computational physics, transportation, and structural optimization [16–18].

The construction of a multifidelity model using the Gaussian process (GP) holds great promise and has found applications in various fluid mechanics domains [19–22], including the optimization of trajectories and shapes of single flapping foil [9,23–25], as well as the optimization of trajectories for tandem flapping wings [26]. Bayesian optimization is a powerful strategy for function optimization that specifically benefits from the uncertainty estimates provided by the GP [27]. A key strength of the GP is the inherent ability to provide predictions with quantified uncertainty. This uncertainty quantification proves to be of great value in active learning. In GP-based active learning, this uncertainty is harnessed by Bayesian optimization using acquisition functions to wisely determine where to sample next. This usually results in a more efficient sampling strategy. In contrast, although the deep neural networks have advantages in dealing with large-scale data sets and complex function mappings, they lack the innate ability to quantify prediction uncertainties. This makes them rely on more detailed and often uniform sampling strategies, which may be inefficient, especially when dealing with limited or costly data such as high-precision simulations and experiments. Combined with active learning, the adaptive strategy provided by the GP allows the use of relatively small but more informative data sets to improve model accuracy. However, the previous research focused on the analysis of the hydrodynamic performance of flapping foils, ignoring the exploration of the multifidelity method. In addition, it did not consider the cooperative training of multioutputs in the process of multifidelity training, so it is difficult to tackle the multiobjective optimization problem of flapping foils.

This study aims to develop a multitask and multifidelity Gaussian process (MMGP) model for accurate prediction and optimization of the multiobjective hydrodynamic performance. This is achieved by integrating low-fidelity response surfaces with a minimal number of high-fidelity data points within an information-fusion framework. The generation of numerous low-fidelity samples is accomplished through various approaches, including data fitting, simplified physics-based models, and numerical simulations conducted on coarser grids or with looser criteria, akin to a hierarchical proxy model [28]. Moreover, the blackbox matrix-matrix Gaussian process inference [29] is employed at different fidelity levels to provide a predicted posterior distribution, enabling the inference of relevant physical quantities along with quantitative uncertainty estimation. To effectively balance exploration (local search) and exploitation (global search) of the design space, the Bayesian optimization algorithm with a multiacquisition function (multi-AF) is utilized. Furthermore, the optimization process considers the accurate modeling of each subtask GP model while addressing the multiobjective

nature of the problem, thus ensuring a balanced representation of different accuracy levels.

The structure of this paper is organized as follows: Section II is divided into six subsections, which include the description of the physical and numerical flapping foil models, the fundamentals of the singleGP model, the proposed comprehensive MMGP model, and the multi-AF framework for the multiobjective task. Section III provides an analysis of the hydrodynamic performance of the flapping foil based on the high-fidelity ergodic database. In Sec. III and the Appendix, a comparison and evaluation of various kernels and acquisition functions are presented, focusing on their mean absolute error (MAE) and their ability to predict optimal values for the foil optimization problem. The effectiveness of the MMGP model is then validated through a comparison with a traditional multifidelity framework. Additionally, the results of the multiobjective optimization using the MMGP model are demonstrated through a practical implementation. Finally, the paper concludes with a summary of findings and implications in Sec. IV.

II. MATERIALS AND METHODS

In this section, physical and numerical models for the flapping foil are first set up. The framework of the MMGP model is then established and also how the predictions from two different levels of fidelity are combined. This gives a multifidelity stochastic response surface for optimizing the kinematic parameter of the flapping foil.

A. Physical model

The motion of a flapping foil NACA0015 with the foil thickness L can be characterized by several parameters. The nondimensional Strouhal number St , denoting the imposed frequency, is defined as

$$St = \frac{fL}{U}, \quad (1)$$

where f and U are the oscillation frequency and the inflow velocity, respectively. The Reynolds number $Re = \rho U c / \mu$ takes 500 in this paper, where ρ and μ denote the density and dynamic viscosity of the fluid, and c is the chord length as shown in Fig. 1(b).

The flapping foil motion is usually designed by a coupling of heaving and pitching with the same circular frequency $\omega = 2\pi f$, which are defined by.

$$\begin{aligned} y(t) &= y_0 \sin(\omega t), \\ \theta(t) &= \theta_0 \sin(\omega t + \psi), \end{aligned} \quad (2)$$

where ψ is the phase difference, and y_0 and θ_0 are the heaving and pitching amplitudes, respectively. In the present study, the Strouhal number St , the heaving amplitude y_0 , the pitching amplitude θ_0 , and the phase difference ψ are chosen as independent motion parameters of the flapping foil, with the sketch shown in Fig. 1.

In general, the mean thrust coefficient C_T , the mean input power coefficient C_p , and the propulsion efficiency η are used to describe the performance of the flapping foil. The above

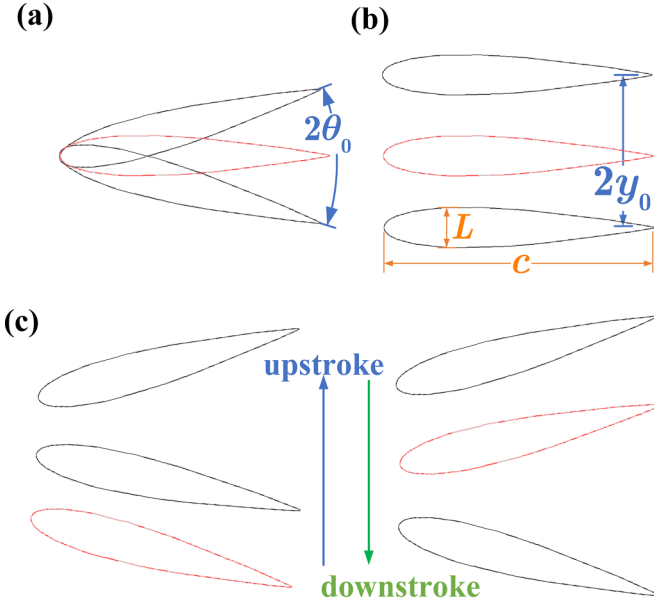


FIG. 1. Schematics of (a) pitching motion, (b) heaving motion, and (c) combined-heaving-and-pitching motion of a foil.

three nondimensional parameters are defined as follows:

$$\begin{aligned}
 C_T &= \frac{\bar{F}_x}{0.5\rho U^2 cL}, & \bar{F}_x &= \frac{1}{T} \int_0^T F_x(t) dt, \\
 C_p &= \frac{\bar{P}}{0.5\rho U^3 cL}, \\
 \bar{P} &= \frac{1}{T} \left[\int_0^T F_y(t) \dot{y}(t) dt + \int_0^T M_\theta(t) \dot{\theta}(t) dt \right], \\
 \eta &= \frac{C_T}{C_p},
 \end{aligned} \tag{3}$$

where $F_x(t)$ and $F_y(t)$ are the instantaneous forces in the x and y directions, $M_\theta(t)$ is the instantaneous torque around the pitching point, $\dot{y}(t)$ and $\dot{\theta}(t)$ are the first-order derivatives of the heaving and pitching motions, and T is the motion period with $T = 2\pi/\omega$. Given the relationship between the three parameters, C_T and η are chosen to describe the performance of the flapping foil in this paper.

B. Numerical model

The primary objective of this study is to establish an efficient framework for the accurate prediction and optimization of foil performance. In pursuit of this goal, the simulation platform we employed is the Lilypad, which is specifically designed based on the boundary data immersion method (BDIM). The BDIM method exhibits remarkable capabilities in simulating the entire domain with quadratic convergence by effectively solving the viscous time-dependent Navier-Stokes equations. This is achieved through the utilization of a kernel function that seamlessly integrates the dynamics of both the moving body and the surrounding fluid. The effectiveness and reliability of the BDIM method have been extensively verified across a diverse range of fluid problems,

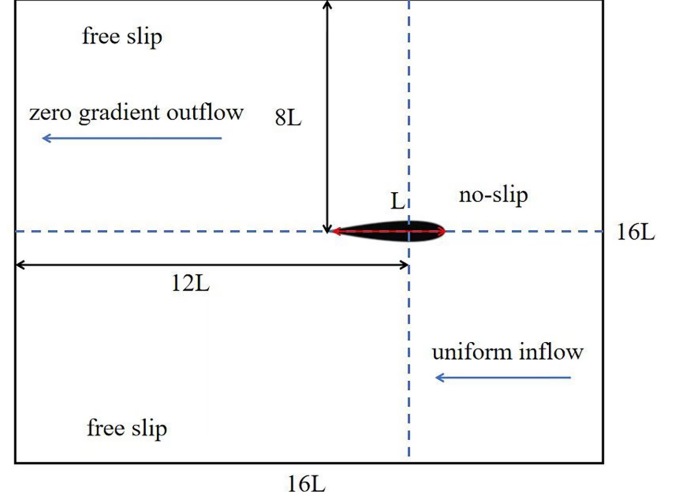


FIG. 2. Sketch of the calculation domain.

employing a comprehensive array of numerical simulations [30–32].

