



Adaptive N-Fidelity Metamodels for Noisy CFD Data

Jeroen Wackers, Michel Visonneau, Simone Ficini, Riccardo Pellegrini, Andrea Serani, Matteo Diez

► To cite this version:

Jeroen Wackers, Michel Visonneau, Simone Ficini, Riccardo Pellegrini, Andrea Serani, et al.. Adaptive N-Fidelity Metamodels for Noisy CFD Data. AIAA Aviation Forum 2020, Jun 2020, online, United States. ⟨hal-02613470⟩

HAL Id: hal-02613470

<https://hal.science/hal-02613470v1>

Submitted on 20 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Adaptive N-Fidelity Metamodels for Noisy CFD Data

Jeroen Wackers * and Michel Visonneau †
LHEEA Lab, Ecole Centrale de Nantes, CNRS UMR 6598, Nantes, France

Simone Ficini ‡
Roma Tre University, Department of Engineering, Rome, Italy

Riccardo Pellegrini §, Andrea Serani ¶, and Matteo Diez ||
CNR-INM, National Research Council-Institute of Marine Engineering, Rome, Italy

An adaptive N -fidelity approach to metamodeling from noisy data is presented for design-space exploration and design optimization. Computational fluid dynamics (CFD) simulations with different numerical accuracy (spatial discretization) provides metamodel training sets affected by unavoidable numerical noise. The N -fidelity approximation is built by an additive correction of a low-fidelity metamodel with metamodels of differences (errors) between higher-fidelity levels whose hierarchy needs to be provided. The approach encompasses two core metamodeling techniques, namely: i) stochastic radial-basis functions (SRBF) and ii) Gaussian process (GP). The adaptivity stems from the sequential training procedure and the auto-tuning capabilities of the metamodels. The method is demonstrated for an analytical test problem and a CFD-based optimization of a NACA airfoil, where the fidelity levels are defined by an adaptive grid refinement technique of a Reynolds-averaged Navier-Stokes (RANS) solver. The paper discusses: i) the effect of using more than two fidelity levels; ii) the use of least squares regression as opposed to exact interpolation; iii) the comparison between SRBF and GP; and iv) the use of two sampling approaches for GP. Results show that in presence of noise, the use of more than two fidelity levels improves the model accuracy with a significant reduction of the number of high-fidelity evaluations. Both least squares SRBF and GP provide promising results in dealing with noisy data.

I. Introduction

THE design of complex engineering systems (such as the aerodynamic/hydrodynamic and structural design of aerial and water-borne vehicles) benefits from simulation-based design optimization (SBDO) formulations and solutions. SBDO supports decision-making process and helps designers to make sound decisions, especially for innovative configurations. It is usually implemented as an iterative process combining design modification methods, numerical simulations, and optimization algorithms to identify new optimized designs. In order to achieve accurate final solutions, high-fidelity physics-based solvers (as implemented for computational fluid dynamics, structural analysis, etc.) are needed, resulting in computationally expensive analyses. Furthermore, the integration of these solvers with optimization algorithms (which may require a large number of function evaluations to converge to the final solution) makes the computational cost very high and the SBDO procedure a technological challenge.

To reduce the computational cost in SBDO, supervised machine learning/metamodeling methods have been developed and successfully applied in several engineering fields [1]. Metamodels provides an approximated model of expensive simulations based on a few of them, then the optimization is performed over the metamodel. The performance of the metamodels in providing accurate approximation of the design space, is problem-dependent and determined by several concurrent issues, such as the presence of nonlinearities, the problem dimensionality, and the approach used for its training [2]. In the last decade, research has moved to function-adaptive approaches, also known as dynamic/adaptive metamodels [3], which are able to improve their fitting capability by adaptive sampling. In this context, the design of

*Research Scientist, ECN, 1 rue de la Noe, 44321 Nantes, Cedex 3

†Research Director, ECN, 1 rue de la Noe, 44321 Nantes, Cedex 3

‡Ph.D. student, Roma Tre University, Via Vito Volterra 62, 00146 Rome

§Postdoctoral Research Fellow, INM Rome, Via di Vallerano 139, 00128, Email: riccardo.pellegrini@inm.cnr.it

¶Research Scientist, INM Rome, Via di Vallerano 139, 00128

|| Senior Research Scientist, INM Rome, Via di Vallerano 139, 00128, AIAA Member

experiments used for metamodel training is not defined a priori but dynamically updated, exploiting the information that becomes available during the analysis process. Thus, training points are added where they are most useful, reducing the number of function evaluations required to represent/optimize the desired function [4]. Among others, an adaptive sampling method based on the prediction uncertainty has been presented in [3], the expected improvement adaptive sampling has been introduced in [5], a multi-criteria adaptive sampling for parallel-infill has been presented in [6], and the lower confidence bounding method has been presented in [7]. Unfortunately, the adaptive sampling/learning process is affected by the computational-output noise. Adaptive sampling methods may react to noise by adding many training points in noisy regions, rather than selecting new points in unseen regions. This deteriorates the model quality/efficiency [8].

In addition to dynamic/adaptive metamodels and with the aim of reducing further the computational cost of SBDO, multi-fidelity (MF) approximation methods have been developed, with the aim of combining the accuracy of high-fidelity solvers with the computational cost of low-fidelity solvers [9]. MF metamodels use mainly low-fidelity simulations and only few high-fidelity simulations are used to improve the model accuracy [10]. Additive and/or multiplicative correction methods, also known as “bridge functions” [11, 12], can be used to build MF metamodels. Several metamodels have been used in the literature with MF data, such as non-intrusive polynomial chaos [13], Gaussian process [14], co-kriging [15] and stochastic radial basis functions [16]. In CFD-based analysis/optimization, different fidelity levels may be obtained by varying the physical model (*e.g.* potential flow or RANS solvers), the numerical accuracy (*e.g.* grid discretization and convergence tolerances), and data coverage [17]. Recently, [18] proposed a MF metamodel with an arbitrary number of fidelity levels (N -fidelity), based on SRBF. Adaptive sampling methods have been recently combined with MF metamodels: an augmented expected improvement for multi-fidelity kriging has been presented in [19], a multi-fidelity version of the lower confidence bounding method has been presented by [20], and a multi-fidelity adaptive sampling method based on cross validation error and Voronoi partition has been presented in [21]. The use of MF models with noisy data and the assessment of the effect of noise associated with different fidelity levels is still little discussed and requires rigorous formulations and implementations.

In the present work, N -fidelity (NF, with $N = 1, 2$ and 3) metamodeling is used in combination with two metamodels, namely stochastic RBF (SRBF) and Gaussian process (GP). SRBF are used with exact interpolation (I-SRBF) and least squares regression (LS-SRBF), and GP with two sampling methods. The NF methods with SRBF are developed based on authors’ previous work [23, 24] that considered N -fidelity and noise reduction separately. The SBDO is here performed with a Reynolds-averaged Navier-Stokes (RANS) solver for a shape optimization problem. An adaptive grid refinement method is adopted, which naturally produces results spanning a range of fidelity levels. The simulation output is affected by numerical noise.

The objective of the present work is to address four topics of interest for multi-fidelity metamodels, applied to shape optimization problems affected by numerical noise, specifically:

- 1) the effect of using three fidelities instead of one or two;
- 2) the effect of using SRBF with least squares regression (LS-SRBF) as opposed to exact interpolation (I-SRBF);
- 3) the effect of using LS-SRBF versus GP;
- 4) the effect of including an explicit treatment of the training set noise in the adaptive sampling method for GP.

The first topic is investigated comparing I-SRBF with a number of fidelities $N = 1, 2$ and 3 ; the second comparing I-SRBF with LS-SRBF; the third comparing LS-SRBF with GP; and the fourth comparing two approaches for the adaptive sampling with GP. The NF method for noisy data is first demonstrated for an analytical problem. Then, it is applied to the shape optimization of a NACA four-digit airfoil operating at constant lift addressing the reduction of the drag coefficient. CFD simulations are performed with the RANS code ISIS-CFD [25], developed at ECN/CNRS and integrated in the FINE/Marine suite (NUMECA Int). The CFD-solver fidelity is varied by using three computational grid levels, defined by an adaptive grid refinement technique [26]. The capability of the methods are discussed of identifying the global/reference minimum position and value, approximating the high-fidelity function, and identifying the numerical noise in the data.

II. N -Fidelity Approach

Consider $\mathbf{x} \in \mathbb{R}^D$ as the design variables vector of dimension D . Let the true function $f(\mathbf{x})$ be assessed by N fidelities: the highest-fidelity level is $f_1(\mathbf{x})$, the lowest-fidelity is $f_N(\mathbf{x})$, and arbitrary intermediate fidelity levels are $\{f_i\}_{i=2}^{N-1}(\mathbf{x})$. Using $\tilde{\cdot}$ to denote metamodel prediction (given by the metamodels in Section III), the NF approximation $\hat{f}_i(\mathbf{x})$ of $f_i(\mathbf{x})$ is the summation of the lowest-level metamodel and metamodels of the errors (inter-level errors or

bridge-functions) between subsequent levels ($i = 1, \dots, N - 1$)

$$\hat{f}_i(\mathbf{x}) = \tilde{f}_N(\mathbf{x}) + \sum_{k=i}^{N-1} \tilde{\varepsilon}_k(\mathbf{x}). \quad (1)$$

For each i -th fidelity level the training set is $\mathcal{T}_i = \{\mathbf{y}_j, f_i(\mathbf{y}_j)\}_{j=1}^{J_i}$, with J_i the training set size. The resulting inter-level error training set is defined as $\mathcal{E}_i = \{(\mathbf{y}_j, \varepsilon_i(\mathbf{y}_j))\}_{j=1}^{J_i}$, where

$$\varepsilon_i(\mathbf{y}_j) = f_i(\mathbf{y}_j) - \hat{f}_{i+1}(\mathbf{y}_j). \quad (2)$$

This approach is based on the idea that in the presence of noise, a noise-filtered metamodel \hat{f}_{i+1} is a better representation of the low-fidelity response than the data themselves.

