A Non-Nested Infilling Strategy for Multi-Fidelity based Efficient Global Optimization

Matthieu Sacher[a,f,*], Olivier Le Maître[b], Régis Duvigneau[c], Frédéric Hauville[a], Mathieu Durand[d], Corentin Lothodé[e]

[a]*Naval Academy Research Institute - IRENAV CC600, 29240 BREST Cedex 9, France*
[b]*CMAP, CNRS, Inria, Ecole Polytechnique, Route de Saclay, 91128 Palaiseau, France*
[c]*Université Côte d'Azur, Inria, CNRS, LJAD, 2004 Route des Lucioles, 06902 Valbonne, France*
[d]*Sirli Innovations, Pornichet, France*
[e]*K-Epsilon, 1300 Route des Crêtes, 06560 Valbonne, France*
[f]*Present affiliation: ENSTA Bretagne, CNRS, IRDL, Brest, France*

Abstract

Efficient Global Optimization (EGO) has become a standard approach for the global optimization of complex systems with high computational costs. EGO uses a training set of objective function values computed at selected input points to construct a statistical surrogate model, with low evaluation cost, on which the optimization procedure is applied. The training set is sequentially enriched, selecting new points, according to a prescribed infilling strategy, in order to converge to the optimum of the original costly model. Multi-fidelity approaches combining evaluations of the quantity of interest at different fidelity levels have been recently introduced to reduce the computational cost of building a global surrogate model. However, the use of multi-fidelity approaches in the context of EGO is still a research topic. In this work, we propose a new effective infilling strategy for multi-fidelity EGO. Our infilling strategy has the particularity of relying on non-nested training sets, a characteristic that comes with several computational benefits. For the enrichment of the multi-fidelity training set, the strategy selects the next input point together with the fidelity level of the objective function evaluation. This characteristic is in contrast with previous nested approaches, which require estimation all lower fidelity levels and are more demanding to update the surrogate. The resulting EGO procedure achieves a significantly reduced computational cost, avoiding computations at useless fidelity levels whenever possible, but it is also more robust to low correlations between levels and noisy estimations. Analytical problems are used to test and illustrate the efficiency of the method. It is finally applied to the optimization of a fully nonlinear fluid-structure interaction system to demonstrate its feasibility on real large-scale problems, with fidelity levels mixing physical approximations in the constitutive models and discretization refinements.

Keywords: Global optimization, Gaussian process model, multi-fidelity, non-nested datasets

1. Introduction

Efficient Global Optimization (EGO) strategies [13] have been used in various application domains, and many works demonstrated their applicability for the optimization of complex systems with costly objective function evaluation [28]. Examples of applications are the aerodynamic drag reduction [12], the vibration reduction for rotating aircraft [9], the optimization of Fluid-Structure Interactions (FSI) problems [1] and the sail trimming optimization [27]. In a nutshell, the EGO method builds a statistical surrogate model of the costly objective function f to be optimized, usually a Gaussian Processes [16] (GP), and a statistical criterion accounting for the Expected Improvement (EI), often referred to as *merit function*, is maximized to select a new design point. The objective function is then evaluated at the new-design point to enrich the training set, the surrogate is updated, and the algorithm proceeds with the selection of a new design point or eventually stops if it has converged.

*Corresponding author
Email address: matthieu.sacher@ensta-bretagne.fr (Matthieu Sacher)

Numerical models with different fidelity level and computation cost can be used to approximate a physical problem. As an example, the fidelity levels may correspond to finite element models using different mesh refinements, therefore having variable computational cost and accuracy. In the perspective of mitigating the computational complexity, the authors of [15] proposed a multi-fidelity (MF) surrogates combining estimations from different fidelity models. However, the use of MF approaches in the EGO framework is not yet widespread. The authors of [8] introduced an EGO formulation in the MF framework. However, this approach accounts for two-fidelity levels only (also called *co-kriging*), and the merit function does not account for the fidelity level as the method uses the prediction of the highest fidelity surrogate to estimate the merit function. The work of [7] investigates the resolution of an optimization problem involving seven design variables and using two Computational Fluid Dynamics (CFD) models having different fidelity. The authors of the paper propose five infilling strategies, all focusing on the global reduction of the MF-surrogates uncertainty and the selection of the CFD model to use for the estimation of the objective function at the new point. Therefore, this approach is not a formal EGO method, as it does not account for the optimization objective. Moreover, infilling strategies based on the global reduction of the surrogate uncertainty are known to result in sub-optimal approaches as they waste resources to improve the surrogate in regions of poor objective value instead of focusing on areas with substantial expected improvement [21].

In [17], the authors have proposed a nested infilling strategy for the optimation based on MF-surrogates. The nested character means that the sequence of training sets at successive fidelity levels are nested; specifically, the set of training points at level l is a subset of the training points at level $l-1$. In their work, the merit function uses the predictive variance of the highest fidelity surrogate. Then, an approximation of the integrated variance predictor [22] is employed to compute the potential variance reduction of the highest fidelity surrogate, after the enrichment of the training sets up to level l, with the same training point (thus requiring the evaluation of all models with fidelity level $\leq l$). This maximum fidelity level l is determined by balancing the resulting integrated variance reduction with the computational cost of the l model evaluation. To sum up, the method in [17] has the advantage of incorporating a fidelity level selection in the infilling strategy, but it has several drawbacks. First, this strategy is constrained to nested training sets, and so it may not be optimal. Besides, the method tends to favor low fidelity levels (with low computational cost) in addition to use a merit function only based on the variance that is known to be of poor efficiency for optimization problem applications.

Recently, the problem of applying EGO to airfoil shape optimization in a multi-fidelity framework was considered in [3]. Nested datasets are used to build MF-surrogates, and the Expected Improvement merit function is computed on the highest fidelity level. This work also proposed an infilling criterion for the selection of the fidelity level. The criterion corresponds to the ratio of the variance reduction to the associated squared cost of the highest fidelity level to evaluate. From our perspective, this strategy presents several drawbacks. Indeed, the use of nested datasets may not be optimal. Besides, the merit function and the fidelity level selection are decoupled. Finally, the ratio of variance to cost is not a dimensionless number, and that may raise difficulties if the costs are not well known.