In our study, we utilize a multifidelity approach by leveraging both high- and low-fidelity data. High-fidelity data are obtained from simulations using finer grids, ensuring higher accuracy but also incurring greater computational costs. In contrast, low-fidelity data come from simulations with coarser grids, which provide faster results but reduce accuracy. The mesh configuration consists of dense uniform grids near the body and in the immediate wake, transitioning to exponentially stretching grids in distant regions. The simulation domain, under a right-to-left uniform inflow, incorporates a zero-gradient outflow, free-slip on the top and bottom, and no-slip condition on the foil surface, as shown in Fig. 2. For high-fidelity simulations, the domain size is set at $16L \times 16L$, with a grid size of $\delta x = \delta y = L/32$. These simulations take about 0.33 h on average for computation. Low-fidelity simulations reduce the domain size to $8L \times 8L$ and use a grid size of $\delta x = \delta y = L/8$, reducing the computation time to an average of 8.2 s. Grid convergence tests were implemented and are presented in Table I. All simulations were performed on hardware configurations with an Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz processor, 16.0 GB of RAM, and a NVIDIA GeForce GT 730 graphics card.

TABLE I. Grid convergence tests with $St = 0.25$, $y_0 = 1$, $\theta_0 = 30^\circ$, $\psi = -\pi/2$, and $Re = 1100$.

Resolution	Domain size	C_T	Time
8	$8L \times 8L$	0.181	8.2 s
32	$16L \times 16L$	1.048	0.33 h
32	$32L \times 32L$	1.061	0.53 h
64	$16L \times 16L$	1.252	2.41 h
64	$32L \times 32L$	1.275	4.58 h
80	$16L \times 16L$	1.271	4.47 h
80	$32L \times 32L$	1.296	9.83 h

C. Gaussian process regression model

Single-fidelity GP (SF-GP). A Gaussian process (GP) is a stochastic process that consists of random variables with a joint Gaussian distribution. A GP is fully characterized by its mean function $m(x)$ and covariance function $k(x, x')$, which describe the expected value and the correlation of the outputs for any two inputs x and x' . A common choice for the mean function is zero, and a common choice for the covariance function is the radial basis function (RBF) kernel with the lengthscale Θ and the variance σ^2 :

$$k_{\text{RBF}}(x, x') = \exp\left(-\frac{1}{2}(x - x')^\top \Theta^{-2}(x - x')\right). \quad (4)$$

Given a set of training data (X, y) , where X is an $n \times d$ matrix of inputs and y is an $n \times 1$ vector of outputs, we can use the GP to make predictions for a new input x_* by using the conditional distribution of y_* given (X, y) and x_* , which is also Gaussian with mean and variance:

$$\begin{aligned} \mu(x_*) &= k(x_*, X)[K(X, X) + \sigma_n^2 I]^{-1} y, \\ \sigma^2(x_*) &= k(x_*, x_*) - k(x_*, X)[K(X, X) + \sigma_n^2 I]^{-1} k(X, x_*), \end{aligned} \quad (5)$$

where $K(X, X)$ is the Gram matrix with entries $K_{ij} = k(x_i, x_j)$, σ_n^2 is the noise variance, and I is the identity matrix [33].

To fit a GP to the data, we need to choose appropriate mean and covariance functions and optimize their hyperparameters, which influence the shape and smoothness of the functions. Different methods to optimize hyperparameters include heuristic algorithms (e.g., GA and particle swarm optimization) and stochastic gradient descent (SGD). Heuristic algorithms are global optimization methods that avoid local optima, but they are computationally expensive and require Cholesky-based approaches for matrix operations. SGD is a local optimization method that finds a solution quickly by using conjugate gradients for matrix-matrix multiplications, though it may be prone to local optima and instability. Therefore, we implement GP regression using the GPYTORCH software platform [29], which employs SGD and CG for hyperparameter optimization. This method is advantageous because it offers faster computation than a heuristic algorithm and can deal with data of large size and dimensionality.

Multitask GP. We use the multitask GP model to build a multioutput model that can predict both thrust and efficiency. The multitask GP model is a type of GP model that can handle multiple tasks (e.g., thrust and efficiency) simultaneously by exploiting the shared information and the specific information of each task. Given inputs x and x' , and tasks i and j , the covariance between two data points and two tasks is

$$k([x, i], [x', j]) = k_{\text{inputs}}(x, x') * k_{\text{tasks}}(i, j), \quad (6)$$

where k_{inputs} is a standard kernel (e.g., RBF) that operates on the inputs, and k_{tasks} is a lookup table containing intertask covariance.

There are different ways to construct a multitask GP model, such as the correlated multitask GP, the independent multitask GP, and the Hadamard multitask GP [34]. The correlated multitask GP assumes that the tasks are correlated and models the cross-covariance between them. The independent multi-

task GP assumes that the tasks are independent and models each task separately. The Hadamard multitask GP assumes that the tasks are conditionally independent given some latent variables and models the joint distribution of the tasks and the latent variables.

Multifidelity GP. The multifidelity GP is a type of GP model that can handle data from different sources of varying fidelity (e.g., low-fidelity simulations and high-fidelity experiments). The idea is to combine the information from different sources in a coherent way and leverage the cheaper low-fidelity data to improve the accuracy and efficiency of the high-fidelity model. A common way to construct a multifidelity GP model is to use a linear relationship between the outputs of different sources, such as

$$f_2(x) = \hat{\rho} f_1(x) + \delta_2(x), \quad (7)$$

where f_1 and f_2 symbolize the low-fidelity and high-fidelity models, respectively. $\hat{\rho}$ acts as a scaling factor, describing the slope of the relationship between the high-fidelity and the low-fidelity outputs. Although in many cases, $\hat{\rho}$ can be a function that depends on the input x , it is seen as a deterministic scalar that is learned from the data via maximum-likelihood estimation in the present method. The term δ_2 represents the residual, encapsulating the discrepancies between the two models. This residual can be further modeled using another GP, equipped with distinct mean and covariance functions. Consequently, the multifidelity GP model becomes proficient at offering predictions for both fidelity models at any input point, by leveraging the joint distribution of f_1 and f_2 .

For the Gaussian process regression model, the choice of kernels can highly affect the performance. The kernels in GP are divided into the stationary kernel, the dot product kernel, the neural network kernel, etc. [35]. A stationary kernel is invariant to translations of the inputs and is mostly a function of $\gamma = x - x'$ for any pair of inputs x and x' .

The Fourier transform of popular stationary kernels such as the RBF kernel and the Matern kernel corresponds to a very small corner of the set of possible stationary kernels [36]. While, the spectral mixture (SM) kernel whose spectral density is arbitrary Gaussian mixtures, are much more expressive. It contains many stationary kernels with a simple form [37],

$$k_{\text{SM}}(\gamma) = \sum_{a=1}^A w_a^2 \exp\{-2\pi^2 \gamma^2 \sigma_a^2\} \cos(2\pi \gamma \mu_a). \quad (8)$$

The classic Matern52 kernel is expressed as

$$k_{\text{Matern}}(\gamma) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu}d)^\nu K_\nu(\sqrt{2\nu}d), \quad (9)$$

where $d = \gamma^\top \Theta^{-2} \gamma$. ν is a smoothness parameter (5/2 for Matern52). The SM kernel together with RBF in formula (4) and the Matern kernel are applied and compared for the foil optimization to find the kernel that behaves best. The results are shown in the next section.

D. Multitask GP with multiacquisition function

In each simulation, we observe both thrust and efficiency as outputs for a given input. However, when using two separate SF-GPs to model these two tasks, each model selects the

inputs based on its own acquisition function and only obtains information about its corresponding output. This means that half of the information from each simulation remains unused. Therefore, a more efficient multitask model is required to fully exploit the available information. Additionally, an infill criteria for the multitask model is needed to balance the iteration process.