For metamodels that provide the uncertainty associated to the prediction, assuming that the uncertainty associated to the prediction of the lowest-fidelity $U_{\tilde{f}_N}$ and the inter-level errors $U_{\tilde{\varepsilon}_k}$ are uncorrelated, the NF prediction uncertainty $U_{\hat{f}_i}$ reads ($i = 1, \dots, N - 1$)

$$U_{\hat{f}_i}(\mathbf{x}) = \sqrt{U_{\tilde{f}_N}^2(\mathbf{x}) + \sum_{k=i}^{N-1} U_{\tilde{\varepsilon}_k}^2(\mathbf{x})}. \quad (3)$$

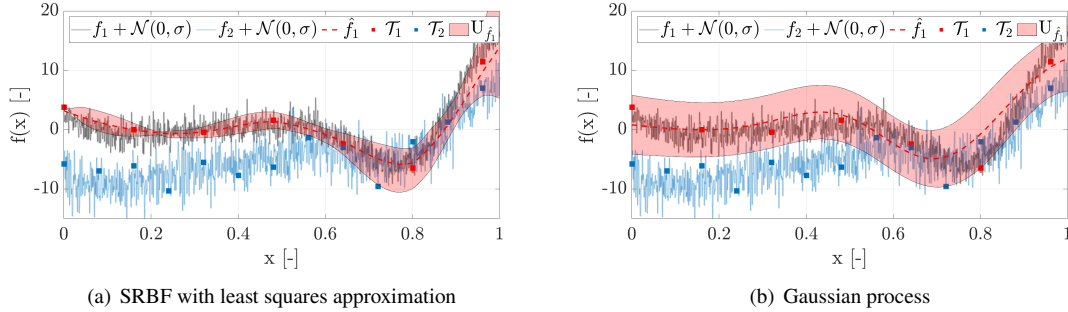


Fig. 1 Example of NF ($N = 2$) metamodels for noisy data.

III. Metamodels for Noisy Data

A. Stochastic Radial-Basis Functions

Given a training set $\mathcal{T} = \{\mathbf{y}_i, f(\mathbf{y}_i)\}_{i=1}^J$, normalizing the variables domain in a unit hypercube, the RBF metamodel prediction $\tilde{f}(\mathbf{x})$ is here computed as the expected value (EV) over a stochastic tuning parameter of the metamodel [3], $\tau \sim \text{unif}[1, 3]$

$$\tilde{f}(\mathbf{x}) = \text{EV}[g(\mathbf{x}, \tau)]_{\tau}, \quad \text{with} \quad g(\mathbf{x}, \tau) = \sum_{j=1}^M w_j \|\mathbf{x} - \mathbf{c}_j\|^{-\tau}, \quad (4)$$

where w_j are unknown coefficients, $\|\cdot\|$ is the Euclidean norm and \mathbf{c}_j are the M RBF centers, whose coordinates are defined via k -means clustering of the training points [27]. Noise reduction is achieved by choosing $M \leq J$. Hence, w_j are determined with least squares regression by solving $\mathbf{w} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{f}$, where $\mathbf{w} = \{w_j\}$ and $a_{ij} = \|\mathbf{y}_i - \mathbf{c}_j\|^{-\tau}$. When $M = J$ then exact interpolation of the training set is imposed, the corresponding metamodel is I-SRBF and the coefficients w_j are computed by solving $\mathbf{A} \mathbf{w} = \mathbf{f}$, with $\mathbf{c}_j = \mathbf{y}_j$. An example of LS-SRBF with two fidelity levels is shown in Fig. 1a.

In this work the optimal number of RBF centers (M) is defined by minimizing a leave-one-out cross-validation (LOOCV) metric [28]. Let $\tilde{h}(\mathbf{x})$ be a metamodel trained by all points but the i -th point, then M is defined as:

$$M = \underset{M_c}{\operatorname{argmin}}(\operatorname{RMSE}), \quad (5)$$

where $M_c \in \mathbb{N}^+$ and the root mean squared error (RMSE) is defined as

$$\operatorname{RMSE} = \sqrt{\frac{1}{J} \sum_{i=1}^J (f(\mathbf{y}_i) - \tilde{h}(\mathbf{y}_i))^2}. \quad (6)$$

The uncertainty $U_{\tilde{f}}(\mathbf{x})$ associated with the SRBF prediction is quantified by the 95%-confidence band of the cumulative density function (CDF) of $g(\mathbf{x}, \tau)$, evaluated using a Monte Carlo sampling over τ [3]

$$U_{\tilde{f}}(\mathbf{x}) = \operatorname{CDF}^{-1}(0.975; \mathbf{x}) - \operatorname{CDF}^{-1}(0.025; \mathbf{x}), \quad (7)$$

with

$$\operatorname{CDF}(\lambda; \mathbf{x}) = \frac{1}{S} \sum_{i=1}^S H[\lambda - f(\mathbf{x}, \tau_i)]. \quad (8)$$

where $H(\cdot)$ is the Heaviside step function and S is the number of Monte Carlo samples. The noise eventually present in the training set is assessed by evaluating the NRMSE of the training set with respect to the corresponding LS-SRBF metamodel as

$$\operatorname{NRMSE} = \frac{1}{R} \sqrt{\frac{1}{J} \sum_{i=1}^J (f(\mathbf{y}_i) - \tilde{f}(\mathbf{y}_i))^2}. \quad (9)$$

where R is the function range.

B. Gaussian process (GP)

Given a training set $\mathcal{T} = \{\mathbf{y}_i, f(\mathbf{y}_i)\}_{i=1}^J$, normalizing the variables domain in a unit hypercube, the GP prediction with zero mean, $\tilde{f}(\mathbf{x})$ and its variance ($\operatorname{Var}[\tilde{f}(\mathbf{x})]$) can be written as [29]

$$\tilde{f}(\mathbf{x}) = \mathbf{k}(\mathbf{x}, \mathbf{y}) \mathbf{K}(\mathbf{y}, \mathbf{y})^{-1} \mathbf{f}(\mathbf{y}), \quad (10)$$

$$\operatorname{Var}[\tilde{f}(\mathbf{x})] = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{k}(\mathbf{x}, \mathbf{y}) \mathbf{K}(\mathbf{y}, \mathbf{y})^{-1} \mathbf{k}(\mathbf{x}, \mathbf{y}), \quad (11)$$

where $\mathbf{K}(\mathbf{x}, \mathbf{x})$ is the covariance matrix with elements $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ and $k(\cdot, \cdot)$ is the covariance function defined as [29]

$$k(\mathbf{x}, \mathbf{y}) = \sigma_F^2 e^{(-\boldsymbol{\gamma}^\top (\mathbf{x} - \mathbf{y}) \circ^2)} + \sigma_n^2 \delta(\mathbf{x}, \mathbf{y}), \quad (12)$$

with " \circ " the Hadamard product, $\boldsymbol{\gamma} \in \mathbb{R}^D$ are the inverse length scale parameters, δ is the Kronecker delta. Finally, σ_F^2 and σ_n^2 are two parameters introduced to explicitly model the noise in the training set [29], namely the signal and noise variance, respectively.

To estimate the variance associated to the noise in the training set (if not known a priori), therefore σ_n^2 and σ_F^2 are evaluated beside γ by maximizing the log marginal likelihood l [29]

$$\{\sigma_n^{*2}, \sigma_F^{*2}, \boldsymbol{\gamma}^*\} = \underset{\sigma_n^2, \sigma_F^2, \boldsymbol{\gamma}}{\operatorname{argmax}} [l], \quad (13)$$

where

$$l = \log p(f(\mathbf{y})|\mathbf{y}) = -\frac{J}{2} \log 2\pi - \frac{1}{2} f(\mathbf{y})^T \mathbf{K}(\mathbf{y}, \mathbf{y})^{-1} f(\mathbf{y}) - \frac{1}{2} \log |\mathbf{K}(\mathbf{y}, \mathbf{y})|. \quad (14)$$

Finally, the prediction uncertainty $U_{\tilde{f}}$ is quantified as

$$U_{\tilde{f}} = 4\sqrt{\operatorname{Var}[\tilde{f}(\mathbf{x})]}. \quad (15)$$

It is worth to mention that the prediction uncertainty as defined in Eq. 15 takes into account also the variance associated to the noise in the training set.

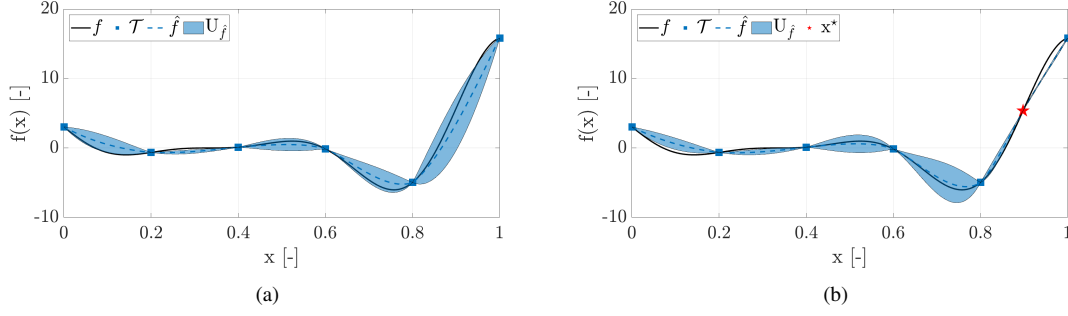


Fig. 2 Example of the adaptive sampling methods using one fidelity without noise: (a) is the initial metamodel with the associated prediction uncertainty and training set and (b) shows the position of the new training point.