In the present work, we propose to extend the original EI-based merit function of the EGO method to the case of objective functions estimated using multiple fidelity models at non-nested sets of training points. We propose a novel merit function that selects the next design point and the fidelity level of the model to be evaluated (enrichment level). This infilling method accounts for the computational cost and the potential reduction of the predictive variance of the highest fidelity surrogate, within the region of interest (susceptible to contain the optimum). Several advantages arise from this new approach. The main one concerns a potential reduction of the computational cost compared to the single-fidelity EGO (*i.e.*, all evaluations based on the highest fidelity model) and MF methods based on nested training sets. Using non-nested training sets presents the positive aspect of naturally selecting the optimal fidelity level, regarding cost/variance-reduction trade-off, while avoiding the computations of useless lower fidelity models. Besides, our numerical tests also demonstrate that our approach is more robust than the nested approach when the correlation between the successive levels is weak or in the presence of evaluation noise.

The paper structure is as follows. Section 2 concerns the optimization methods; we review the constituents of the classical single-fidelity EGO (SF-EGO) in Section 2.1. We detail the construction of MF surrogates in Section 2.2 and our infilling strategy for non-nested training sets in Section 3. Section 4 then presents several analytical test problems to compare the efficiency of the single-fidelity EGO with the nested and non-nested MF-EGO approaches. The optimization of a complex nonlinear FSI model is subsequently considered in Section 5, with the maximization of the aerodynamic performance of a Figaro racing yacht through the optimization of the two sail trimming variables. The multi-fidelity framework involves five numerical models for the computation of the objective function, which

differ in their discretization levels and physical models simplifications of both the structural and fluid solvers involved in the FSI problem. Finally, the major conclusions of this work are proposed in Section 6.

2. GP-based optimization

Our objective is to determine the optimum design variable $\boldsymbol{x}_{\text{opt}}$, the solution of the following generic optimization problem

$$\boldsymbol{x}_{\text{opt}} = \arg\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x}$ is the vector of design variables, $\Omega \subset \mathbb{R}^d$ is the admissible optimization domain and $f : \Omega \mapsto \mathbb{R}$ is the objective function. The global optimum of f can be very costly to find when, in particular, individual evaluations of $f(\boldsymbol{x})$ are numerically expensive. A classical approach is then to use a surrogate model in place of f to reduce the computational burden related to the evaluations of f [28]. In this work, we consider the use of Gaussian processes (GP) [16], which, owing to their statistical nature, provide both a prediction of the objective function value and a measure of the uncertainty (variance) in this prediction. These features are appealing in the optimization context, as they can be exploited to derive rigorous optimization strategies based on the maximization of an Expected Improvement (EI) criterion, and leading to approaches globally referred to as Efficient Global Optimization (EGO) [13] methods.

We summarize in Section 2.1 the construction of single-fidelity GP surrogate of an objective function f, and the resulting SF-EGO strategy. Section 2.2 introduces the extension of the GP surrogate construction to the multi-fidelity case, while our new non-nested infilling strategy for MF-EGO is detailed in Section 3.

2.1. Single-fidelity EGO

2.1.1. SF-*surrogate construction*

Consider a set of n training points $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$, each in Ω. The training points are associated to the vector $\boldsymbol{Y} = (y_1, \ldots, y_n)$ of noisy observations of the objective function f. It is assumed that $y_i = f(\boldsymbol{x}_i) + \varepsilon_i$, where the ε_i are independent and identically distributed Gaussian random variables having zero-mean and variance σ_ϵ^2, *i.e*, $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$.

The GP construction considers that $f(\boldsymbol{x})$ is a realization of a zero-mean multivariate Gaussian process with covariance function C. Without loss of generality, we consider here the multidimensional squared exponential covariance functions defined by

$$C(\boldsymbol{x}, \boldsymbol{x}'; \Theta) \doteq \theta_1 \prod_{i=1}^{d} \exp\left(\frac{-(x_i - x_i')^2}{2l_i^2}\right) + \theta_2, \tag{2}$$

where $\Theta = \{\theta_1, \theta_2, l_1, l_2, \ldots, l_d\}$ is the vector of covariance hyper-parameters to be inferred from the observations. Denoting $\boldsymbol{C} \in \mathbb{R}^{n\times n}$ the symmetric covariance matrix of the training points, with entries $C_{i,j} \doteq C\big(\boldsymbol{x}_i, \boldsymbol{x}_j; \Theta\big)$, for $1 \le i, j \le n$, the joint Gaussian distribution of the noisy observations $\boldsymbol{Y}$ and prediction of a new observation $Y(\boldsymbol{x})$ at $\boldsymbol{x} \notin \mathcal{X}$ is given by

$$\begin{pmatrix} \boldsymbol{Y} \\ Y(\boldsymbol{x}) \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{C} + \sigma_\epsilon^2 \boldsymbol{I}_n & \boldsymbol{k}(\boldsymbol{x}) \\ \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}) & \kappa(\boldsymbol{x}) + \sigma_\epsilon^2 \end{bmatrix}\right). \tag{3}$$

In (3) we have denoted $\kappa(\boldsymbol{x}) \doteq C(\boldsymbol{x}, \boldsymbol{x}; \Theta)$, $\boldsymbol{k}(\boldsymbol{x}) \doteq (C(\boldsymbol{x}, \boldsymbol{x}_1; \Theta) \cdots C(\boldsymbol{x}, \boldsymbol{x}_n; \Theta))^{\mathrm{T}}$ and $\boldsymbol{I}_n$ the identity matrix of $\mathbb{R}^n$. From the conditional rules of joint Gaussian distributions [24], the *best* prediction $\widehat{f}(\boldsymbol{x})$ of $f(\boldsymbol{x})$, *i.e.* the mean of $Y(\boldsymbol{x})$, and the prediction variance $\widehat{\sigma}^2(\boldsymbol{x})$ are given by

$$\widehat{f}(\boldsymbol{x}) = \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}) \left(\boldsymbol{C} + \sigma_\epsilon^2 \boldsymbol{I}_n\right)^{-1} \boldsymbol{Y}, \tag{4}$$

$$\widehat{\sigma}^2(\boldsymbol{x}) = \kappa(\boldsymbol{x}) + \sigma_\epsilon^2 - \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}) \left(\boldsymbol{C} + \sigma_\epsilon^2 \boldsymbol{I}_n\right)^{-1} \boldsymbol{k}(\boldsymbol{x}). \tag{5}$$