For active learning based on the Bayesian optimization, a series of points $\bar{x} \in \mathbf{D}$ are sampled and calculated based on the acquisition function, and the top n points are selected to infill and update the model. In the multitask model, how to flexibly keep balance in improving all submodels is vital for the training result. In general, a pair of factors is expected that can adjust itself to add more weights to the less convergent model. A multi-AF with factor α is proposed, and the convergence of two models can be measured by the relative prediction errors m and n . The multi-AF can be expressed as follows:

$$\begin{aligned} \delta(\bar{x}) &= \alpha \frac{\delta_{\text{ct}}(\bar{x})}{\max[\delta_{\text{ct}}(\bar{x})]} + (1 - \alpha) \frac{\delta_{\eta}(\bar{x})}{\max[\delta_{\eta}(\bar{x})]}, \\ \alpha &= \frac{m}{m + n}, \\ m &= \frac{\mathcal{M}_{\text{ct}}}{\max(\hat{f}_{\text{ct}}) - \min(\hat{f}_{\text{ct}})}, \\ n &= \frac{\mathcal{M}_{\eta}}{\max(\hat{f}_{\eta}) - \min(\hat{f}_{\eta})}, \end{aligned} \quad (10)$$

where \hat{f}_{η} and \hat{f}_{ct} are the predicted values by the GP model, \mathcal{M}_{ct} and \mathcal{M}_{η} are the mean absolute errors (represented by \mathcal{M} in mathematical expressions hereafter) that can be estimated by the infill points in the last iteration. Denoting the k th infill point as (\bar{x}^k, f^k) , we have

$$\begin{aligned} \mathcal{M}_{\text{ct}} &= \frac{1}{k} \sum_{i=1}^k (\hat{f}_{\text{ct}}^i - f_{\text{ct}}^i), \\ \mathcal{M}_{\eta} &= \frac{1}{k} \sum_{i=1}^k (\hat{f}_{\eta}^i - f_{\eta}^i). \end{aligned} \quad (11)$$

Generally, $\delta_{\text{ct}}(\bar{x})$ and $\delta_{\eta}(\bar{x})$ can be any acquisition function. The most commonly used acquisition function includes probability of improvement (PI), expected improvement (EI), and upper confidence bound (UCB). The above three acquisition functions are applied to the multi-AF to verify its universality in the following section. The EI, which balances exploitation and exploration, performs better than the other two functions in test results as shown in Figs. 10(a) and 10(b). Therefore, EI is adopted as the priority in the implementation. For more details of EI, see Ref. [38].

E. Multifidelity based on Hadamard multitask GP

With the use of the low-fidelity model mentioned in Sec. II C and the high-fidelity ergodic database in Sec. II A, we can establish the multifidelity model based on the Hadamard multitask GP.

The Hadamard multitask GP is used for each input corresponding to a single task. Except for the standard kernel

$k_{\text{input}}(x, x')$ for the input, it uses an index kernel $k_{\text{task}}(i, j)$ to measure the covariance between tasks.

The correlation for the multifidelity in formula 7 can be equivalently written in the following compact form:

$$\begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} \sim \mathcal{GP} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_1(\cdot) & \hat{\rho}k_1(\cdot) \\ \hat{\rho}k_1(\cdot) & \hat{\rho}^2k_1(\cdot) + k_2(\cdot) \end{bmatrix} \right), \quad (12)$$

where (\cdot) is short for (x, x') . Therefore, the multifidelity model can be constructed based on the Hadamard multitask GP by two designed index kernels:

$$k_{\text{task1}} = \begin{bmatrix} 1 & \hat{\rho} \\ \hat{\rho} & \hat{\rho}^2 \end{bmatrix}, \quad k_{\text{task2}} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \quad (13)$$

The whole covariance in formula (6) between two tasks and inputs yields the following:

$$\begin{aligned} k([x, i], [x', j]) &= k_{\text{input1}}(x, x')k_{\text{task1}}(i, j) + k_{\text{input2}}(x, x')k_{\text{task2}}(i, j) \\ &= k_{\text{input1}}(x, x') \begin{bmatrix} 1 & \hat{\rho} \\ \hat{\rho} & \hat{\rho}^2 \end{bmatrix} + k_{\text{input2}}(x, x') \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned} \quad (14)$$

F. MMGP for multiobjective flapping foil prediction and optimization

Based on the basic work prepared above, the MMGP framework for predicting and optimizing performance of the flapping foil can be established, as shown in Fig. 3. We use a MMGP model to combine multifidelity and multitask GP. The MMGP model consists of two components: an independent multitask GP with a multi-AF and a multifidelity GP based on the Hadamard multitask GP. The independent multitask GP employs the multi-AF to balance the convergence of two separate outputs of the SF-GP model, which are thrust and efficiency, the optimization objectives. Thus the low-fidelity multitask submodel can be constructed by utilizing the low-fidelity data that capture the main trends of the objectives. The multifidelity GP fuses the information from the low-fidelity and high-fidelity multitask submodels, assuming that they are conditionally independent given some latent variables that represent the underlying functions of the objectives.

Prediction and optimization are two main functions for surrogate-based methods. We evaluate the effects of prediction by qualitative analysis of the three-dimensional color maps and quantitative analysis of MAE. The three-dimensional color maps can show the feature that the GP model learns from the training set. Meanwhile, the MAE shows the accuracy of the model prediction with the test set. On the other hand, optimization effects are evaluated by comparing the difference between the prediction point of maximum and the actual maximum point in the ergodic database.

The objective is to predict and optimize the performance by using the least high-fidelity points. Therefore, a well-trained low-fidelity multitask submodel at the beginning of the MMGP training process is applied to assist the high-fidelity submodel converging. The details of the training iteration process are outlined in Algorithm 1.

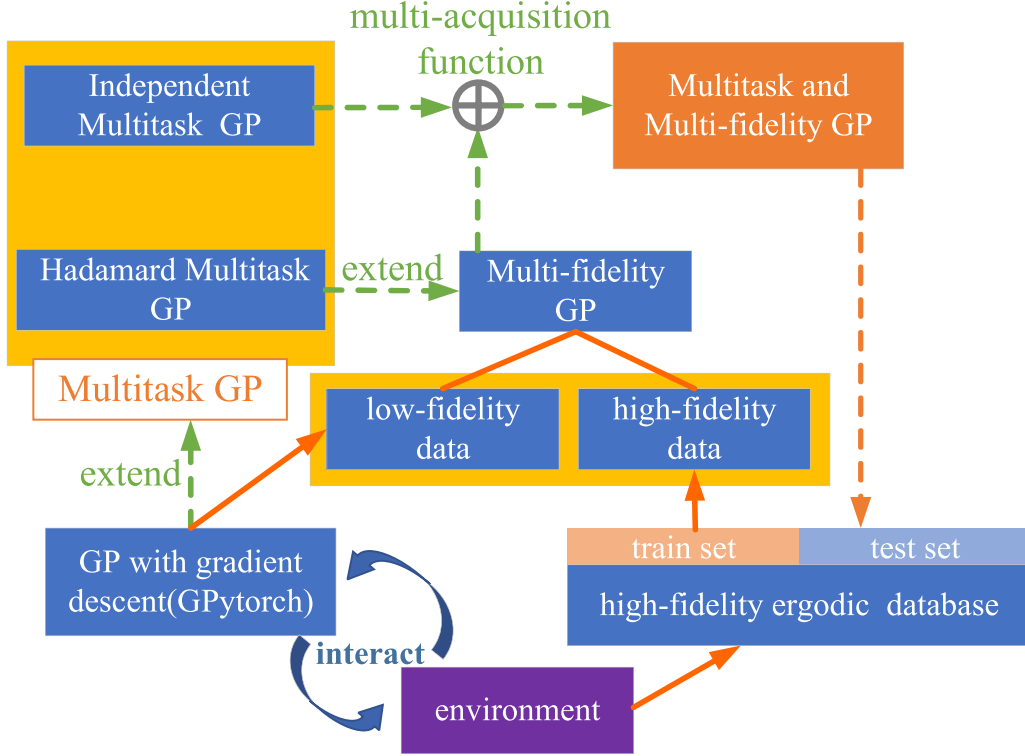


FIG. 3. Composition and workflow of the MMGP framework. The MMGP combines the multifidelity GP based on the Hadamard multitask GP and the independent multitask GP with multi-AF. The Hadamard multitask GP and the independent multitask GP belong to the multitask GP, which is extended from the SF-GP. The multifidelity GP works with low-fidelity data sourced from interacting with the environment and high-fidelity data derived from the high-fidelity ergodic database. The database is divided into the training set and the test set based on whether the data are used by the MMGP model.

The $\mathcal{M}_{\text{loop}}$ is a function of the points chosen at each iteration, which are indirectly influenced by the infill criterion. The infill criterion typically selects points with high uncertainty according to the current model, leading to overestimated $\mathcal{M}_{\text{loop}}$ compared to the actual \mathcal{M} of model. The ratio of estimated values can thus indicate the relative convergence

level of the two models; moreover, if the estimated values converge, the actual $\mathcal{M}_{\text{loop}}$ of the model will also converge.