IV. Adaptive Sampling Procedure

The training points for the metamodels are defined with an adaptive sampling strategy. To avoid abrupt changes in the number of RBF centers in the problem of Eq. 5, M is constrained to $M_i^{t-1} - 2 \leq M_i^t \leq M_i^{t-1} + 2$, where t is the adaptive sampling iteration. The maximum uncertainty adaptive sampling (MUAS) [4] criterion is then used, where new training points are chosen where the prediction uncertainty is largest. Accordingly, MUAS identifies new training points (see Fig. 2b) by solving the single-objective maximization problem

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}} [U_{\hat{f}}(\mathbf{x})]. \quad (16)$$

The new point \mathbf{x}^* is added to the training sets \mathcal{T}_i with $i = k, \dots, N$, where k is identified as $k = \maxloc(\mathbf{U})$ where

$$\mathbf{U} \equiv \{U_{\tilde{\varepsilon}_1}/\beta_1, \dots, U_{\tilde{\varepsilon}_{N-1}}/\beta_{N-1}, U_{\tilde{f}_N}/\beta_N\}, \quad (17)$$

with β_i the computational cost associated to the i -th level. Using this definition of \mathbf{U} with GP is referred to as GP1 in the following.

Since GP allows to explicitly model the noise in the training set, a second definition of \mathbf{U} is proposed where the noise standard deviation (σ_n^* , computed as per Eq. 13) is subtracted from the total prediction uncertainty

$$\mathbf{U} \equiv \{(U_{\tilde{\varepsilon}_1} - 4\sigma_{n,\tilde{\varepsilon}_1}^*)/\beta_1, \dots, (U_{\tilde{\varepsilon}_{N-1}} - 4\sigma_{n,\tilde{\varepsilon}_{N-1}}^*)/\beta_{N-1}, (U_{\tilde{f}_N} - 4\sigma_{n,\tilde{f}_N}^*)/\beta_N\}. \quad (18)$$

Using this definition of \mathbf{U} with GP is referred to as GP2 in the following.

Finally, Algorithm 1 and 2 show the NF approach with SRBF and GP1/GP2, respectively.

V. Analytical Test Problems

One analytical test problem is selected to assess the proposed approach performance in presence of noise. Here a mono-dimensional and multi-modal example is shown. The functions f_1 and f_2 , without noise, are taken from [31], whereas f_3 is introduced in this work. Figure 3a shows the highest-fidelity level (f_1), whereas Figs. 3b and 3c show, respectively, the same analytical function along with one (f_3) and two ($f_{2,3}$) lower-fidelities and the corresponding errors ($\varepsilon_{1,2}$). The analytical test is defined in Tab. 1, where χ_i is the noise associated to the i -th fidelity level. The noise bandwidth (at 95% confidence) is set to $a_i R_1$ using a normal distribution with zero mean and standard deviation equal to $a_i R_1/4$; $\mathbf{a} = \{a_i\}$ is an arbitrary coefficient set and R_1 is the function range for the 1-st level. The coefficient sets used here are $\mathbf{a} = \{10, 40\}\%$ for $N = 2$ and $\mathbf{a} = \{10, 20, 40\}\%$ for $N = 3$.

Algorithm 1: N -fidelity implementation with SRBF.

```

Define the initial training set for each fidelity level  $\{\mathcal{T}_i\}_{i=1}^N = \{\mathbf{y}_j, f_i(\mathbf{y}_j)\}_{j=1}^{J_i}$ ;
while stopping criteria are not met do // iterate the adaptive sampling process with index  $t$ 
    Set  $\{M_i^t = J_i^t\}_{i=1}^N$ ;
    if Least squares approximation is used then // LS-SRBF auto-tuning based on LOOCV
        for  $i=N, 1$  do
            Find  $M_i^t = \underset{M_c}{\operatorname{argmin}}(\operatorname{RMSE})$  as per Eq. (5) with  $M_i^{t-1} - 2 \leq M_i^t \leq M_i^{t-1} + 2$ ;
        end
    end
    Construct the SRBF metamodels  $\tilde{f}_N$  as per Eq. (4); // Low-fidelity approximation
    Compute the prediction uncertainty  $U_{\tilde{f}_N}$  as per Eq. (7);
    for  $i=N-1, 1$  do // Evaluate bridge functions
        Compute the intra-level error  $\{(\mathbf{y}_j, \varepsilon_i(\mathbf{y}_j))\}_{j=1}^{J_i^t}$  as per Eq. (2);
        Construct the SRBF metamodels of the inter-level errors  $\tilde{\varepsilon}_i$  as per Eq. (4);
        Compute the prediction uncertainty  $U_{\tilde{\varepsilon}_i}$  as per Eq. (7);
    end
    Construct the  $N$ -fidelity approximation  $\hat{f}_1(\mathbf{x})$  as per Eq. (1); //  $N$ -fidelity approximation
    Compute the  $N$ -fidelity prediction uncertainty  $U_{\hat{f}_1}$  as per Eq. (3);
    Find  $\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmax}}[U_{\hat{f}_1}(\mathbf{x})]$ ; // Perform adaptive sampling
    Find  $k = \max_{\operatorname{loc}}(\mathbf{U})$  with  $\mathbf{U}$  as per Eq. (17);
    for  $i=k, N$  do // Perform new function evaluations
        Evaluate  $\{\mathbf{x}^*, f_i(\mathbf{x}^*)\}$ ;
        Update the training sets  $\{\mathcal{T}_i\} \cup \{\mathbf{x}^*, f_i(\mathbf{x}^*)\}$ ;
        Update the training sets size  $\{J_i^{t+1}\} = \{J_i^t\} + 1$ ;
    end
     $t = t + 1$ ; // Move to the next adaptive sampling iteration
end

```

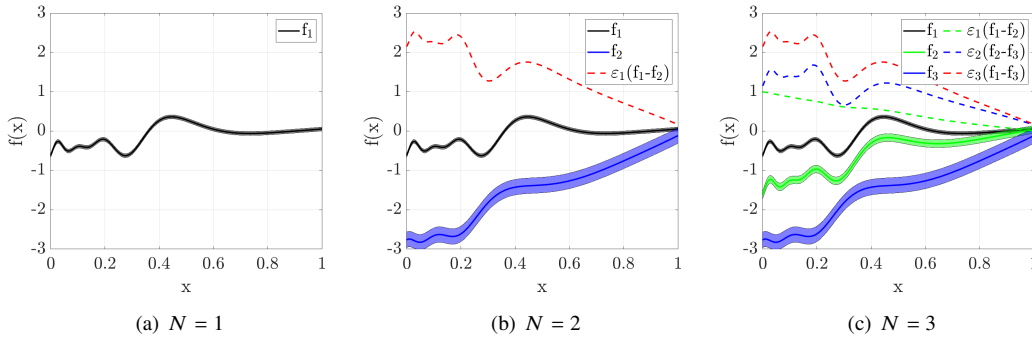


Fig. 3 Analytical test problem with different number of fidelities (N) with noise bands.

Algorithm 2: N -fidelity implementation with GP.

```

Define the initial training set for each fidelity level  $\{\mathcal{T}_i\}_{i=1}^N = \{\mathbf{y}_j, f_i(\mathbf{y}_j)\}_{j=1}^{J_i}$ ;
while stopping criteria are not met do // iterate the adaptive sampling process with index  $t$ 
    Find  $\{\sigma_{Nn}^{*2}, \sigma_{NF}^{*2}, \gamma_N^*\} = \text{argmax}[I]$  as in (13); // Evaluate GP parameters
    Construct the GP metamodels  $\hat{f}_N$  as in (10); // Low-fidelity approximation
    Compute the prediction uncertainty  $U_{\hat{f}_N}$  as in (15);
    for  $i=N-1, 1$  do // Evaluate bridge functions
        Compute the intra-level error  $\{(\mathbf{y}_j, \varepsilon_i(\mathbf{y}_j))\}_{j=1}^{J_i^t}$  as in (2);
        Find  $\{\sigma_{in}^{*2}, \sigma_{iF}^{*2}, \gamma_i^*\} = \text{argmax}[I]$  as in (13); // Evaluate GP parameters
        Construct the GP metamodels of the inter-level errors  $\hat{\varepsilon}_i$  as in (10);
        Compute the prediction uncertainty  $U_{\hat{\varepsilon}_i}$  as in (15);
    end
    Construct the  $N$ -fidelity approximation  $\hat{f}_1(\mathbf{x})$  as in (1); //  $N$ -fidelity approximation
    Compute the  $N$ -fidelity prediction uncertainty  $U_{\hat{f}_1}$  as in (15);
    Find  $\mathbf{x}^* = \text{argmax}_{\mathbf{x}}[U_{\hat{f}_1}(\mathbf{x})]$  with  $\mathbf{U}$  as per Eq. (17) or Eq. (18); // Perform adaptive sampling
    for  $i=k, N$  do // Perform new function evaluations
        Evaluate  $\{\mathbf{x}^*, f_i(\mathbf{x}^*)\}$ ;
        Update the training sets  $\{\mathcal{T}_i\} \cup \{\mathbf{x}^*, f_i(\mathbf{x}^*)\}$ ;
        Update the training sets size  $\{J_i^{t+1}\} = \{J_i^t\} + 1$ ;
    end
     $t = t + 1$ ; // Move to the next adaptive sampling iteration
end

```

VI. NACA Airfoil Shape Optimization

This problem addresses the shape optimization of a NACA four-digit airfoil operating at constant lift [24]. The following minimization problem is solved

$$\begin{aligned}
 &\text{minimize} && C_D(\mathbf{x}), \\
 &\text{subject to} && C_L(\mathbf{x}) = 0.6, \\
 &&& \text{and to } \mathbf{l} \leq \mathbf{x} \leq \mathbf{u},
 \end{aligned} \tag{19}$$

where \mathbf{x} is the design variable vector, C_D and C_L are respectively the drag and lift coefficient, and \mathbf{l} and \mathbf{u} are the lower and upper bound of \mathbf{x} . The equality constraint on the lift coefficient is necessary to compare different geometries at the same lift force (typically equal to the weight of the object), since the drag depends strongly on the lift.