In the following, we shall denote $\boldsymbol{K} \doteq \left(\boldsymbol{C} + \sigma_\epsilon^2 \boldsymbol{I}_n\right)$ the matrix of the GP model. This matrix depends only on the set $\mathcal{X}$ of training points and the hyper-parameters of the covariance. The inverse of $\boldsymbol{K}$ appears in the expressions of the prediction expectation and variance. As predictions may have to be evaluated at multiple points $\boldsymbol{x}$, the inverse of $\boldsymbol{K}$ or its factorization is usually computed and stored for subsequent use. The hyper-parameters Θ of the GP covariance and the noise variance σ_ϵ^2 can be determined by maximizing the log-marginal likelihood of the GP model [24]; in the present work, we used an evolution strategy algorithm [10] to determine the optimal hyperparameters. More details on GP models can be found in [24].

2.1.2. SF-*infilling strategy*

Let $\widehat{\boldsymbol{x}}_{\text{opt}}$ be the optimum of $\widehat{f}(\boldsymbol{x})$. It is expected that $\widehat{\boldsymbol{x}}_{\text{opt}} \approx \boldsymbol{x}_{\text{opt}}$ if the approximation error $\widehat{f} - f$ is small enough, in particular in the neighborhood of $\boldsymbol{x}_{\text{opt}}$. The advantage of minimizing $\widehat{f}$ instead of f is that GP models are usually inexpensive to evaluate compared to the original objective function. In practice, a sequential enrichment of the training set, with new training points $\boldsymbol{x}_i \in \Omega$ selected in areas of interest where the optimum is likely to be found, reduces the approximation error. A deterministic optimization procedure would choose the next training point $\boldsymbol{x}_{n+1}$ as the optimal point of $\widehat{f}$. However, the GP model provides probabilistic information that can be exploited to propose more robust strategies based on merit functions combining the prediction $\widehat{f}(\boldsymbol{x})$ and its variance $\widehat{\sigma}^2(\boldsymbol{x})$. In this work, we use the merit functions based on the Augmented Expected Improvement (AEI) [11]. The AEI estimates the expected progress in the objective value at a new point, given the prediction variance $\widehat{\sigma}^2$ and observations noise in $\boldsymbol{Y}$. The AEI is written as

$$\text{AEI}(\boldsymbol{x}) = \text{EI}(\boldsymbol{x}) \left(1 - \frac{\sigma_\epsilon}{\sqrt{\widehat{\sigma}^2(\boldsymbol{x}) + \sigma_\epsilon^2}} \right), \tag{6}$$

where the Expected Improvement $\text{EI}(\boldsymbol{x})$ is defined by

$$\text{EI}(\boldsymbol{x}) \doteq \widehat{\sigma}(\boldsymbol{x}) \left[u(\boldsymbol{x}) \Phi\left(u(\boldsymbol{x})\right) + \phi\left(u(\boldsymbol{x})\right) \right], \quad u(\boldsymbol{x}) \doteq \frac{\widehat{f}(\boldsymbol{x}_{\text{best}}) - \widehat{f}(\boldsymbol{x})}{\widehat{\sigma}(\boldsymbol{x})}, \tag{7}$$

with Φ and ϕ are the cumulative and density functions of the standard Gaussian distribution, and $\boldsymbol{x}_{\text{best}} \in \mathcal{X}$ is the current effective best solution (see [11]). The optimum point of the AEI is added to the training set $\mathcal{X}$, and f is evaluated at this new point to complement the observation vector $\boldsymbol{Y}$. A new iteration can then start updating the GP model with the new training point. We remark that when the observation noise is zero, $\sigma_\epsilon^2 \to 0$, the AEI reduces to the EI while, for finite observation noise, the multiplicative factor tends to penalize areas of Ω where the prediction variance is comparatively small ($\widehat{\sigma}^2(\boldsymbol{x}) \ll \sigma_\epsilon^2$). Overall, each iteration requires one computation of the objective function f and the resolution of two optimization problems: a first one for the hyper-parameters of the GP model, and a second one to find the AEI optimum. These iterations of the GP-based optimization continue until a stopping criterion is satisfied or the resources allocated to the optimization problem have been exhausted.

2.2. Multi-fidelity surrogates

We consider the availability of a sequence of L models $f^{(1)}, \ldots, f^{(\text{L})}$ of increasing fidelity for the evaluation of the objective function, such that

$$f^{(l)}(\boldsymbol{x}) \xrightarrow{l \to \text{L}} f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \Omega.$$

Typically, the computational cost associated to the evaluation of $f^{(l)}(\boldsymbol{x})$ increases with the fidelity level l. The objective of the MF-surrogates is to incorporate observations corresponding to fidelity levels $l < \text{L}$ (thus associated with a lower computational cost) to improve the predictive capability of the surrogate model of $f^{(\text{L})}(\boldsymbol{x})$.

In this work, we use the recursive formulation proposed in [18], which reduces the complexity of the original MF approach proposed in [15]. The recursive formulation breaks down the determination of the MF-surrogate in successive GP process constructions for each intermediate level. The construction at a level involves the observations of that level only, contrary to the direct approach of [15], where all observations at all levels are treated at once, leading to the inversions of much larger covariance matrices. The recursive approach also drastically eases the determination

of the hyper-parameters associated with the GP models, by breaking down the global optimization in a sub-problem per level. Also, the extension to non-nested sets of training points of the recursive approach is immediate and has the advantage of providing explicit expressions for the predictive variances of the MF-surrogates at every intermediate level. We exploit the latter advantage in the following Section 3 to derive a novel non-nested infilling strategy.