III. RESULT AND DISCUSSION

A. Hydrodynamic analysis of flapping foil

To validate the efficacy of the MMGP framework in the optimization of flapping foil, a meticulously constructed high-fidelity ergodic database is established, encompassing performance evaluations across the design space with four key parameters. Leveraging Bayesian optimization, the MMGP intelligently selects and incorporates a substantial number of data points from the high-fidelity database to enrich the model. Consequently, the remaining data points serve as a comprehensive test set for robust evaluation. Additionally, meticulous measures are taken to identify and eliminate any outliers present within the high-fidelity database, ensuring the integrity and reliability of the data set.

Once the high-fidelity ergodic database is established, a comprehensive analysis of the maximum thrust and efficiency, as well as the thrust curve and instantaneous flow field, as depicted in Fig. 4, is conducted. Specifically, for $St = 0.26$ and $St = 0.18$, we present five isosurfaces, color-coded by the scaled values of f_{ct} or f_{η} , with the labels y_0 , θ_0 , and ψ representing the heaving amplitude, the pitching amplitude, and the phase angle, respectively. The color bar serves to indicate the magnitude of these values, with blue denoting low values and red denoting high values. Notably, a gradual

Algorithm 1. MMGP for flapping foil.

Require: input $\vec{x} = [St, y_0, \theta_0, \psi]$
output $\vec{f} = [f_{ct}^{\text{low}}, f_{\eta}^{\text{low}}, f_{ct}^{\text{high}}, f_{\eta}^{\text{high}}]$

- 1: Train a single-fidelity multitask GP with i points as low-fidelity of f_{ct}^{low} and f_{η}^{low} .
- 2: Initial sampling j points by latin hypercube sampling for high-fidelity of f_{ct}^{high} and f_{η}^{high} .
- 3: **while** episodes **do**
- 4: **while** iteration **do**
- 5: train $\vec{f} = \text{model}(\vec{x})$
- 6: **end while**
- 7: **end while**
- 8: Use the rest of the points ($\vec{x}_{\text{rest}}, \vec{f}_{\text{rest}}$) in the high-fidelity database as the test set to calculate \mathcal{M} .

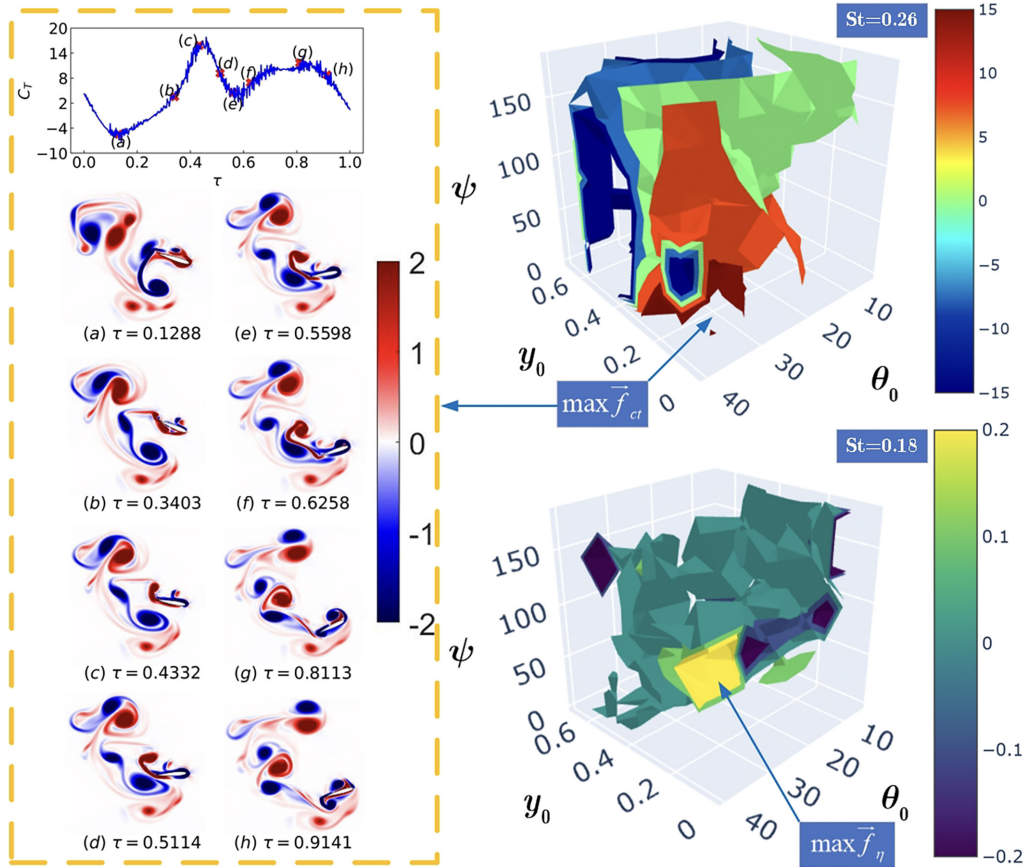


FIG. 4. The isosurfaces under St corresponding to the $\max \bar{f}_{ct}$ and $\max \bar{f}_\eta$. Inside the yellow-contour, we present the thrust curve and the instantaneous flow field diagram under the $\max \bar{f}_{ct}$ case. Subplots (a)–(h) represent the instantaneous vorticity fields corresponding to the specific instants (a)–(h) on the thrust curve, as denoted by the normalized time parameter $\tau = \omega t / 2\pi$.

convergence of f_{ct} and f_η towards the maximum isosurface is observed.

To provide a comprehensive understanding of the thrust curve and corresponding characteristic instantaneous flow fields in Fig. 4, we use the normalized time parameter $\tau = \omega t / 2\pi$. Specifically, the points marked (a)–(h) on the thrust curve correspond to the eight characteristic instantaneous vorticity diagrams shown in Figs. 4(a)–4(h). During the flapping period, the trailing edge of the foil gradually moves upward and shears at the tail, generating a counterclockwise rotating vortex, as shown in Fig. 4(a). The form drag increases with the larger angle of attack (AOA), leading to a negative thrust coefficient. As the foil begins to move downward, the trailing edge vortex (TEV) gradually moves from the lower side to the upper side of the foil under the interaction of the downstroke and the free stream. The trailing edge shear velocity compresses the TEV below the foil, which makes the vorticity of the TEV rotating clockwise above the foil increase gradually. This kind of wake mode can be used to reduce resistance and even transfer to the propulsive mode. With the AOA decreasing, the form drag decreases and the thrust coefficient increases. The thrust generated by the TEV gradually dominates, and the thrust coefficient changes from negative to positive, reaching its peak at $\tau = 0.4332$ as shown in Fig. 4(c).

As the foil continues to flap downward, the TEV below and above the foil move away from the trailing edge, reducing the horizontal velocity component of the jet generated by the TEV. The thrust coefficient decreases significantly, but it is still positive as shown in Figs. 4(c) and 4(d).

As the foil flaps downward, the form drag gradually increases, and the thrust coefficient continues to decrease. The small clockwise-rotating vortex sandwiched between the trailing edge shows positive vorticity. Under the influence of trailing edge velocity attenuation, the strength of the TEV also decreases and the thrust decreases significantly as illustrated in Fig. 4(e). It can be seen that the leading edge vortex (LEV) produces slowly and its intensity increases. The suction generated by the leading edge of the flapping foil can provide part of the propulsion, so the thrust generated by the TEV and the leading edge suction lead to the gradual increase of the thrust coefficient, as shown in Fig. 4(f).

As the TEV moves upward after shedding, the intensity of the LEV significantly increases and its coverage reaches nearly half the chord length. Then a secondary vortex opposite to the direction of the LEV is generated at the leading edge. With development of the LEV strength, the leading edge suction remains constant [39,40]. The thrust generated by the TEV maintains the stability in Fig. 4(g). However, as the trailing edge moves upward, the TEV that rotates coun-

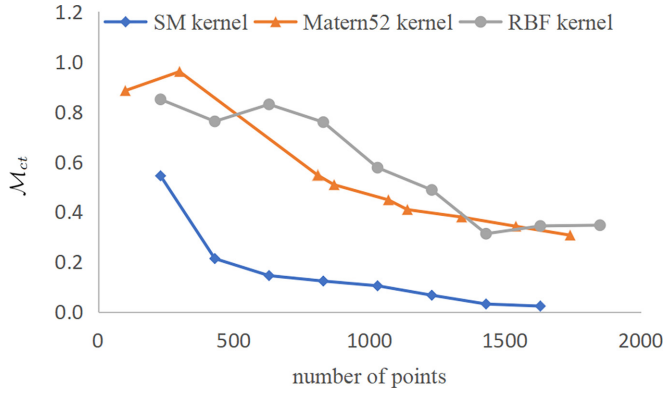


FIG. 5. The M_{ct} of different kernels during iterations.

terclockwise is produced on the lower surface of the trailing edge. The decrease of the AOA leads to the reduction of leading edge suction and cannot produce obvious thrust. At the same time, the boundary layer stops transporting energy to develop the vorticity of the secondary vortex, which means that the vorticity of “ λ ” structure reaches its peak, cutting off the further growth of the LEV and leading to the separation [41]. The thrust coefficient decreases significantly when the TEV and the small vortex that rotates clockwise are far away from the foil, as shown in Fig. 4(h).