The airfoil shape (see Fig. 4) is defined by the general equation for four-digit NACA foils. The upper (y_u) and lower

Table 1 Analytical test problem: f_1 and f_2 are taken from [31].

N	Problem			Domain
1	$f_1(x)$	=	$\sin(30(x - 0.9)^4) \cos(2(x - 0.9)) + (x - 0.9)/2 + \chi_1$	$[0, 1]$
2	$f_1(x)$	=	$\sin(30(x - 0.9)^4) \cos(2(x - 0.9)) + (x - 0.9)/2 + \chi_1$	$[0, 1]$
	$f_2(x)$	=	$(f_1(x) - 1 + x)/(1 + 0.25x) + \delta_2$	
3	$f_1(x)$	=	$\sin(30(x - 0.9)^4) \cos(2(x - 0.9)) + (x - 0.9)/2 + \chi_1$	$[0, 1]$
	$f_2(x)$	=	$(f_1(x) - 1 + x)/(1 + 0.25x) + \delta_2$	
	$f_3(x)$	=	$\sin(20(x - 0.87)^4) \cos(2(x - 0.87)) + (x - 0.87)/2 - (2.5 - (0.7x - 0.14)^2) + 2x + \delta_3$	

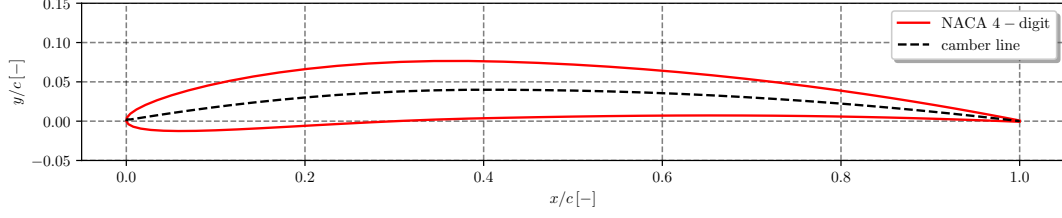


Fig. 4 NACA 4-digit airfoil.

(y_l) airfoil surfaces are computed as

$$\begin{cases} \xi_u = \xi - y_t \sin \theta \\ \xi_l = \xi + y_t \sin \theta \\ y_u = y_c + y_t \cos \theta \\ y_l = y_c - y_t \cos \theta \end{cases} \quad \text{with } y_c = \begin{cases} \frac{m}{p^2} \left[2p \frac{\xi}{c} - \left(\frac{\xi}{c} \right)^2 \right], & 0 \leq \xi < pc \\ \frac{m}{(1-p)^2} \left[(1-2p) + 2p \frac{\xi}{c} - \left(\frac{\xi}{c} \right)^2 \right], & pc \leq \xi \leq c \end{cases} \quad (20)$$

where ξ is the position along the chord, c the chord length, y_c the mean camber line, p the location of the maximum camber, m the maximum camber value, t the maximum thickness, and y_t the half thickness distribution given by

$$y_t = 5t \left(0.2969\sqrt{\xi} - 0.1260\xi - 0.3516\xi^2 + 0.2843\xi^3 - 0.1015\xi^4 \right). \quad (21)$$

In this work, the design variables vector is defined as $\mathbf{x} = \{t, m\}$ with $t \in [0.030, 0.120]$ and $m \in [0.025, 0.070]$. The maximum camber position is fixed at $p = 0.4$. The simulation conditions are: velocity $U = 10$ m/s, chord $c = 1$ m, fluid density $\rho = 1,026$ kg/m³, and Reynolds number $Re = 8.41 \cdot 10^6$ based on the chord length.

CFD simulations for the NACA airfoil are performed with the Navier-Stokes solver ISIS-CFD developed at ECN – CNRS [25], available in the FINETM/Marine computing suite from NUMECA Int.

A. ISIS-CFD

ISIS-CFD is an incompressible unstructured finite-volume solver for multifluid flow. The velocity field is obtained from the momentum conservation equations and the pressure field is extracted from the mass conservation constraint transformed into a pressure equation. These equations are similar to the Rhie and Chow SIMPLE method [32], but have been adapted for flows with discontinuous density fields. Free-surface flow is simulated with a conservation equation for the volume fraction of water, discretised with specific compressive discretisation schemes. The method features sophisticated turbulence models, such as the anisotropic EASM model and DES models.

The unstructured discretisation is face-based. While all unknown state variables are cell-centered, the systems of equations used in the implicit time stepping procedure are constructed face by face. Therefore, cells with an arbitrary number of arbitrarily-shaped constitutive faces are accepted. The code is fully parallel using the message passing

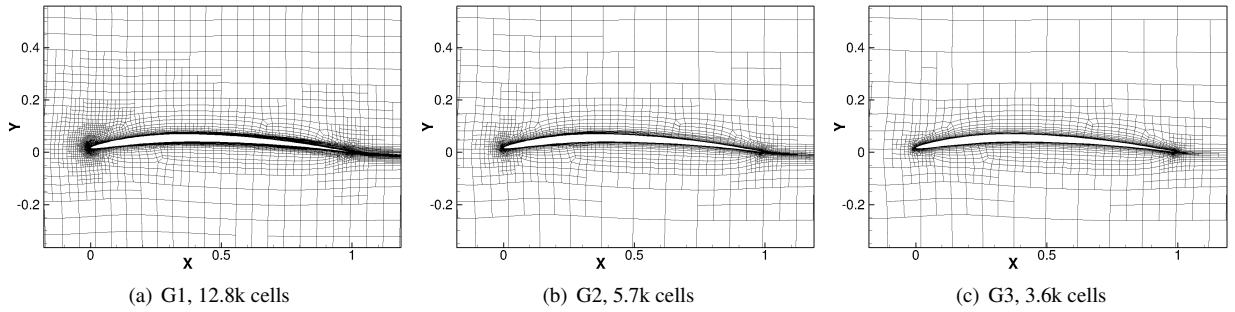


Fig. 5 NACA airfoil computational grids for ISIS-CFD.

interface (MPI) protocol. A detailed description of the solver is given by [25]. Information on the interface-capturing scheme can also be found in [33].

Computational grids are created through adaptive grid refinement [26, 34], to optimize the efficiency of the solver and to simplify the automatic creation of suitable grids. The adaptive grid refinement method adjusts the computational grid locally, during the computation, by dividing the cells of an original coarse grid. The decision where to refine comes from a refinement criterion, a tensor field $C(x, y, z)$ computed from the flow. The tensor is based on the water surface position and on second derivatives of pressure and velocity, which gives a crude indication of the local truncation errors. The grid is refined until the dimensions $\mathbf{d}_{p,j}$ ($j = 1, 2, 3$) of each hexahedral cell p satisfy

$$\|C_p \mathbf{d}_{p,j}\| = T_r. \quad (22)$$

The refinement criterion based on the second derivatives of the flow is not very sensitive to grid refinement [26], so the cell sizes everywhere are proportional to the constant threshold T_r .

For the NF optimization, grid adaptation is used to take into account the need for several fidelities. The interest of this procedure is that different fidelity results can be obtained by running the same simulations and simply changing the threshold T_r . Thus, it is straightforward to automate the NF simulations.

The computational domain runs from $11c$ in front of the leading edge to $16c$ behind the airfoil and from $-10c$ to $10c$ vertically. Dirichlet conditions on the velocity are imposed, except on the outflow side which has an imposed pressure condition. The hydrofoil surface is treated with a wall law and $y^+ = 60$ for the first layer. Turbulence is modeled with the standard $k - \omega$ SST model. To obtain the same lift for all geometries (see Eq. 19), the angle of incidence α for the airfoil is adjusted dynamically during the simulations.

In this work, the NACA airfoil shape optimization is here solved with I-SRBF varying the number of fidelities and LS-SRBF, GP1, and GP2 with $N = 3$. The initial computational grid has 2,654 cells, the refinement threshold value T_r is set equal to 0.1, 0.2, and 0.4 from highest- to lowest-fidelity. This results in a cell size ratio of 4 : 1 between the refined fine and coarse grids. The final grids (G1, G2, and G3) have about 12.8k, 5.7k, and 3.7k cells, respectively (see Fig. 5). Note that when $N = 2$, G1 and G3 grids are used. Highest- to lowest-fidelity simulations require about 17, 9, and 5 minutes, respectively, of wall-clock time to converge. The resulting computational costs are about $\beta = \{1, 0.3\}$ for $N = 2$ and $\beta = \{1, 0.5477, 0.3\}$ for $N = 3$.

VII. Numerical Results

A deterministic single-objective formulation of the particle swarm optimization algorithm [35], is used for the metamodel-based optimization, as well as for the solution of the minimization sampling problem of Eq. 13 and Eq. 16.

The performance of the NF approach with LS-SRBF, GP1, and GP2 and $N = 3$ fidelities are compared with the NF approach with I-SRBF with a number of fidelities $N = 1, 2$, and 3. Three metrics are used to assess the NF performance as follows

$$E_p = \frac{\hat{f}(\mathbf{x}_{\min}) - f(\check{\mathbf{x}})}{R_1}, \quad E_x = \|\mathbf{x}_{\min} - \check{\mathbf{x}}\|, \quad \text{and} \quad E_v = \frac{f(\mathbf{x}_{\min}) - f(\check{\mathbf{x}})}{R_1}, \quad (23)$$

where E_p is the prediction error quantifying the accuracy of the metamodel prediction at the location of the minimum; E_x and E_v are the position and validation errors respectively, which quantify the performance in identifying the global optimum; $\{\check{\mathbf{x}}, f(\check{\mathbf{x}})\}$ is the reference/global optimum; $\hat{f}(\mathbf{x}_{\min})$ the minimum predicted by the metamodel, and $f(\mathbf{x}_{\min})$ is the verified minimum by high-fidelity evaluation. Please note that a normalized domain is used.