We first consider the case of nested training sets. For $l = 1, \dots, \mathrm{L}$, let us denote $\mathcal{X}^{(l)}$ the set of training points for model $f^{(l)}$, such that

$$\mathcal{X}^{(l+1)} \subseteq \mathcal{X}^{(l)}, \quad l = 1, \dots, \mathrm{L} - 1. \tag{8}$$

As a result, the number of observations of a model decreases with its fidelity level. We shall denote $n_{(l)} = |\mathcal{X}^{(l)}|$ the size of the training set. Let $\boldsymbol{Y}^{(l)}$ be the associated vector of noisy observations of $f^{(l)}$ at the corresponding training points in $\mathcal{X}^{(l)}$, and $\sigma^2_{\epsilon_{(l)}}$ the variance of the observation noise on $f^{(l)}$.

2.2.1. MF *recursive construction*

The key idea of the recursive construction [18] is to define the MF-surrogate at level l, denoted $Y^{(l)}$, as a correction of the MF-surrogate at level $l-1$. Specifically, we write for $l = 1, \dots, \mathrm{L}$

$$Y^{(l)}(\boldsymbol{x}) \doteq \rho_{(l-1)} Y^{(l-1)}(\boldsymbol{x}) + \delta Y^{(l)}(\boldsymbol{x}), \quad Y^{(0)}(\boldsymbol{x}) \doteq 0, \tag{9}$$

where $\rho_{(l-1)} \in \mathbb{R}$ accounts for the correlation between the two successive levels, and $\delta Y^{(l)}(\boldsymbol{x})$ is the Gaussian correction process orthogonal to $Y^{(l-1)}(\boldsymbol{x})$. This particular form ensures a Markovian property between levels [15]. In particular, it can be shown that the covariance between $Y^{(l)}(\boldsymbol{x})$ and $Y^{(l-1)}(\boldsymbol{x}')$, conditioned on $Y^{(l-1)}(\boldsymbol{x})$, vanishes for all $\boldsymbol{x}, \boldsymbol{x}' \in \Omega^2$.

The MF-surrogate $Y^{(l)}$ is then a GP that dependents on the previous level MF-surrogate and on the noisy observations of $f^{(l)}(\boldsymbol{x})$, that is

$$\left[Y^{(l)}(\boldsymbol{x}) | Y^{(l-1)}(\boldsymbol{x}), \boldsymbol{Y}^{(l)} \right] \sim \mathcal{N}\left(\widehat{f^{(l)}}(\boldsymbol{x}), \widehat{\sigma}^2_{(l)}(\boldsymbol{x}) \right), \tag{10}$$

with mean and variance respectively given by

$$\widehat{f^{(l)}}(\boldsymbol{x}) = \rho_{(l-1)} \widehat{f^{(l-1)}}(\boldsymbol{x}) + \boldsymbol{k}^{\mathrm{T}}_{(l)}(\boldsymbol{x}) \left(\boldsymbol{C}_{(l)} + \sigma^2_{\epsilon_{(l)}} \boldsymbol{I}_{n_{(l)}} \right)^{-1} \Delta \boldsymbol{Y}^{(l)}(\rho_{(l-1)}), \tag{11}$$

$$\widehat{\sigma}^2_{(l)}(\boldsymbol{x}) = \rho^2_{(l-1)} \widehat{\sigma}^2_{(l-1)}(\boldsymbol{x}) + \kappa_{(l)}(\boldsymbol{x}) + \sigma^2_{\epsilon_{(l)}} - \boldsymbol{k}^{\mathrm{T}}_{(l)}(\boldsymbol{x}) \left(\boldsymbol{C}_{(l)} + \sigma^2_{\epsilon_{(l)}} \boldsymbol{I}_{n_{(l)}} \right)^{-1} \boldsymbol{k}_{(l)}(\boldsymbol{x}). \tag{12}$$

In the previous expressions, we have denoted $\Delta \boldsymbol{Y}^{(l)}$ the vector of noisy observation residuals at training points in $\mathcal{X}^{(l)}$ and defined as

$$\Delta \boldsymbol{Y}^{(l)}(\rho_{(l-1)}) = \begin{cases} \boldsymbol{Y}^{(l)}, & l = 1, \\ \boldsymbol{Y}^{(l)} - \rho_{(l-1)} \boldsymbol{Y}^{(l-1)}_{\mapsto(l)}, & l \geq 1, \end{cases} \tag{13}$$

where $\boldsymbol{Y}^{(l-1)}_{\mapsto(l)}$ is the sub-vector of $\boldsymbol{Y}^{(l-1)}$ corresponding to the elements associated with training points also belonging to $\mathcal{X}^{(l)}$. Therefore, $\Delta \boldsymbol{Y}^{(l)}$ is the vector of noisy observation residuals (relative to the previous level observations) at the training points in $\mathcal{X}^{(l)}$. We also reused in (11) and (12) the notations of the SF case, adding subscripts (l) to stress the dependencies on the considered level of $\boldsymbol{k}_{(l)}(\boldsymbol{x})$, $\boldsymbol{C}_{(l)}$, $\kappa_{(l)}(\boldsymbol{x})$ and $\sigma_{\epsilon_{(l)}}$. These dependencies arise from the variable size of the training sets, which decreases with the level l, but also because of the hyper-parameters $\Theta_{(l)}$, which are allowed to change from a level to another, as discussed below.

Observing the expression of the mean in (11), one recognizes the contribution from the mean prediction at the previous fidelity level scaled by the correlation coefficient $\rho_{(l-1)}$, while the second contribution is due to the observation residual $\Delta \boldsymbol{Y}^{(l)}$. Similarly, the expression of the variance of the MF-surrogate prediction, in (12), makes evident the contribution of the variance at the previous level (scaled by $\rho^2_{(l-1)}$) which, consistently with the Markovian assumption, directly adds up with the predictive variance of the correction $\delta Y^{(l)}(\boldsymbol{x})$. In fact, an expression of the two-point covariance of the MF-surrogate at level l can be derived in terms of the lower level covariances. It comes [15]

$$\mathrm{cov}\left(Y^{(l)}(\boldsymbol{x}), Y^{(l)}(\boldsymbol{x}') \right) = \sum_{j=1}^{l} \left(\prod_{i=j}^{l-1} \rho^2_{(i)} \right) C\left(\boldsymbol{x}, \boldsymbol{x}'; \Theta_{(j)} \right). \tag{14}$$