B. Thrust optimization: Kernel selection for single-fidelity GP

Consider an ergodic search with $St \in [0.1, 0.3]$, $y_0 \in [0.1, 0.6]$, $\theta_0 \in [5, 40]$, and $\psi \in [0, 180]$. Supposing to set up a database of $6 \times 6 \times 8 \times 7 = 2016$ points by equal-distance sampling, the low-fidelity data are less complex than the high-fidelity data and are much easier to train. Therefore, the low-fidelity output f_{ct} is taken as an implementation. The M_{ct} of different kernels is illustrated in Fig. 5.

The M_{ct} of the SM kernel reaches 0.14 ($f_{ct} \in [-15, 12]$) with 600 points, almost an order of magnitude more accurate than the standard kernels like the Matern52 kernel and the RBF kernel. The SM kernel is more suitable for the flapping foil’s motion response which involves many complex features since it has a more abundant spectrum. Therefore, the SM kernel is adopted as the base kernel of the present GP framework. After training for 600 points, the predictions of the parameter space compared with the raw data are shown in Fig. 6 for $St = 0.1$ and $St = 0.3$, where the y_0 , θ_0 , and ψ labels denote the heave amplitude, the pitch amplitude, and the phase angle, respectively. It can be observed that the SF-GP learns sufficient features with only 30% of the ergodic data set, significantly saving the number of simulations with adequate accuracy.

C. Validation and implementation for MMGP

The efficiency of a multifidelity GP for the flapping foil by the BDIM has been validated in Ref. [9]. Meanwhile, the

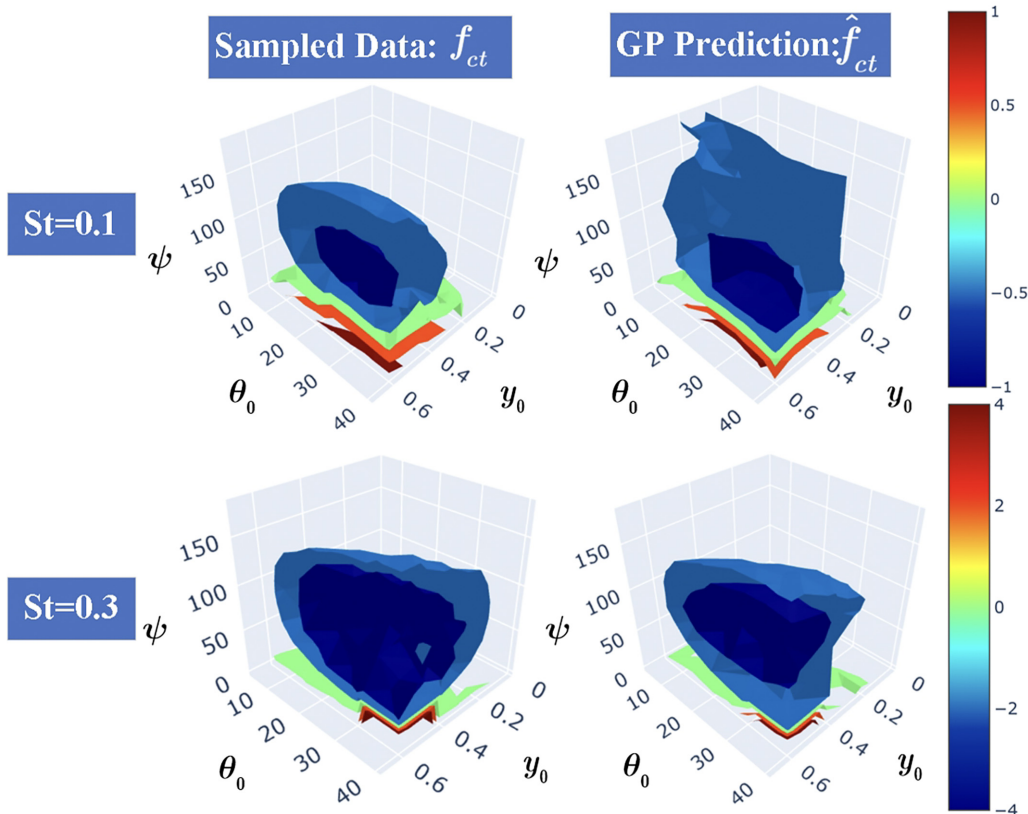


FIG. 6. The three-dimensional view of low-fidelity C_T comparison between the uniformly sampled result (left) and the GP model with 600 points (right).

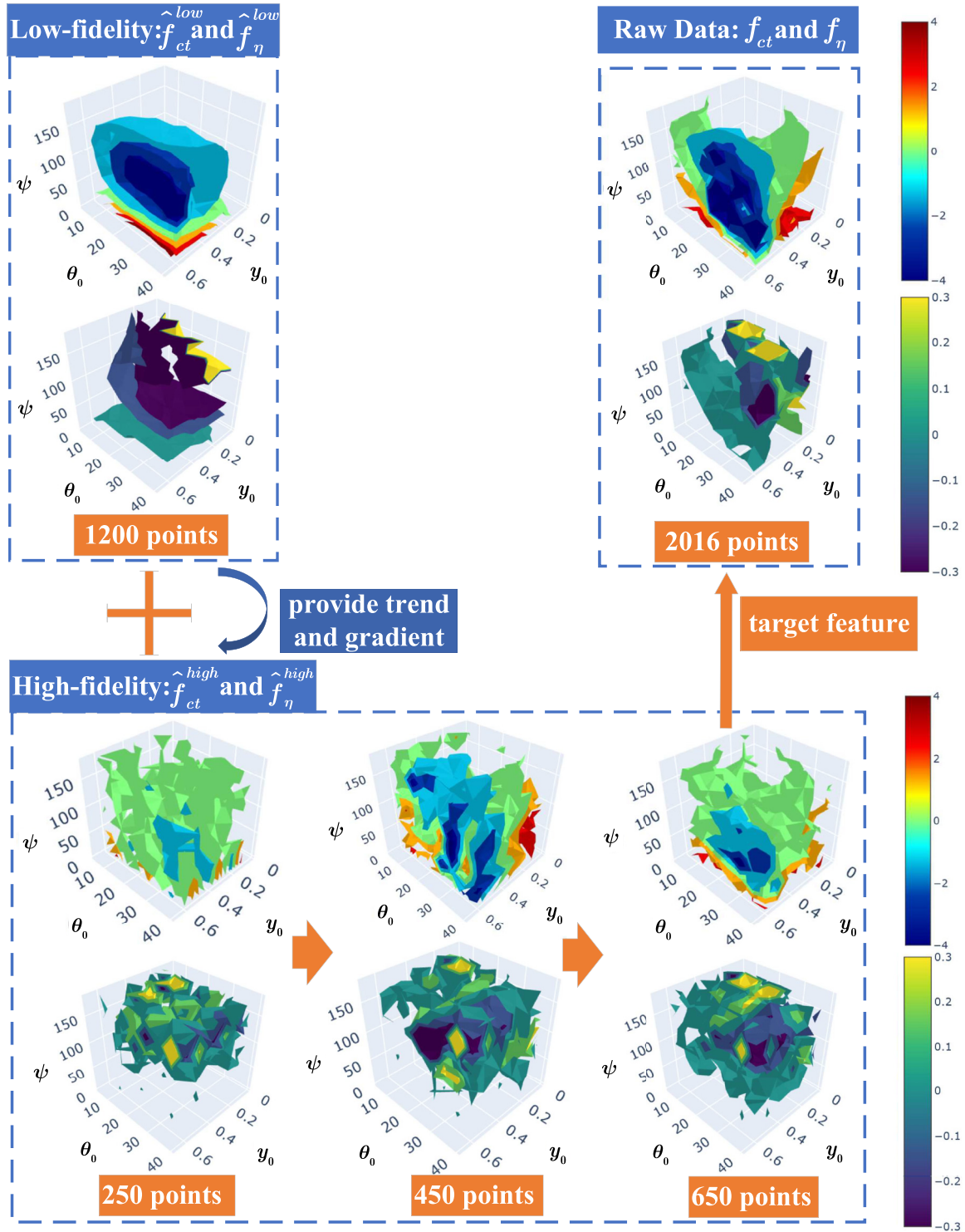


FIG. 7. The training process of the MMGP model. We fuse two fidelities of the multitask GP model in the MMGP model. The low-fidelity model provides trend and gradient information with 1200 points. The high-fidelity model initially resembles the low-fidelity model and gradually converges to the raw high-fidelity database during the training process. The color bars correspond to the values represented by the isosurfaces in the respective plots, with different color maps being employed for f_{ct} and f_{η} for clearer distinction.

advantage, the robustness, and the capability of the proposed multitask GP with multi-AF have been verified in the Appendix. In this subsection, the advantage of prediction and multiobjective optimization performance of the MMGP compared with the multifidelity GP will be shown.