A. Analytical Test Problem

The normalized root mean square error (NRMSE, between the metamodel prediction and a high-fidelity reference, evaluated using 2500 points with a uniform distribution) is used with the analytical test problem to assess the accuracy of the metamodel approximation. The normalization is defined with respect to R_1 . Furthermore, the capability of LS-SRBF, GP1, and GP2 in assessing the training set noise is investigated. LS-SRBF identifies the noise as the NRMSE of the training set against the metamodel prediction, whereas GP directly computes the standard deviation σ_n of the noise. The noise prediction ($\tilde{\mathbf{a}}R_1$) of LS-SRBF, GP1, and GP2 is compared with the nominal value of $\mathbf{a}R_1$ and the value of the maximum prediction uncertainty U_M . $R_1 = 0.895$ is the high-fidelity range.

The analytical problem has global minimum $\{\check{\mathbf{x}}, f(\check{\mathbf{x}})\} = \{0.275, -0.625\}$. The β values are $\beta = \{1, 0.05\}$ for $N = 2$ and $\beta = \{1, 0.1, 0.05\}$ for $N = 3$. The initial training set is represented by three samples, for all the fidelity levels. Finally, the available computational budget is the equivalent of 20 high-fidelity evaluations.

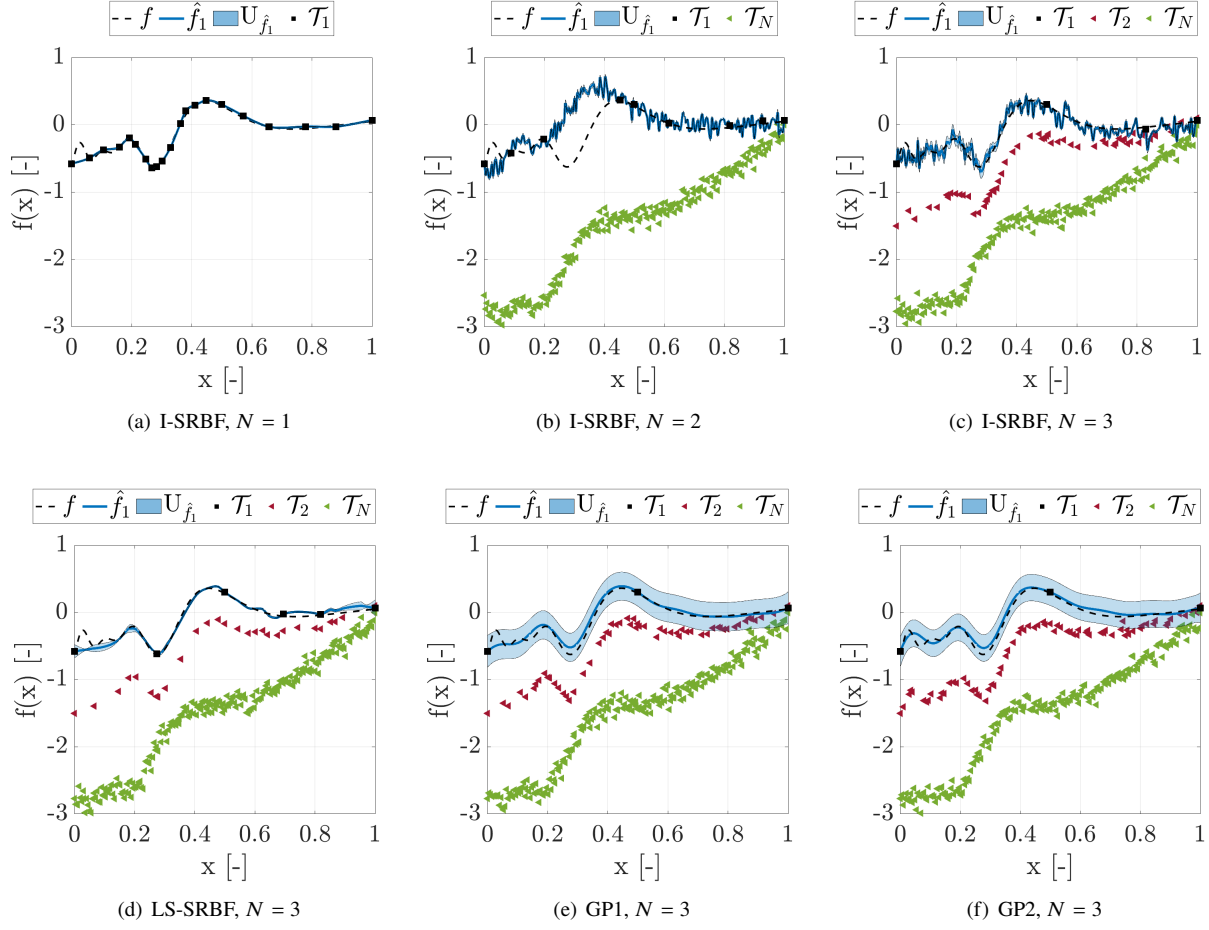


Fig. 6 Analytical test results: final iteration of the adaptive sampling procedure.

Table 2 Analytical test problem: minimum value of the function prediction $\hat{f}(x_{\min})$, its position x_{\min} , the associated prediction uncertainty $U_{\hat{f}}(x_{\min})$, the prediction $|E_p|$, validation $|E_v|$, and position $|E_x|$ errors, the NRMSE and the training sets size J . The noise bandwidth (at 95% confidence) of the noise imposed to the training set is $a = \{10, 20, 40\}\%R_1$.

Metamodel	N	x_{\min}	$U_{\hat{f}}(x_{\min}) [\%R_1]$	$\hat{f}(x_{\min})$	$f(x_{\min})$	$ E_p [\%R_1]$	$ E_v [\%R_1]$	$ E_x [\%R_1]$	NRMSE	J_1	J_2	J_3
I-SRBF	1	0.272	0.91	-0.643	-0.623	2.23	0.22	0.30	5.56	20	-	-
	2	0.000	5.01	-0.790	-0.618	19.2	0.77	27.5	33.9	9	-	221
	3	0.278	19.7	-0.708	-0.624	9.39	0.11	0.30	10.9	4	58	204
LS-SRBF	3	0.278	3.59	-0.625	-0.624	0.11	0.11	0.30	7.38	6	28	226
GP1	3	0.000	13.2	-0.577	-0.618	4.58	0.77	27.5	6.49	3	59	226
GP2	3	0.000	12.3	-0.577	-0.618	4.58	0.77	27.5	5.76	3	75	195
Reference	-	0.265	-	-	-0.625	-	-	-	-	-	-	-

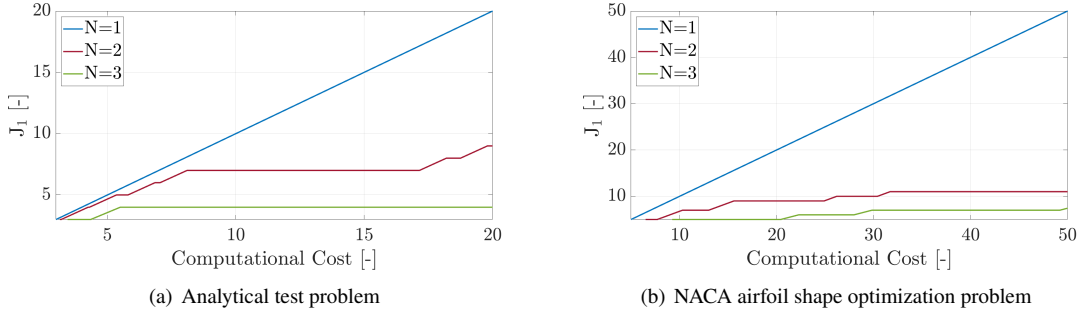


Fig. 7 I-SRBF high-fidelity training set size (J_1) versus computational cost for $N = 1$, $N = 2$, and $N = 3$.

1. Effect of N -fidelity

Figures 6a-c show the final iteration of the adaptive sampling for the I-SRBF approach with $N = 1, 2$, and 3 . For $N = 1$ the sampling is quite uniform, and the approximation of the desired function is reasonable, although the neighborhood of $x = 0.1$ is not sampled, thus missing the local maxima. For $N = 2$ the presence of the noise significantly reduces the metamodel prediction accuracy. Finally, for $N = 3$ the metamodel prediction achieves a reasonable accuracy, although the presence of the noise significantly affects the results. It is worth noting that (for $N = 3$) medium-fidelity samples are placed in the local maxima. Table 2 summarizes the performance of the I-SRBF approach. The lowest validation error is achieved for $N = 3$. The results show that the use of an intermediate fidelity level improves the performance of the method. Finally, Fig. 7a depicts the I-SRBF high-fidelity training set size and shows how the increase of the number of fidelities significantly reduces the number of high-fidelity evaluations.

2. Comparison Between Interpolation and Least Squares Approximation

Figures 6c and 6d show the final iteration for $N = 3$ of the adaptive sampling for I-SRBF and LS-SRBF, respectively. LS-SRBF provides a significantly smoother approximation of the desired function, clearly showing that the least squares approximation is effective in filtering out the noise. The performance achieved by I-SRBF and LS-SRBF are summarized in Tab. 2. LS-SRBF achieves the same validation error $|E_v|$ than I-SRBF, whereas the prediction error and NRMSE values are significantly improved: from 9.4 to 0.1% and from 10.9 to 7.4%, respectively. Finally, LS-SRBF achieves a lower prediction uncertainty in the location of the global minimum \check{x} .

3. Comparison Between LS-SRBF and GP

Figures 6d and 6e show the final iteration for $N = 3$ of the adaptive sampling for LS-SRBF and GP1, respectively. Both metamodels achieve a smooth representation of the desired function. GP1 shows a higher prediction uncertainty, because such uncertainty includes the contribution of σ_n . The results in Tab. 2 show that GP1 achieves higher values of the validation, position, and prediction errors, whereas it achieves lower values of NRMSE than LS-SRBF. Finally, GP1 significantly reduces the number of high-fidelity evaluations and increases the number of medium-fidelity training points in comparison to LS-SRBF.