As mentioned before, one advantage of the recursive approach is that the hyper-parameters $\Theta_{(l)}$, noise variances $\sigma^2_{\epsilon_{(l)}}$ and regression coefficients $\rho_{(l-1)}$, must not be determined all together for all fidelity levels. Instead, owing to the Markovian structure of the construction, they are sequentially determined, from $l = 1$ to $l = \mathrm{L}$. As in the SF case, the hyper-parameters of level l are selected by maximizing the log-marginal likelihood of $Y^{(l)}$ [24],

$$\mathcal{L}\left(\rho_{(l-1)}, \Theta_{(l)}, \sigma^2_{\epsilon_{(l)}}\right) = -\frac{n_{(l)}}{2}\log(2\pi) - \frac{1}{2}\log\left|\boldsymbol{C}_{(l)}(\Theta_{(l)}) + \sigma^2_{\epsilon_{(l)}}\boldsymbol{I}_{n_{(l)}}\right| \tag{15}$$

$$- \frac{1}{2}(\Delta\boldsymbol{Y}^{(l)}(\rho_{(l-1)}))^{\mathrm{T}}\left(\boldsymbol{C}_{(l)}(\Theta_{(l)}) + \sigma^2_{\epsilon_{(l)}}\boldsymbol{I}_{n_{(l)}}\right)^{-1}\Delta\boldsymbol{Y}^{(l)}(\rho_{(l-1)}). \tag{16}$$

2.2.2. Elementary example

As an elementary illustrative example, taken from [8], consider the case of two fidelity-levels with functions $f^{(1)}$ and $f^{(2)}$ defined as

$$f^{(1)}(x) = \frac{1}{2}(6x-2)^2\sin(12x-4) + 10(x-1), \quad f^{(2)}(x) = 2f^{(1)}(x) - 20(x-1). \tag{17}$$

The two models are reported with the solid blue ($f^{(1)}$) and red ($f^{(2)}$) curves shown in Fig. 1. The low fidelity model $f^{(1)}$ is evaluated at 11 points (blue circles), while the high fidelity model $f^{(2)}$ is only evaluated at a subset of 4 points (red squares). In Fig. 1, the dashed green curve is the SF-surrogate based only on the 4 observations of $f^{(2)}$. The SF-surrogate is seen to badly approximate $f^{(2)}$: the location and value of the SF-surrogate minimum are far from those of $f^{(2)}$. In contrast, the MF-surrogate constructed as described below, and shown with a dashed black line, provides a much better estimation of $f^{(2)}$ with improved minimum value and location.

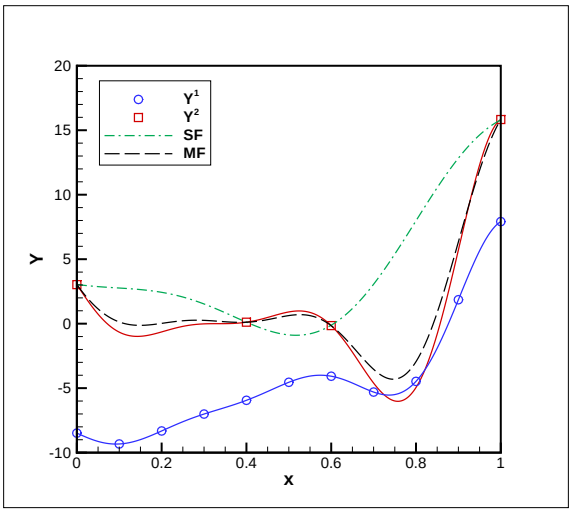


Figure 1: One-variable multi-fidelity example. The MF-surrogate uses both low ($\boldsymbol{Y}^{(1)}$) and high ($\boldsymbol{Y}^{(2)}$) fidelity observations.

3. Non-Nested MF-infilling strategy

3.1. Non-Nested MF*-surrogate construction*

As discussed in the introduction, we prefer to consider non-nested sets of points for the MF-EGO, in order to reduce the computational cost and time associated with the resolution of lower fidelity models.

The first step is then to extend the previous MF-surrogate construction to non-nested training sets. This turns to be quite natural within the recursive formulation of the MF-surrogate, simply substituting in the definition of the

observation residual (13) the previous level observations $\boldsymbol{Y}^{(l-1)}_{\mapsto(l)}$ with the corresponding vector of surrogate predictions $\widehat{f}_{(l-1)}(\boldsymbol{x} \in \mathcal{X}^{(l)})$. This change affects the (16) and (11), for which we now have

$$\Delta\boldsymbol{Y}^{(l)}(\rho_{(l-1)}) = \begin{cases} \boldsymbol{Y}^{(l)}, & l = 1, \\ \boldsymbol{Y}^{(l)} - \rho_{(l-1)}\widehat{f}_{(l-1)}(\boldsymbol{x} \in \mathcal{X}^{(l)}), & l \geq 1. \end{cases} \tag{18}$$

3.2. Non-Nested MF merit function

We now present our infilling method for MF-EGO that uses non-nested multi-fidelity surrogates. We start from the Augmented Expected Improvement (AEI) [11] merit function discussed in the SF-EGO section. We recall that this merit function, in (6), consists of the product of two terms: the expected improvement EI and a penalization term favoring areas with substantial prediction uncertainty.

A straight-forward MF-extension of (6) consists in defining the EI on the highest fidelity surrogate, *i.e.* using $\widehat{f}^{(\mathrm{L})}(\boldsymbol{x})$, to obtain

$$\mathrm{EI}(\boldsymbol{x}) = \widehat{\sigma}_{(\mathrm{L})}(\boldsymbol{x})\left[u(\boldsymbol{x})\Phi\left(u(\boldsymbol{x})\right) + \phi\left(u(\boldsymbol{x})\right)\right], \quad u(\boldsymbol{x}) = \frac{\widehat{f}^{(\mathrm{L})}(\boldsymbol{x}_{\mathrm{best}}) - \widehat{f}^{(\mathrm{L})}(\boldsymbol{x})}{\widehat{\sigma}_{(\mathrm{L})}(\boldsymbol{x})}, \tag{19}$$

with, as before, Φ and ϕ the cumulative and density functions of the standard Gaussian distribution, and where $\boldsymbol{x}_{\mathrm{best}} \in \cup_l \mathcal{X}^{(l)}$ the current effective best solution of all the sets of training points. Concerning the augmentation factor, one can compare the variance of the MF-prediction variance with the inferred noise in the observation residual at the highest-fidelity level ($\Delta Y^{(\mathrm{L})}$). This would lead to a first expression of merit function:

$$\mathrm{AEI}_{(\mathrm{L})}(\boldsymbol{x}) = \mathrm{EI}(\boldsymbol{x})\left(1 - \frac{\sigma_{\epsilon_{(\mathrm{L})}}}{\sqrt{\widehat{\sigma}^2_{(\mathrm{L})} + \sigma^2_{\epsilon_{(\mathrm{L})}}}}\right). \tag{20}$$

Contrary to the SF case, we have to select a new point $\boldsymbol{x}_*$ and the fidelity level l_* of the model to be evaluated. The SF case would correspond to selecting systematically the highest level (L).