Figure 7 shows the isosurfaces of the prediction with the MMGP model for $St = 0.1$, using the same isosurface configuration as in Fig. 4. The labels y_0 , θ_0 , and ψ indicate the heave amplitude, the pitch amplitude, and the phase difference, respectively. The variance of the high-fidelity data is larger than

TABLE II. The number of high-fidelity points that the MMGP and the multifidelity GP needed to converge.

	Objective	Objective	Total points
	f_{ct}	f_η	
MMGP	650	650	650
Multifidelity GP	610	270	821 (duplicates removed)

that of the low-fidelity data, which means the former has more features. Meanwhile, the MMGP needs to predict both f_{ct} and f_η , which is more complex than the previous task in Sec. II C. In the training process of 650 high-fidelity points, the MMGP learns most of the features of both f_{ct} and f_η . The relative prediction errors according to formula (10) are $m = 5.174\%$ and $n = 5.729\%$, illustrating that the multi-AF balances the convergence of two outputs well. The relative prediction errors are further defined as the training stop conditions, or rather the conditions of convergence.

To demonstrate the prediction efficiency of the MMGP, the multifidelity framework is reproduced to make a comparison between the MMGP and the multifidelity based on the stopping condition. The MMGP necessitates the training of just one model to achieve convergence. The multifidelity approach separately and independently trains two models, one for the objective f_{ct} and another for f_η . Once both models converge, their data sets, obtained through the active learning, are deduplicated and merged. The combined data set is then compared with the dataset required by the MMGP to achieve convergence for both f_{ct} and f_η . As seen in Table II, although the MMGP might not converge as efficiently as each individual task in the multifidelity approach, the MMGP utilizes around 20% fewer high-fidelity points to attain the same accuracy when handling both tasks simultaneously. In essence, the MMGP proves to be more adept than the multifidelity GP for such multiobjective optimization problems.

To show the optimization performance of the MMGP, a multiobjective optimization problem is set up as follows:

$$\begin{aligned}
 \max \quad & P = [f_{ct}, f_\eta], \\
 \text{s.t.} \quad & \text{St} \in [0.1, 0.3], \\
 & y_0 \in [0.1, 0.6], \\
 & \theta_0 \in [5, 40], \\
 & \psi \in [0, 180].
 \end{aligned} \tag{15}$$

The linear weighting model is widely used for the multiobjective optimization. The simple additive weighting (SAW) [42], a classical class of linear weighted summing methods, is applied in this study. The SAW includes two steps, adding zoom factors β and adding weights w to different objectives. The target function can be expressed as

$$P = w_{ct}\beta_{ct}f_{ct} + w_\eta\beta_\eta f_\eta, \tag{16}$$

where zoom factors β_η and β_{ct} satisfy

$$\frac{\beta_\eta}{\beta_{ct}} = \frac{\max f_\eta - \min f_\eta}{\max f_{ct} - \min f_{ct}}. \tag{17}$$

In the present study, we set $\beta_{ct} = 1$ and $\beta_\eta = 40$ as f_{ct} is nearly 1 order larger than f_η . Three tasks are designed and

TABLE III. The parameter configurations of task P_1 and P_2 .

	w_{ct}	β_{ct}	w_η	β_η
P_1	0.5	1	0.5	40
P_2	0.8	1	0.2	40

listed in Table III. Additionally, the optimal efficiency near the rated thrust is of interest in engineering practice. Therefore, another task, $P_3 = -\beta_{ct}|f_{ct} - 10| + \beta_\eta f_\eta$, is further considered.

The prediction of loss $L_i = \max \hat{P}_i - \max P_i$ is shown in Fig. 8. It can be observed that different tasks converge at different points (P_2 converges at 250, while P_1 and P_3 converge at 950), while the trend is similar. Though predictions get better or worse during the training process, the overall trend is gradually close to the optimum. Taking P_2 , for example, the flow field of the predicted optimal point \hat{x} is constantly optimized in the training process.

We use the nondominated sorting genetic algorithm 2 (NSGA-II) to find the Pareto-optimal solutions for the multiobjective optimization problem. These solutions are not dominated by any other feasible solutions in terms of the two objectives. Figure 9 shows the Pareto front of the optimal solutions obtained by NSGA-II. Three optimal points are selected to illustrate the corresponding values of tasks P_1 to P_3 . P_1 represents a balanced motion model. P_2 represents the high-thrust motion model. P_3 represents the high-efficient motion model around the foil's rated thrust. As the parameter ψ decreases gradually, the foil's motion mode transitions from

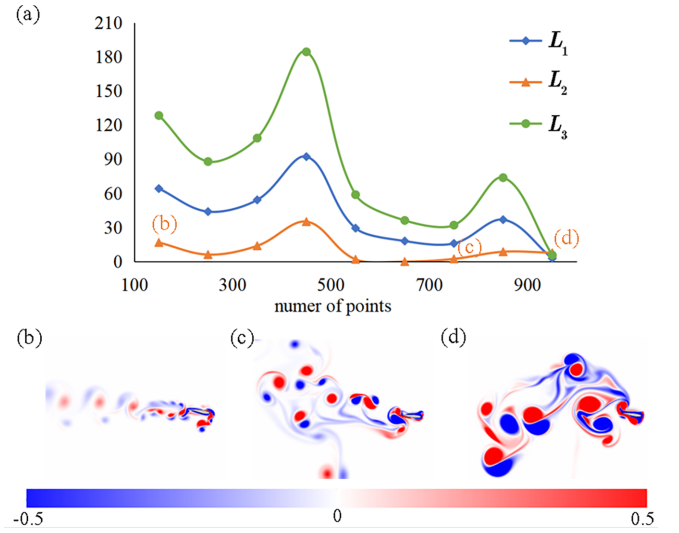


FIG. 8. (a) The prediction loss of multiobjective function L_i during the training process. (b)–(d) The instantaneous vorticity diagram's optimization shown in the early, middle, and late stages of training are also marked on the curve. (b) For $[\text{St}, y_0, \theta_0, \psi] = [0.26, 0.2, 15, 150]$, $[C_T, \eta, P_2] = [-1.06, 0.1, -0.38]$. (c) For $[\text{St}, y_0, \theta_0, \psi] = [0.26, 0.1, 15, 30]$, $[C_T, \eta, P_2] = [6.17, 0.049, 5.3]$. (d) For $[\text{St}, y_0, \theta_0, \psi] = [0.3, 0.4, 20, 0]$, $[C_T, \eta, P_2] = [21.3, 0.052, 17.49]$. The color bar below represents the vorticity magnitude in panels (b)–(d).

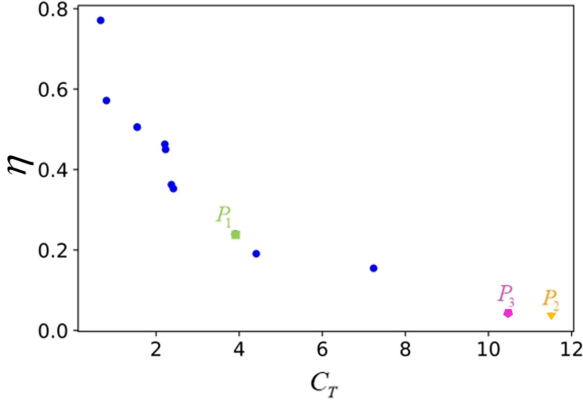


FIG. 9. The Pareto front in the objective space. The square, triangle, and pentagon points indicate the selected optimal points corresponding to tasks P_1 , P_2 and P_3 , where P_1 is for $[\text{St}, y_0, \theta_0, \psi] = [0.22, 0.5, 15, 90]$, P_2 is for $[\text{St}, y_0, \theta_0, \psi] = [0.22, 0.3, 35, 0]$, and P_3 is for $[\text{St}, y_0, \theta_0, \psi] = [0.26, 0.4, 20, 60]$.

high efficiency to high thrust, along with other subtle changes in the parameters. This method can facilitate the design of different motion modes for flapping-wing robots in the future.