4. Effect of σ_n on the GP Adaptive Sampling Method

Figures 6e and 6f show the final iteration for $N = 3$ of the adaptive sampling for GP1 and GP2, respectively. GP2 achieves a slightly better accuracy of the prediction, especially in the neighborhood of $x = 0.1$. Table 2 summarizes the performance of GP1 and GP2. GP2 achieves the same values of the validation, position and prediction errors, but a lower NRMSE than GP1. Finally, using GP2 as opposed to GP1 does not increase the number of high-fidelity evaluations, leads to more medium-fidelity samples with a more uniform distribution, and finally requires less low-fidelity samples. In the presence of low-fidelity noise, this improves the prediction accuracy.

Table 3 summarizes the predicted noise \tilde{a}_i and the maximum prediction uncertainty $U_{M,i}$ for each fidelity level i . For $N = 3$ LS-SRBF, GP1, and GP2 achieve a reasonable prediction of the nominal noise for the lower fidelity ($i = 3$). GP2 is the only method able to achieve a reasonable prediction also for the medium-fidelity noise. This is because GP1 is sampling more the low-fidelity and therefore does not have enough medium-fidelity sample to model the associated

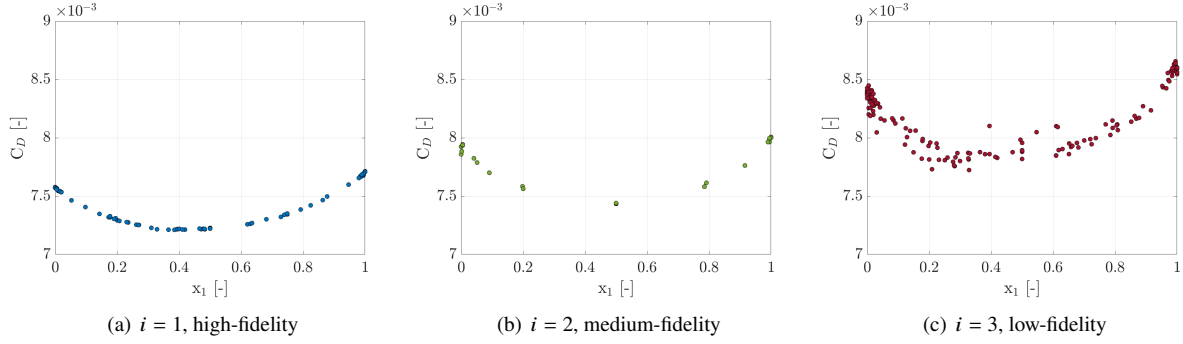


Fig. 8 NACA airfoil: samples laying in $x_2 = 0$, combined over all the datasets.

noise. It is worth noting that the maximum prediction uncertainty of LS-SRBF is always lower than the noise prediction, whereas for GP the maximum prediction uncertainty is always higher than the noise prediction.

B. NACA Airfoil Shape Optimization

The available computational budget is the equivalent of 50 high-fidelity evaluations. $R_1 = 1.5257 \cdot 10^{-3}$ is the high-fidelity range.

Figure 8 shows the high-, medium-, and low-fidelity samples laying in $x_2 = 0$, combining the data for all the metamodels and sampling methods. The numerical noise affects significantly the low-fidelity training set, see Fig. 8c. A lower numerical noise is also visible in the extrema of the medium-fidelity level, whereas the high-fidelity shows very small numerical noise, see Figs. 8b and 8a respectively.

Table 3 Analytical test problem with $N = 3$: predicted noise \tilde{a}_i and maximum prediction uncertainty in brackets ($U_{M,i}$) for the i -th fidelity level. The noise bandwidth (at 95% confidence) of the noise imposed to the training set is $\mathbf{a} = \{10, 20, 40\}\%R_1$.

Approach	$\tilde{a}_1 (U_{M,1}) [\%R_1]$	$\tilde{a}_2 (U_{M,2}) [\%R_1]$	$\tilde{a}_3 (U_{M,3}) [\%R_1]$
LS-SRBF	75.3 (16.0)	80.6 (22.8)	41.3 (18.4)
GP1	0.00 (4.20)	25.7 (26.2)	40.4 (41.2)
GP2	0.00 (15.3)	18.2 (23.5)	37.5 (42.3)

Table 4 NACA airfoil: minimum value of the drag coefficient prediction $\hat{C}_D(\mathbf{x}_{\min})$, its position \mathbf{x}_{\min} , the associated prediction uncertainty $U_{\hat{C}_D}(\mathbf{x}_{\min})$, the prediction $|E_P|$, validation $|E_V|$, and position $|E_X|$ errors, and the training sets size J .

Metamodel	N	$x_{\min 1} [-]$	$x_{\min 2} [-]$	$U_{\hat{C}_D}(\mathbf{x}_{\min}) [\%R_1]$	$\hat{C}_D(\mathbf{x}_{\min})$	$C_D(\mathbf{x}_{\min})$	$ E_P [\%R_1]$	$ E_V [\%R_1]$	$ E_X [\%R_1]$	J_1	J_2	J_3
I-SRBF	1	0.3897	0.0000	0.02	7.2169E-03	7.2207E-03	0.25	0.59	1.21	50	-	-
	2	0.2080	0.0000	1.78	7.0082E-03	7.2988E-03	19.0	5.71	16.7	11	-	134
	3	0.2805	0.0011	2.41	7.1486E-03	7.2397E-03	5.97	1.84	9.71	8	26	96
LS-SRBF	3	0.3903	0.0000	1.25	7.2023E-03	7.2182E-03	1.04	0.43	1.27	7	24	104
GP1	3	0.4797	0.0000	40.4	7.2312E-03	7.2261E-03	0.33	0.95	10.2	6	9	134
GP2	3	0.4999	0.0000	135	7.2048E-03	7.2303E-03	1.67	1.22	12.2	9	20	102
Reference	1	0.3776	0.0000	-	-	7.2116E-03	-	-	-	150	-	-

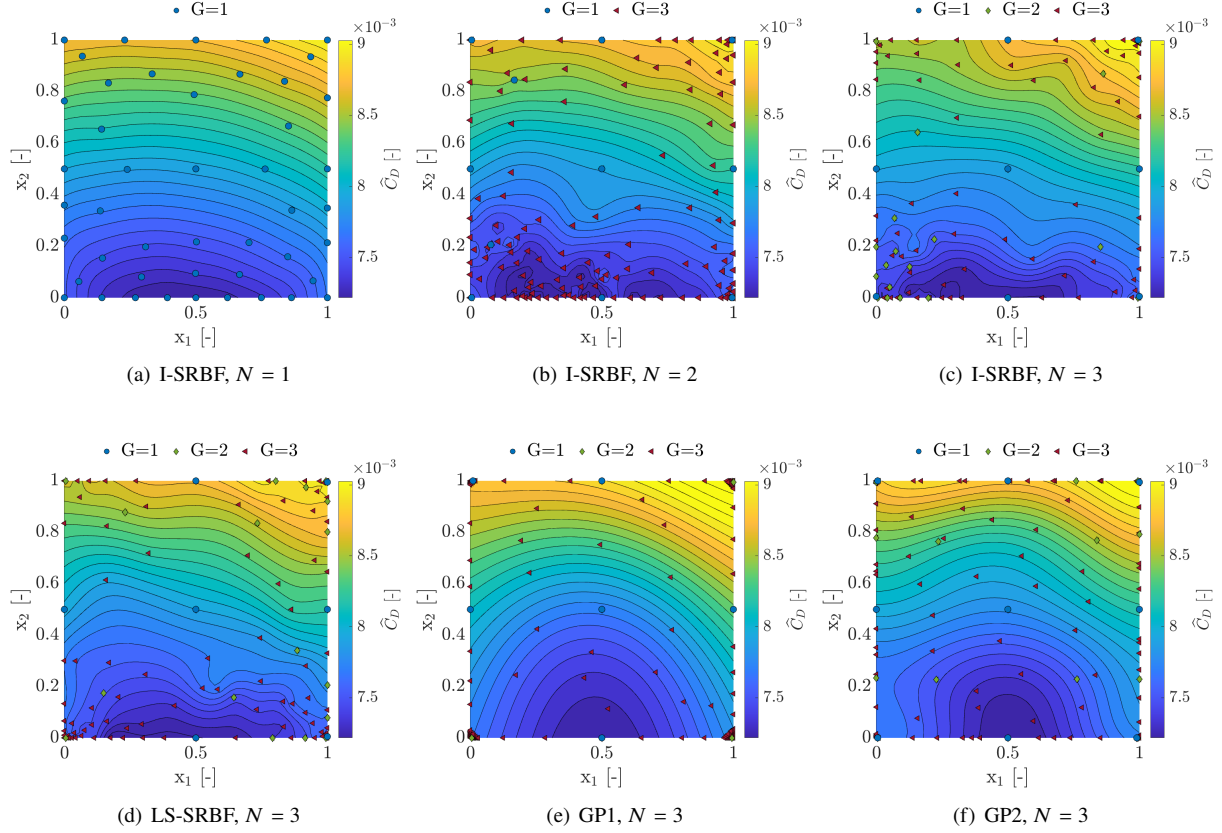


Fig. 9 NACA airfoil: contour plots of \hat{C}_D .