A possible point of view for selecting the fidelity level, whence the new point $\boldsymbol{x}_*$ is decided, consists in answering the question: what is the effect of updating the MF-surrogate with a new observation of $f^{(l)}$ at $\boldsymbol{x}_*$, with $l < \mathrm{L}$, compared to an evaluation of the highest fidelity model $f^{(\mathrm{L})}$? To address this question, we have to specify the measure of this effect. In the present work, we propose to measure the impact of the new observation $\boldsymbol{x}_*$, and its associated fidelity level l_*, through the induced reduction of the MF-surrogate predictive variance. To this end, we denote $\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}_*, l_*)$ the predictive variance at the highest resolution level, given a new observation $f^{(l_*)}(\boldsymbol{x}_*)$. The interest of using the predictive variance as an criterion for the fidelity level selection is that the evaluation of $\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}_*, l_*)$ does not need to evaluate $f^{(l_*)}(\boldsymbol{x}_*)$, provided that the hyper-parameters and regression parameters of the MF-GPs are not too affected by the new observation. In this case, we can exploit the recursive expression of the predictive variance in (12). In particular, the predictive variances $\widehat{\sigma}^2_{(l)}$ of models $f^{(l)}$ with $l < l_*$ are not affected by a new observation of $f^{(l_*)}$.

We first express the change in the predictive variance of $f^{(l_*)}$ induced by the new observation. From (12) we see that $\widehat{\sigma}^2_{(l_*)}(\boldsymbol{x})$ has 3 contributions: a first contribution, $\rho_{(l_*-1)}\widehat{\sigma}^2_{(l_*-1)}(\boldsymbol{x})$, from the predictive variance at the previous level; a second contribution, $\kappa_{(l_*)}(\boldsymbol{x}) + \sigma^2_{\epsilon_{(l_*)}}$, independent of the observation points in $\mathcal{X}^{(l_*)}$; and finally a third contribution $-\boldsymbol{k}^{\mathrm{T}}_{(l)}(\boldsymbol{x})\left(\boldsymbol{C}_{(l)} + \sigma^2_{\epsilon_{(l)}}\boldsymbol{I}_{n_{(l)}}\right)^{-1}\boldsymbol{k}_{(l)}(\boldsymbol{x})$ which depends on the observation points in $\mathcal{X}^{(l_*)}$. Only the last contribution is affected by adding a new observation, and the updated prediction variance of $f^{(l_*)}(\boldsymbol{x})$ can be expressed as

$$\widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}|\boldsymbol{x}_*, l_*) = \widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}) + \Delta\widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}|\boldsymbol{x}_*, l_*), \tag{21}$$

where $\Delta\widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}|\boldsymbol{x}_*, l_*)$ is the predictive variance update. Reusing the previous notations the predictive variance update can be written as

$$\Delta\widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}; \boldsymbol{x}_*, l_*) = \begin{bmatrix} \boldsymbol{k}^{\mathrm{T}}_{(l)}(\boldsymbol{x}) & C_{(l)}(\boldsymbol{x}, \boldsymbol{x}_*) \end{bmatrix} \begin{bmatrix} \boldsymbol{K}_{(l)} & \boldsymbol{k}_{(l)}(\boldsymbol{x}_*) \\ \boldsymbol{k}^{\mathrm{T}}_{(l)}(\boldsymbol{x}_*) & \kappa_{(l)}(\boldsymbol{x}_*) + \sigma^2_{\epsilon_{(l)}} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{k}_{(l)}(\boldsymbol{x}) \\ C_{(l)}(\boldsymbol{x}, \boldsymbol{x}_*) \end{bmatrix} - \boldsymbol{k}^{\mathrm{T}}_{(l)}(\boldsymbol{x})\boldsymbol{K}^{-1}_{(l)}\boldsymbol{k}_{(l)}(\boldsymbol{x}), \tag{22}$$

where, consistently with the notations of Section 2.1.1, we have denoted the GP model matrix

$$\boldsymbol{K}_{(l)} = \left[\boldsymbol{C}_{(l)} + \sigma^2_{\epsilon_{(l)}} \boldsymbol{I}_{n_{(l)}}\right].$$

Since for fixed hyper-parameters a new observation can only reduce the predictive variance of the GP, it holds for all $\boldsymbol{x}$ and $\boldsymbol{x}_*$:

$$0 \leq -\Delta\widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}|\boldsymbol{x}_*, l_*) \leq \widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}).$$

Returning to the recursive expression for the predictive variance, we find that the updated MF predictive variance of $f^{(\mathrm{L})}(\boldsymbol{x})$, for the new observation $f^{(l_*)}(\boldsymbol{x}_*)$, is

$$\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}_*, l_*) = \widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}) + R^2_{l_*}\Delta\widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}|\boldsymbol{x}_*, l_*), \quad R^2_{l_*} = \begin{cases} \prod_{l=l_*}^{l<\mathrm{L}} \rho^2_{(l)}, & l_* < \mathrm{L}, \\ 1, & l_* = \mathrm{L}. \end{cases}$$