IV. CONCLUSION

A multitask and multifidelity Gaussian process (MMGP) model, based on Bayesian optimization, is proposed to accurately predict and optimize the multiobjective performance of flapping foils using minimal high-fidelity data. By integrat-

ing the multifidelity submodel and the multitask submodel, the MMGP effectively leverages low-fidelity data and efficiently addresses multiobjective problems. The Bayesian optimization with the multi-AF algorithm is employed to select candidate samples, representing multiple outputs, which are automatically simulated to update the MMGP model.

To determine the most suitable prior GP model for describing the hydrodynamic responses of the flapping foil, an ergodic database is constructed and analyzed. Initially, the basic single-fidelity GP model is investigated, revealing that the SM kernel exhibits faster convergence than the RBF kernel and the Matern kernel for the present flapping foil case. Moreover, the single-fidelity GP with the SM kernel can capture sufficient features using only 30% of the ergodic database. For the multitask GP based on Bayesian optimization, it is observed that the proposed multiacquisition function is compatible with classic basic acquisition functions such as EI, PI, and UCB. The optimizing and predicting performance of the MMGP model is evaluated using optimal value prediction and mean absolute error. The results demonstrate that the multi-AF, employing a pair of factors aimed at improving less convergent models, exhibits better robustness in optimization ability when compared with fixed factors. Finally, the prediction task indicates that the proposed MMGP requires fewer total points to converge compared to the multifidelity GP, suggesting its high efficiency as an optimization framework for flapping foil motion. Additionally, the MMGP demonstrates its capability as a paradigm for addressing multiobjective problems in the context of flapping foils.

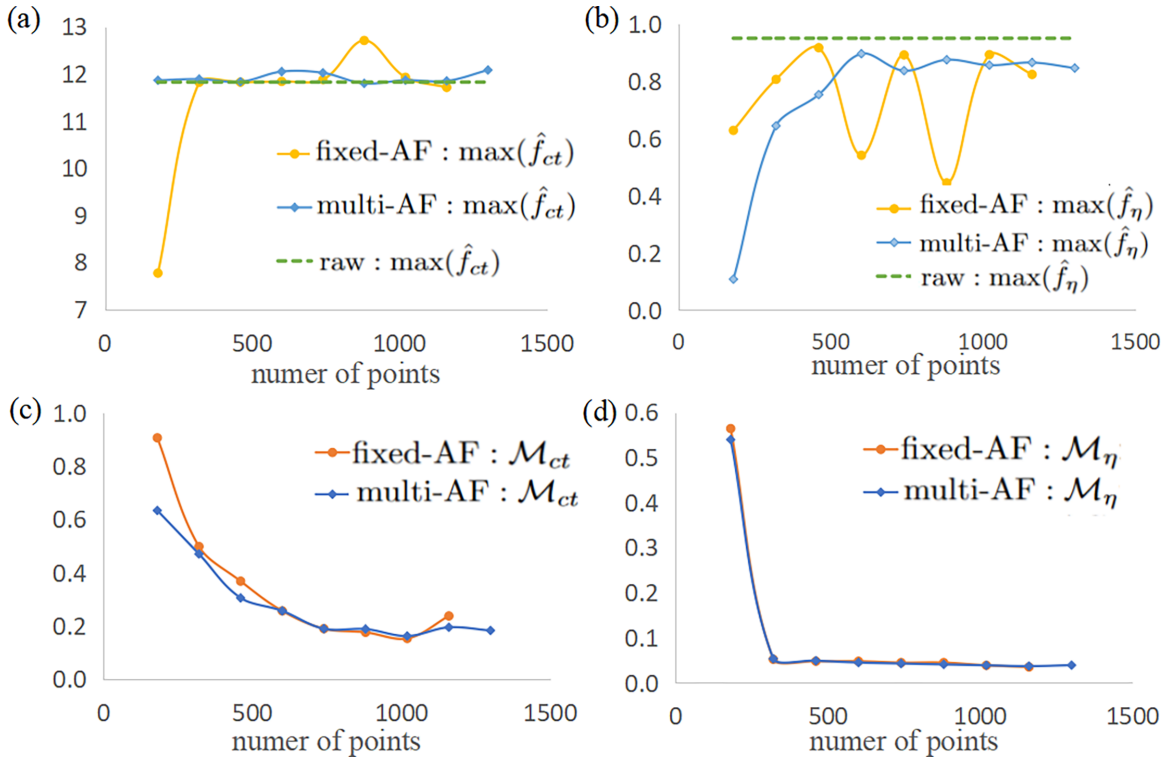


FIG. 10. Compare the two models' global optimization ability and convergence by predicting the two tasks. (a) The prediction of maximum \hat{f}_{ct} during the training process. (b) The prediction of maximum \hat{f}_η during the training process. (c) The \mathcal{M}_{ct} of \hat{f}_{ct} during the training process. (d) The \mathcal{M}_η of \hat{f}_η during the training process.

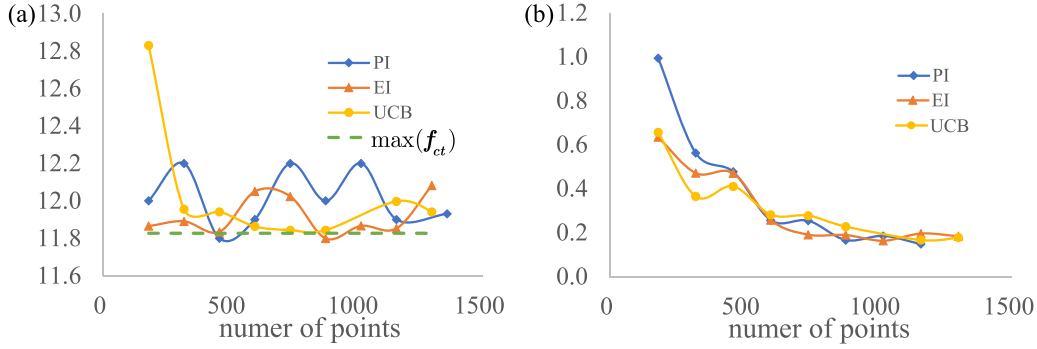


FIG. 11. Compare the performance of multi-AF with different basic acquisition functions. (a) The prediction of maximum \hat{f}_{ct} during the training process. (b) The \mathcal{M}_{ct} of \hat{f}_{ct} during the training process.

As demonstrated in previous sections, while the simplified problem is far from practical engineering applications, it is believed that the MMGP holds the potential for handling high-dimensional and complex problems. Exploring neural network approaches to represent the nonlinear component of different fidelities would be of significant interest in future studies. Furthermore, although the optimization efficiency of the MMGP may not match that of heuristic algorithms, the MMGP is capable of inferring the mean and the variance. Hence, merging the MMGP and the GA holds promise for further exploration and investigation.

ACKNOWLEDGMENTS

The authors would like to acknowledge the funding by the National Key Research and Development Program (Grant No. 2022YFC2805200), the Key Research and Development Program of Zhejiang Province (Grant No. 2023C03133), the Innovative Research Foundation of Ship General Performance (Grant No. 31422225), and the research initiation grant provided by Westlake University (Grant No. 103110556022101).

APPENDIX: VALIDATION FOR MULTITASK GP WITH MULTIAQUISITION FUNCTION

To set up the MMGP, a low-fidelity multitask GP model needs to be trained. Therefore, the multi-AF is proposed to construct a high-efficiency multitask GP. In this Appendix, we

demonstrate the advantage, the robustness, and the capability of the multi-AF for the multitask GP.

To compare with the multi-AF, we use a fixed factor of $\alpha = 0.5$ in formula (10) as the control group fixed AF. Then we measure the global optimization ability and convergence of the two multitask models. In Figs. 10(a) and 10(b), the green line denotes the raw maximum point in the ergodic database. The $\max \hat{f}_{\eta}$ and $\max \hat{f}_{ct}$ of multi-AF (acquisition function) have a variance of 0.01 and 0.06, while the control group has 2.29 and 0.06. Combined with the analysis in Figs. 10(a) and 10(b), the prediction is more stable when the multi-AF converges near the optimal raw value. Thus, it has better robustness in optimization ability. In Figs. 10(c) and 10(d), the \mathcal{M}_{ct} and \mathcal{M}_{η} of multi-AF decreased with the training process and the declining trend gradually slowed down, just slightly lower than the control group.

To verify the compatibility of the multi-AF, we replace $\delta_{ct}(\vec{x})$ and $\delta_{\eta}(\vec{x})$ with a common single acquisition function such as PI, EI, and UCB. The result is shown in Fig. 11. The result shows that the multi-AF is a compatible combined policy for the classic basic acquisition functions. The variance of optimization values PI, EI, and UCB in Fig. 11(a) is 0.02, 0.01, and 0.11, respectively. The EI is chosen as the basic acquisition function in the implementations since it has better robustness in optimization. Additionally, Fig. 11(a) shows that UCB has weak robustness early in the training process but has the best robustness in the middle and late stages of training. The converging speed of the three acquisition functions is nearly the same as shown in Fig. 11(b).