1. Effect of N -fidelity

Figures 9a-c show the contour plots of the drag coefficient prediction at the final iteration of the adaptive sampling method for I-SRBF. I-SRBF with $N = 1$ performs an almost uniform sampling of the domain, there are no specific zones where the samples are significantly clustered, which means that the noise at this fidelity level is low. For $N = 2$ the low-fidelity samples are significantly clustered at the corners and in the region of the minimum, thus producing a response surface that is multi-modal with several local minima. Finally, for $N = 3$, the clustering of the low-fidelity samples is reduced, especially in the region of the minimum where also the low-fidelity samples have been reduced in favor of medium-fidelity samples. Figures 10a-c show the contour plot of the drag coefficient prediction uncertainty at the final iteration of the adaptive sampling method for I-SRBF. The prediction uncertainty is low for $N = 1$, $N = 2$, and $N = 3$, with a smooth pattern for $N = 1$. The noise in the low-fidelity training set is evident in the pattern of the prediction uncertainty for $N = 2$, as well as the smoothing effect of introducing a medium-fidelity ($N = 3$). Table 4 summarizes the minimum value of the drag coefficient prediction $\hat{C}_D(\mathbf{x}_{\min})$, its position \mathbf{x}_{\min} , the associated prediction uncertainty $U_{\hat{C}_D}(\mathbf{x}_{\min})$, the prediction $|E_p|$, validation $|E_v|$, and position $|E_x|$ errors, and the training sets size for I-SRBF, LS-SRBF, GP1, and GP2. I-SRBF with $N = 2$ achieves a 78% reduction of the number of high-fidelity evaluations, but the significant noise in the low-fidelity leads to an increase of the prediction, validation, and position errors. The introduction of an intermediate fidelity level ($N = 3$) leads to a further reduction of the high-fidelity evaluations, plus a shift from low- to medium-fidelity evaluations, see Fig. 7b, which significantly improves the prediction, validation, and position errors.

2. Comparison Between Interpolation and Least Squares Approximation

Figure 9d shows the contour plot of the drag coefficient at the final iteration of the adaptive sampling method for LS-SRBF with $N = 3$. The samples are significantly more uniformly distributed than for I-SRBF, with a mild

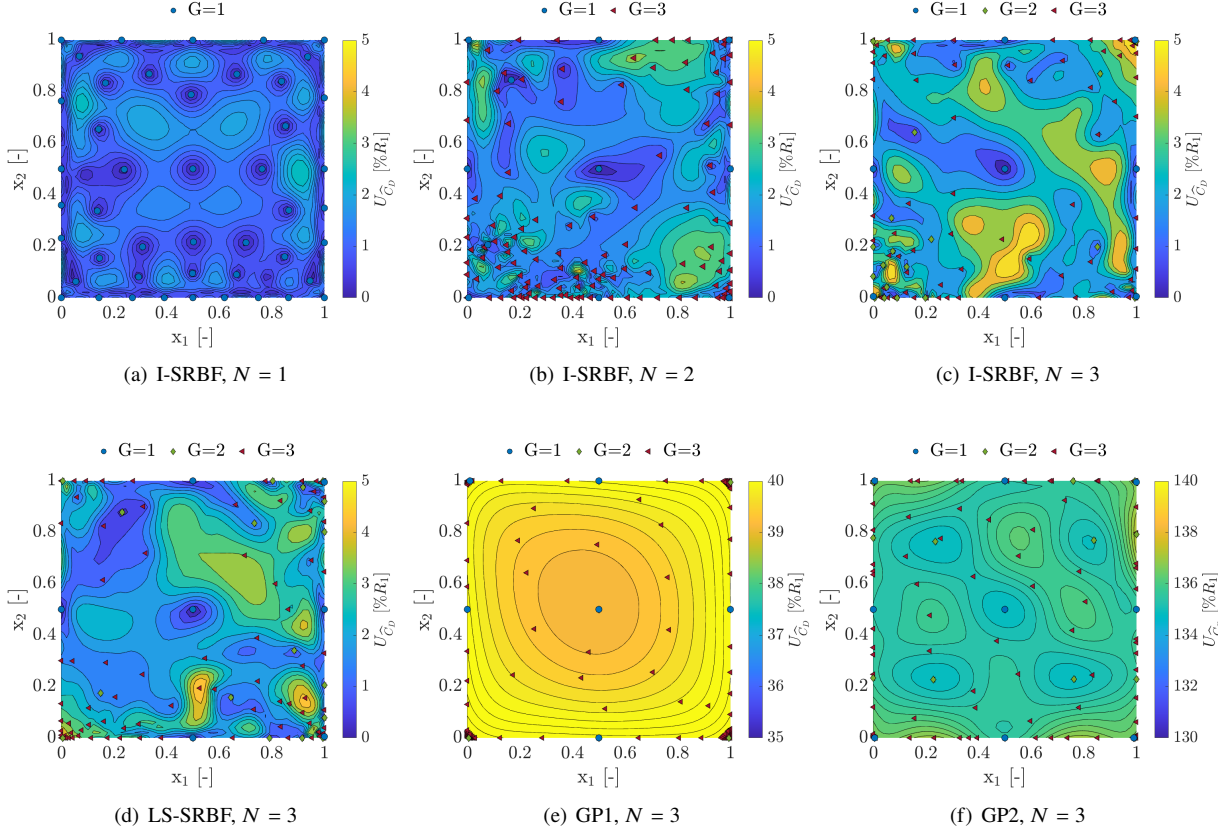


Fig. 10 NACA airfoil: contour plots of the N -fidelity prediction uncertainty $U_{\hat{C}_D}$. Note that the color scales are different among SRBF, GP1, and GP2.

clusterization at the $\{0, 0\}$ corner. LS-SRBF provides a smoother response surface without local minima and with a closer similarity to I-SRBF with $N = 1$, especially in representing the global minimum region. Figure 10d shows the contour plot of the drag coefficient prediction uncertainty at the final iteration of the adaptive sampling method for LS-SRBF. The prediction uncertainty has comparable values to and is more regular than I-SRBF (see, Fig. 10c). The results in Tab. 4 show that the use of LS-SRBF produces a slight reduction of the number of high-fidelity evaluations and leads to a significant reduction of the prediction, validation, and position errors with respect to I-SRBF. Furthermore, the validation error is improved with respect to I-SRBF with $N = 1$.

3. Comparison Between SRBF and GP

Figure 9e shows the contour plot of the drag coefficient at the final iteration of the adaptive sampling method for GP1. A significant clusterization of low-fidelity samples is present at the corners. Figure 10d shows the contour plot of the drag coefficient prediction uncertainty at the final iteration of the adaptive sampling method for GP1. The prediction uncertainty is higher than LS-SRBF because GP1 includes in the prediction uncertainty the noise contribution (proportional to $4\sigma_n$ with $\sigma_n = 4.5\%R_1$), but is also significantly smoother. The results in Tab. 4 show that GP1 achieves similar results to LS-SRBF, with a slight reduction of the number of high-fidelity samples and a significant reduction of the number of medium-fidelity samples. Furthermore, GP1 achieves lower values of the prediction errors.

4. Effect of σ_n on the GP Adaptive Sampling Method

Figure 9e shows the contour plot of the drag coefficient at the final iteration of the adaptive sampling method for GP2. A significantly more uniform sampling than GP1 is shown, with a negligible clusterization at the corners. Figure 10e shows the contour plot of the drag coefficient prediction uncertainty at the final iteration of the adaptive sampling

method for GP2. The overall prediction uncertainty is significantly higher than the other approaches, this is because GP2 evaluates a value of $\sigma_n = 33\%R_1$. For comparison, GP1 evaluates a value of $\sigma_n = 4.5\%R_1$. Considering the noise of Fig. 8c, the GP2 prediction looks more reasonable than that of GP1. Considering the variations of the prediction uncertainty value, GP2 provides a smooth response surface, similar to the one of I-SRBF with $N = 1$ (see, Fig. 10a). The results in Tab. 4 show that GP2 requires more samples from the medium- and high-fidelity levels than GP1. This looks to be strictly connected to the fact that GP1 clusterizes low-fidelity samples at the corners, whereas GP2 performs a more uniform exploration.

VIII. Conclusions and Future Work

An adaptive N -fidelity (NF) metamodel for noisy CFD data has been presented for design space exploration and design optimization. The NF approximation is obtained as the sum of a low-fidelity-trained metamodel and the metamodel of inter-level errors (difference). Two metamodeling techniques have been used, namely stochastic radial basis functions (SRBF) and Gaussian process (GP). SRBF and GP provides the metamodel prediction and the associated uncertainty. The adaptivity stems from the sequential training procedure and the auto-tuning capabilities of the metamodels. New training points are added to the appropriate fidelity level, based on the NF prediction uncertainty and the computational cost. SRBF has been used with exact interpolation (I-SRBF) and least squares regression (LS-SRBF). Two multi-fidelity adaptive sampling methods have been used with GP, namely GP1 and GP2.

The methods have been first demonstrated for an analytical test function and the shape optimization of a NACA four-digit airfoil. CFD simulations have been performed with the RANS code ISIS-CFD, where three fidelity levels have been used based on an adaptive grid refinement technique. The following conclusions have been achieved.

- 1) Effect of N -fidelity: using more than one fidelity level reduces the number of high-fidelity evaluations required to accurate function prediction and optimization; using intermediate fidelity levels has a beneficial effect when dealing with noisy data.
- 2) Effect of using LS-SRBF as opposed to I-SRBF: the adaptive least squares regression significantly improves the function prediction and optimization by effectively filtering out the noise in the training data.
- 3) Effect of using LS-SRBF versus GP: LS-SRBF and GP show very close performance; nevertheless, current GP1 method is found slightly more efficient and accurate than other methods, at least for current problems.
- 4) Effect of using GP1 versus GP2: the performance are very close; nevertheless, GP2 achieves a higher accuracy by distributing more uniformly the training samples in the design space and across the fidelity levels; moreover, GP2 is found the most accurate in assessing the noise in the training data.

Ongoing activities are focused on testing different adaptive sampling methods in addition to MUAS. Specifically, aggregate criteria based on both the maximum prediction uncertainty and the function value are currently investigated [20]. Finally, different strategies for choosing the number and position of the RBF centers will be considered, to further improve the robustness of the regression.

Acknowledgments

CNR-INM is grateful to Dr. Woei-Min Lin, Dr. Elena McCarthy, and Dr. Salahuddin Ahmed of the Office of Naval Research and Office of Naval Research Global, for their support through NICOP grant N62909-18-1-2033. Dr. Riccardo Pellegrini is partially supported through CNR-INM project OPTIMAE.