To introduce a criteria on the fidelity level in the merit function, we can augment (20) to favor couples $(\boldsymbol{x}_*, l_*)$ achieving significant reduction of the predictive variance. The reduction can be measured globally, integrating $R^2_{l_*}\Delta\widehat{\sigma}^2_{(l_*)}(\boldsymbol{x}|\boldsymbol{x}_*, l_*)$ for $\boldsymbol{x}$ in the whole design space Ω, for instance using a Monte-Carlo quadrature. To avoid this computational overhead, we restrict ourselves to a simpler reduction criteria and compare the prior and updated predictive variances locally at $\boldsymbol{x}_*$, where the reduction is the largest. Specifically, we consider the variance ratio $\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}, l)/\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x})$ to augment $\mathrm{AEI}_{(\mathrm{L})}$ into the MF-merit function

$$\mathrm{M}^{\mathrm{MF}}(\boldsymbol{x}, l) \propto \mathrm{AEI}_{(\mathrm{L})}(\boldsymbol{x}) \ \max\left(0, \left(1 - \frac{\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}, l)}{\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x})}\right)\right) = \mathrm{AEI}_{(\mathrm{L})}(\boldsymbol{x}) \max\left(0, \left(-R^2_l \frac{\Delta\widehat{\sigma}^2_{(l)}(\boldsymbol{x}|\boldsymbol{x}, l)}{\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x})}\right)\right). \tag{23}$$

Note that, in principle, $0 \leq \widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}, l) \leq \widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x})$ such that the max operator is theoretically not necessary. In practice, however, we found it necessary to ensure non negative merit value. In the limit $\sigma_\epsilon \to 0$ we have $\Delta\widehat{\sigma}^2_{(l)}(\boldsymbol{x}|\boldsymbol{x}, l) \to -\widehat{\sigma}^2_{(l)}(\boldsymbol{x})$, since the GP processes become interpolatory. Thus, in this limit, the augmented term favors models whose variances yield the largest predictive variance through the recursive rule. Also, models of fidelity level $l < \mathrm{L}$ having large variance are only favored if they correlate well with $f^{(\mathrm{L})}$, *i.e.*, if R^2_l is significant.

The merit function above will generally favor an observation at the highest fidelity level L, for which $R^2_L = 1$, as it is usually the most effective way to reduce the predictive variance. The interest in using lower-fidelity models emerges when the merit function accounts for the computational costs. To this end, let us denote $\mathrm{W}_{(l)} \in \mathbb{R}_+$ a representative computational cost for a single evaluation of $f^{(l)}(\boldsymbol{x})$. We shall assume that the actual computational cost is not departing much from $\mathrm{W}_{(l)}$ for all $\boldsymbol{x} \in \Omega$. These costs can be used to weight the expected reduction of the predictive variance, finally yielding the proposed (Non-Nested) MF-merit function

$$\mathrm{M}^{\mathrm{NN-MF}}(\boldsymbol{x}, l) \doteq \mathrm{AEI}_{(\mathrm{L})}(\boldsymbol{x}) \ \frac{\mathrm{W}_{(\mathrm{L})}}{\mathrm{W}_{(l)}} \ \max\left(0, \left(1 - \frac{\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}, l)}{\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x})}\right)\right). \tag{24}$$

3.3. Nested case

We extend the merit function proposed above to the case of nested sets of training points. In this situation, for a new point $\boldsymbol{x}_*$ with fidelity level l_*, the corresponding reduction of the predictive variance becomes:

$$\Delta_{\mathrm{N}}\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}_*, l_*) = \sum_{l=1}^{l_*} R^2_l \Delta\widehat{\sigma}^2_{(l)}(\boldsymbol{x}|\boldsymbol{x}_*, l).$$

In words, the reduction for the nested approach is given by the sum of all reductions induced by levels $l \leq l_*$. Regarding the cost of the nested case, the new evaluations total cost is $\sum_{l \leq l_*} \mathrm{W}_{(l)}$, which have to be compared to the cost of selecting systematically the highest fidelity level $l_* = \mathrm{L}$, namely $\sum_{l \leq \mathrm{L}} \mathrm{W}_{(l)}$. The nested MF-merit function thus expresses as

$$\mathrm{M}^{\mathrm{N-MF}}(\boldsymbol{x}, l) \doteq \mathrm{AEI}_{(\mathrm{L})}(\boldsymbol{x}) \ \frac{\sum_{l' \leq \mathrm{L}} \mathrm{W}_{(l')}}{\sum_{l' \leq l} \mathrm{W}_{(l')}} \ \max\left(0, \left(-\frac{\sum_{l'=1}^{l} R^2_{l'} \Delta\widehat{\sigma}^2_{(l')}(\boldsymbol{x}|\boldsymbol{x}, l')}{\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x})}\right)\right). \tag{25}$$

An alternative expression for the Nested MF-EGO merit function is

$$\mathrm{M}^{\mathrm{N-MF}}(\boldsymbol{x}, l) = \mathrm{AEI}_{(\mathrm{L})}(\boldsymbol{x}) \, \frac{\sum_{l' \leq \mathrm{L}} \mathrm{W}_{(l')}}{\sum_{l' \leq l} \mathrm{W}_{(l')}} \, \max\left(0, \left(1 - \frac{\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}, l)}{\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x})}\right)\right), \tag{26}$$

where the updated variance $\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}, l)$ is given in the nested case by

$$\widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}, l) = \widehat{\sigma}^2_{(\mathrm{L})}(\boldsymbol{x}) + \sum_{l' \leq l} R^2_{(l')} \Delta\widehat{\sigma}^2_{(l')}(\boldsymbol{x}|\boldsymbol{x}, l'). \tag{27}$$

3.4. Computational details

We see that the MF merit functions call for the resolution of a mixed-type non-linear programming problem for finding the optimum of the functional in (24) and (25). In the present work, where L remains small, it is possible to go through over all possible l and optimize the corresponding new point $\boldsymbol{x}_*(l)$ to select the best couple. The non-nested approach presents the advantage of making feasible the parallel optimizations of $\boldsymbol{x}_*(l)$. A parallel search of $\boldsymbol{x}_*(l)$ is also feasible in the nested approach, but it presents the inconvenience to go through all levels $l' \leq l$ to compute $\mathrm{M}^{\mathrm{N-MF}}(\boldsymbol{x}, l)$. As a result, the determination of the new point and maximal fidelity level is computationally more demanding in the nested case. The computational complexity in the nested case is further impacted by the number of models to evaluate, for each new point, and the update of all the MF-surrogates. Algorithm 1 sketches the two MF-EGO methods, highlighting their main similarities and differences in the updating of the MF-surrogates after the enrichment of the training set(s). Finally, these MF-EGO iterations continue until a stopping criterion is satisfied. This stopping criterion can be based, for instance, on the value of the expected improvement or the exhaustion of a computational budget.