- [1] B. Sun, W. Li, Z. Wang, Y. Zhu, Q. He, X. Guan, G. Dai, D. Yuan, A. Li, W. Cui, and D. Fan, Recent progress in modeling and control of bio-inspired fish robots, *J. Marine Sci. Eng.* **10**, 773 (2022).
- [2] J. Deng, L. Teng, D. Pan, and X. Shao, Inertial effects of the semi-passive flapping foil on its energy extraction efficiency, *Phys. Fluids* **27**, 053103 (2015).
- [3] Y. Pan, X. Dong, Q. Zhu, and D. K. P. Yue, Boundary-element method for the prediction of performance of flapping foils with leading-edge separation, *J. Fluid Mech.* **698**, 446 (2012).
- [4] F. M. Bos, D. Lentink, B. W. Van Oudheusden, and H. Bijl, Influence of wing kinematics on aerodynamic performance in hovering insect flight, *J. Fluid Mech.* **594**, 341 (2008).
- [5] J. Deng, L. Sun, and X. Shao, Dynamical features of the wake behind a pitching foil, *Phys. Rev. E* **92**, 063013 (2015).
- [6] L. Sun, J. Deng, and X. Shao, Three-dimensional instabilities for the flow around a heaving foil, *Phys. Rev. E* **97**, 013110 (2018).
- [7] A. Coppedè, S. Gaggero, G. Vernengo, and D. Villa, Hydrodynamic shape optimization by high fidelity CFD solver and Gaussian process based response surface method, *Appl. Ocean Res.* **90**, 101841 (2019).
- [8] J. S. Izraelevitz and M. S. Triantafyllou, Adding in-line motion and model-based optimization offers exceptional force control authority in flapping foils, *J. Fluid Mech.* **742**, 5 (2014).

- [9] H. Zheng, F. Xie, T. Ji, Z. Zhu, and Y. Zheng, Multifidelity kinematic parameter optimization of a flapping airfoil, *Phys. Rev. E* **101**, 013107 (2020).
- [10] M. A. Esfahani, H. R. Karbasian, and K. C. Kim, Multi-objective optimization of the kinematic parameters of fish-like swimming using a genetic algorithm method, *J. Hydrodyn.* **31**, 333 (2019).
- [11] F. Fish and G. V. Lauder, Passive and active flow control by swimming fishes and mammals, *Annu. Rev. Fluid Mech.* **38**, 193 (2006).
- [12] A. J. Smits, Undulatory and oscillatory swimming, *J. Fluid Mech.* **874**, P1 (2019).
- [13] M. Triantafyllou, A. Techet, and F. Hover, Review of experimental work in biomimetic foils, *IEEE J. Oceanic Eng.* **29**, 585 (2004).
- [14] X. Wu, X. Zhang, X. Tian, X. Li, and W. Lu, A review on fluid dynamics of flapping foils, *Ocean Eng.* **195**, 106712 (2020).
- [15] S. Verma, G. Novati, F. Noca, and P. Koumoutsakos, Fast motion of heaving airfoils, *Procedia Comput. Sci.* **108**, 235 (2017).
- [16] X. Meng, H. Babae, and G. E. Karniadakis, Multi-fidelity Bayesian neural networks: Algorithms and applications, *J. Comput. Phys.* **438**, 110361 (2021).
- [17] Z. Xiong and S. Chen, A multi-fidelity approach for reliability-based risk assessment of single-vehicle crashes, *Acc. Ana. Prevention* **195**, 107391 (2024).
- [18] A. Hebbal, L. Brevault, M. Balesdent, E. Talbi, and N. Melab, Multi-fidelity modeling with different input domain definitions using deep Gaussian processes, *Struct. Multidiscip. Optim.* **63**, 2267 (2021).
- [19] L. Parussini, D. Venturi, P. Perdikaris, and G. E. Karniadakis, Multi-fidelity Gaussian process regression for prediction of random fields, *J. Comput. Phys.* **336**, 36 (2017).
- [20] P. Perdikaris, D. Venturi, J. O. Royset, and G. E. Karniadakis, Multi-fidelity modelling via recursive co-kriging and Gaussian-Markov random fields, *Proc. R. Soc. A* **471**, 20150018 (2015).
- [21] M. Raissi, P. Perdikaris, and G. E. Karniadakis, Inferring solutions of differential equations using noisy multi-fidelity data, *J. Comput. Phys.* **335**, 736 (2017).
- [22] H. Babae, P. Perdikaris, C. Chrysostomidis, and G. E. Karniadakis, Multi-fidelity modelling of mixed convection based on experimental correlations and numerical simulations, *J. Fluid Mech.* **809**, 895 (2016).
- [23] J. A. Esfahani, E. Barati, and H. R. Karbasian, Effect of caudal on hydrodynamic performance of flapping foil in fish-like swimming, *Appl. Ocean Res.* **42**, 32 (2013).
- [24] L. Bonfiglio, P. Perdikaris, S. Brizzolara, and G. E. Karniadakis, Multi-fidelity optimization of super-cavitating hydrofoils, *Comput. Methods Appl. Mech. Eng.* **332**, 63 (2018).
- [25] L. Bonfiglio, P. Perdikaris, G. Vernengo, J. S. de Medeiros, and G. Karniadakis, Improving swath seakeeping performance using multi-fidelity Gaussian process and Bayesian optimization, *J. Ship Res.* **62**, 223 (2018).
- [26] T. Ji, F. Jin, F. Xie, H. Zheng, X. Zhang, and Y. Zheng, Active learning of tandem flapping wings at optimizing propulsion performance, *Phys. Fluids* **34**, 047117 (2022).
- [27] J. Snoek, H. Larochelle, and R. P. Adams, Practical Bayesian optimization of machine learning algorithms, in *26th Annual Conference on Neural Information Processing Systems 2012*, Lake Tahoe, Nevada, USA (Curran Associates, Inc., 2012), Vol. 25.
- [28] D. R. Jones, A taxonomy of global optimization methods based on response surfaces, *J. Global Optim.* **21**, 345 (2001).
- [29] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, Gpytorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration, in *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montreal, Canada, Vol. 31 (MIT, Cambridge, MA, 2018).
- [30] A. P. Maertens and G. D. Weymouth, Accurate Cartesian-grid simulations of near-body flows at intermediate Reynolds numbers, *Comput. Methods Appl. Mech. Eng.* **283**, 106 (2015).
- [31] S. C. Schlanderer, G. D. Weymouth, and R. D. Sandberg, The boundary data immersion method for compressible flows with application to aeroacoustics, *J. Comput. Phys.* **333**, 440 (2017).
- [32] G. D. Weymouth and D. K. Yue, Boundary data immersion method for Cartesian-grid simulations of fluid-body interaction problems, *J. Comput. Phys.* **230**, 6233 (2011).
- [33] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT, Cambridge, MA, 2005).
- [34] E. V. Bonilla, K. Chai, and C. Williams, Multi-task Gaussian process prediction, in *Proceedings of the 21st Annual Conference on Neural Information Processing Systems 2007*, Vancouver, British Columbia, Canada, Vol. 20 (MIT, Cambridge, MA, 2007), pp. 153–160.
- [35] A. G. Wilson and R. P. Adams, Gaussian process kernels for pattern discovery and extrapolation, in *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA, Vol. 28 (PMLR, 2014), pp. 1067–1075.
- [36] A. G. Wilson, Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes, Ph.D. thesis, University of Cambridge, 2014.
- [37] A. G. Wilson, E. Gilboa, A. Nehorai, and J. P. Cunningham, Fast kernel learning for multidimensional pattern extrapolation, in *Proceedings of the 28th Annual Conference on Neural Information Processing Systems 2014*, Montreal, Canada, Vol. 27 (MIT, Cambridge, MA, 2014).
- [38] D. Zhan and H. Xing, Expected improvement for expensive optimization: a review, *J. Global Optim.* **78**, 507 (2020).
- [39] J. Katz, A discrete vortex method for the non-steady separated flow over an airfoil, *J. Fluid Mech.* **102**, 315 (1981).
- [40] K. Ramesh, A. Gopalarathnam, K. Granlund, M. Ol, and J. Edwards, Discrete-vortex method with novel shedding criterion for unsteady aerofoil flows with intermittent leading-edge vortex shedding, *J. Fluid Mech.* **751**, 500 (2014).
- [41] J. D. Eldredge and A. R. Jones, Leading-edge vortices: Mechanics and modeling, *Annu. Rev.* **51**, 75 (2019).
- [42] S. Zions and J. Wallenius, An interactive multiple objective linear programming method for a class of underlying nonlinear utility functions, *Manage. Sci.* **29**, 519 (1983).