References

- [1] Viana, F. A. C., Simpson, T. W., Balabanov, V., and Vasilli, T., "Special Section on Multidisciplinary Design Optimization: Metamodeling in Multidisciplinary Design Optimization: How Far Have We Really Come?" *AIAA Journal*, Vol. 52, No. 4, 2014, pp. 670–690.
- [2] Liu, H., Ong, Y.-S., and Cai, J., "A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design," *Structural and Multidisciplinary Optimization*, Vol. 57, No. 1, 2018, pp. 393–416.
- [3] Volpi, S., Diez, M., Gaul, N. J., Song, H., Iemma, U., Choi, K. K., Campana, E. F., and Stern, F., "Development and validation of a dynamic metamodel based on stochastic radial basis functions and uncertainty quantification," *Structural and Multidisciplinary Optimization*, Vol. 51, No. 2, 2015, pp. 347–368.
- [4] Serani, A., Pellegrini, R., Wackers, J., Jeanson, C.-E., Queutey, P., Visonneau, M., and Diez, M., "Adaptive multi-fidelity

- sampling for CFD-based optimisation via radial basis function metamodels,” *International Journal of Computational Fluid Dynamics*, 2019, pp. 1–19.
- [5] Jones, D. R., Schonlau, M., and Welch, W. J., “Efficient global optimization of expensive black-box functions,” *Journal of Global optimization*, Vol. 13, No. 4, 1998, pp. 455–492.
 - [6] Diez, M., Volpi, S., Serani, A., Stern, F., and Campana, E. F., “Simulation-based design optimization by sequential multi-criterion adaptive sampling and dynamic radial basis functions,” *Advances in Evolutionary and Deterministic Methods for Design, Optimization and Control in Engineering and Sciences*, Springer, 2019, pp. 213–228.
 - [7] Cox, D. D., and John, S., “A statistical method for global optimization,” *[Proceedings] 1992 IEEE International Conference on Systems, Man, and Cybernetics*, IEEE, 1992, pp. 1241–1246.
 - [8] Wackers, J., Pellegrini, R., Serani, A., Diez, M., and Visonneau, M., “Adaptive Multifidelity Shape Optimization based on Noisy CFD Data,” *Proceedings of the 2019 International Conference on Adaptive Modeling and Simulation (ADMOS 2019)*, El Campello (Alicante), Spain, 27-29 May, 2019.
 - [9] Peherstorfer, B., Willcox, K., and Gunzburger, M., “Survey of multifidelity methods in uncertainty propagation, inference, and optimization,” *Siam Review*, Vol. 60, No. 3, 2018, pp. 550–591.
 - [10] Park, C., Haftka, R. T., and Kim, N. H., “Remarks on multi-fidelity surrogates,” *Structural and Multidisciplinary Optimization*, Vol. 55, No. 3, 2017, pp. 1029–1050.
 - [11] Han, Z.-H., Görtz, S., and Zimmermann, R., “Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function,” *Aerospace Science and Technology*, Vol. 25, No. 1, 2013, pp. 177 – 189.
 - [12] Fernandez-Godino, M. G., Park, C., Kim, N.-H., and Haftka, R. T., “Review of multi-fidelity models,” *arXiv preprint arXiv:1609.07196*, 2016.
 - [13] Rumpfkeil, M. P., and Beran, P. S., “Multi-Fidelity, Gradient-enhanced, and Locally Optimized Sparse Polynomial Chaos and Kriging Surrogate Models Applied to Benchmark Problems,” *AIAA Scitech 2020 Forum*, 2020, p. 0677.
 - [14] Coppedè, A., Gaggero, S., Vernengo, G., and Villa, D., “Hydrodynamic shape optimization by high fidelity CFD solver and Gaussian process based response surface method,” *Applied Ocean Research*, Vol. 90, 2019, p. 101841.
 - [15] Baar, J. d., Roberts, S., Dwight, R., and Mallol, B., “Uncertainty quantification for a sailing yacht hull, using multi-fidelity kriging,” *Computers & Fluids*, Vol. 123, 2015, pp. 185 – 201.
 - [16] Serani, A., Pellegrini, R., Wackers, J., Jeanson, C.-E., Queutey, P., Visonneau, M., and Diez, M., “Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function metamodels,” *International Journal of Computational Fluid Dynamics*, Vol. 33, No. 6-7, 2019, pp. 237–255.
 - [17] Beran, P. S., Bryson, D. E., Thelen, A. S., Diez, M., and Serani, A., “Comparison of Multi-Fidelity Approaches for Military Vehicle Design,” *21th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (MA&O), AVIATION 2020*, Virtual Event, June 15-19, 2020.
 - [18] Serani, A., Pellegrini, R., Broglia, R., Wackers, J., Visonneau, M., and Diez, M., “An Adaptive N-Fidelity Metamodel for Design and Operational-Uncertainty Space Exploration of Complex Industrail Problems,” *Proceedings of the 8th International Conference on Computational Methods in Marine Engineering (Marine 2019)*, 2019, pp. 177–188.
 - [19] Huang, D., Allen, T. T., Notz, W. I., and Miller, R. A., “Sequential kriging optimization using multiple-fidelity evaluations,” *Structural and Multidisciplinary Optimization*, Vol. 32, No. 5, 2006, pp. 369–382.
 - [20] Jiang, P., Cheng, J., Zhou, Q., Shu, L., and Hu, J., “Variable-Fidelity Lower Confidence Bounding Approach for Engineering Optimization Problems with Expensive Simulations,” *AIAA Journal*, 2019, pp. 1–15.
 - [21] Cai, X., Qiu, H., Gao, L., Wei, L., and Shao, X., “Adaptive Radial-Basis-Function-Based Multifidelity Metamodeling for Expensive Black-Box Problems,” *AIAA Journal*, Vol. 55, No. 7, 2017, pp. 2424–2436.
 - [22] Pellegrini, R., Serani, A., Diez, M., Wackers, J., Queutey, P., and Visonneau, M., “Adaptive Sampling Criteria for Multi-fidelity Metamodels in CFD-based Shape Optimization,” *7th European Conference on Computational Fluid Dynamics (ECFD 7)*, Glasgow, UK, 11-15 June, 2018.

- [23] Serani, A., Pellegrini, R., Broglia, R., Wackers, J., Visonneau, M., and Diez, M., “An adaptive N-fidelity metamodel for design and operational-uncertainty space exploration of complex industrial problems,” *Proceedings of the 8th International Conference on Computational Methods in Marine Engineering (Marine 2019)*, Göteborg, Sweden, 13-15 May, 2019.
- [24] Wackers, J., Pellegrini, R., Serani, A., Diez, M., and Visonneau, M., “Adaptive Multifidelity Shape Optimization based on Noisy CFD Data,” *Proceedings of the 2019 International Conference on Adaptive Modeling and Simulation (ADMOS 2019)*, El Campello (Alicante), Spain, 27-29 May, 2019.
- [25] Queutey, P., and Visonneau, M., “An Interface Capturing Method for Free-Surface Hydrodynamic Flows,” *Computers & Fluids*, Vol. 36, No. 9, 2007, pp. 1481–1510.
- [26] Wackers, J., Deng, G., Guilmineau, E., Leroyer, A., Queutey, P., Visonneau, M., Palmieri, A., and Liverani, A., “Can adaptive grid refinement produce grid-independent solutions for incompressible flows?” *Journal of Computational Physics*, Vol. 344, 2017, pp. 364 – 380.
- [27] Sparks, D., “Algorithm AS 58: Euclidean Cluster Analysis,” *Applied Statistics*, Vol. 22, No. 1, 1973, pp. 126–130.
- [28] Li, X., Gao, W., Gu, L., Gong, C., Jing, Z., and Su, H., “A cooperative radial basis function method for variable-fidelity surrogate modeling,” *Structural and Multidisciplinary Optimization*, Vol. 56, No. 5, 2017, pp. 1077–1092.
- [29] Williams, C. K., and Rasmussen, C. E., *Gaussian Process for machine learning*, Cambridge, the MIT press, 2006.
- [30] Pellegrini, R., Iemma, U., Leotardi, C., Campana, E. F., and Diez, M., “Multi-fidelity Adaptive global metamodel of expensive computer simulations,” *2016 IEEE Congress on Evolutionary Computation (CEC)*, 2016, pp. 4444–4451.
- [31] Clark, D. L., Bae, H.-r., Gobal, K., and Penmetsa, R., “Engineering Design Exploration Utilizing Locally-Optimized Covariance Kriging,” *18th AIAA Non-Deterministic Approaches Conference*, 2016, p. 0428.
- [32] Rhie, C. M., and Chow, W. L., “A numerical Study of the Turbulent Flow Past an Isolated Airfoil With Trailing Edge Separation,” *AIAA Journal*, Vol. 17, 1983, pp. 1525–1532.
- [33] Wackers, J., Koren, B., Raven, H. C., Ploeg, A. v. d., Starke, A. R., Deng, G. B., Queutey, P., Visonneau, M., Hino, T., and Ohashi, K., “Free-surface viscous flow solution methods for ship hydrodynamics,” *Archives of Computational Methods in Engineering*, Vol. 18, 2011, pp. 1–41.
- [34] Wackers, J., Deng, G., Guilmineau, E., Leroyer, A., Queutey, P., and Visonneau, M., “Combined refinement criteria for anisotropic grid refinement in free-surface flow simulation,” *Computers & Fluids*, Vol. 92, 2014, pp. 209 – 222.
- [35] Serani, A., Leotardi, C., Iemma, U., Campana, E. F., Fasano, G., and Diez, M., “Parameter selection in synchronous and asynchronous deterministic particle swarm optimization for ship hydrodynamics problems,” *Applied Soft Computing*, Vol. 49, 2016, pp. 313 – 334.