Algorithm 1: Multi-fidelity EGO algorithm.

Input: Initial training sets and observation vectors $\{\mathcal{X}^{(l)}, \boldsymbol{Y}^{(l)}, l = 1, \ldots \mathrm{L}\}$, computational costs $\mathrm{W}_{(1 \cdots, \mathrm{L})}$
Output: Updated training sets and observation vectors $\{\mathcal{X}^{(l)}, \boldsymbol{Y}^{(l)}, l = 1, \ldots \mathrm{L}\}$, MF-surrogate models

for $l \leftarrow 1$ **to** L **do**
  Find $\left(\rho_{(l-1)}, \Theta_{(l)}, \sigma^2_{\epsilon_{(l)}}\right)$ minimizing (16) ! Set initial MF-surrogate
  Factorize $\boldsymbol{K}_{(l)} = \left(\boldsymbol{C}_{(l)} + \sigma^2_{\epsilon_{(l)}} \boldsymbol{I}_{n_{(l)}}\right)$! Factorize GP matrix
while *Not converged* **do**
  if *Non-Nested Multifidelity* **then**
    Find $(\boldsymbol{x}_*, l_*)$ minimizing $\mathrm{M}^{\mathrm{NN-MF}}(\boldsymbol{x}, l)$ in (24) ! Find new point and level
    $\mathcal{X}^{(l_*)} \leftarrow \mathcal{X}^{(l_*)} \cup \{\boldsymbol{x}_*\}$, $\boldsymbol{Y}^{(l_*)} \leftarrow \boldsymbol{Y}^{(l_*)} \cup \{f^{(l_*)}(\boldsymbol{x}_*)\}$! Update the training set
    for $l \leftarrow l_*$ **to** L **do**
      Find $\left(\rho_{(l-1)}, \Theta_{(l)}, \sigma^2_{\epsilon_{(l)}}\right)$ minimizing (16) ! Update the MF-surrogate
      Factorize $\boldsymbol{K}_{(l)} = \left(\boldsymbol{C}_{(l)} + \sigma^2_{\epsilon_{(l)}} \boldsymbol{I}_{n_{(l)}}\right)$! Factorize the GP matrix
  else
    Find $(\boldsymbol{x}_*, l_*)$ minimizing $\mathrm{M}^{\mathrm{N-MF}}(\boldsymbol{x}, l)$ in (25)
    for $l \leftarrow 1$ **to** L **do**
      if $l \leq l_*$ **then**
        $\mathcal{X}^{(l)} \leftarrow \mathcal{X}^{(l)} \cup \{\boldsymbol{x}_*\}$, $\boldsymbol{Y}^{(l)} \leftarrow \boldsymbol{Y}^{(l)} \cup \{f^{(l)}(\boldsymbol{x}_*)\}$! Update the training sets
      Find $\left(\rho_{(l-1)}, \Theta_{(l)}, \sigma^2_{\epsilon_{(l)}}\right)$ minimizing (16) ! Update MF-surrogate of $f^{(l)}$
      Factorize $\boldsymbol{K}_{(l)} = \left(\boldsymbol{C}_{(l)} + \sigma^2_{\epsilon_{(l)}} \boldsymbol{I}_{n_{(l)}}\right)$! Factorize the GP matrix

As a final word on computational complexity, we observe that finding the optimum of the merit function calls for multiple estimation of the updated variance $\widehat{\sigma}_{(\mathrm{L})}(\boldsymbol{x}|\boldsymbol{x}, l)$, which in turn calls for the computation of variance reduction

terms, such as $\Delta\widehat{\sigma}_{(l)}(\boldsymbol{x}|\boldsymbol{x}, l)$, defined by (22). The computationally intensive part of this evaluation is the inversion of the GP matrices in (22). The second one, $\boldsymbol{K}_{(l)}$, is not depending on the new point $\boldsymbol{x}_*$ such that it can be factorized and inverted once for all during the search of the couple $(\boldsymbol{x}_*, l_*)$ maximizing the merit function. The first matrix, on the contrary, depends on the new points, and many inversions can be necessary during the search for $\boldsymbol{x}_*$, with a high computational burden as a result. The matrix inversion lemma can alleviate this computational burden. Indeed, dropping temporarily the fidelity level index l, we have

$$\begin{bmatrix} \boldsymbol{K} & \boldsymbol{k}(\boldsymbol{x}_*) \\ \boldsymbol{k}^{\mathrm{T}}(\boldsymbol{x}_*) & \kappa(\boldsymbol{x}_*) + \sigma_\epsilon^2 \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{K}^{-1} + r(\boldsymbol{x}_*)\boldsymbol{K}^{-1}\boldsymbol{k}(\boldsymbol{x}_*)\boldsymbol{k}^T(\boldsymbol{x}_*)\boldsymbol{K}^{-1} & -r(\boldsymbol{x}_*)\boldsymbol{K}^{-1}\boldsymbol{k}(\boldsymbol{x}_*) \\ -r(\boldsymbol{x}_*)\boldsymbol{k}^T(\boldsymbol{x}_*)\boldsymbol{K}^{-1} & r(\boldsymbol{x}_*) \end{bmatrix}, \tag{28}$$

where

$$r(\boldsymbol{x}_*) \doteq \left(\kappa(\boldsymbol{x}_*) + \sigma_\epsilon^2 - \boldsymbol{k}^T(\boldsymbol{x}_*)\boldsymbol{K}^{-1}\boldsymbol{k}(\boldsymbol{x}_*)\right)^{-1}. \tag{29}$$

With this expression, the evaluation of $\Delta\widehat{\sigma}_{(l)}(\boldsymbol{x}|\boldsymbol{x}, l)$ involves only the inverse of the GP model matrix $\boldsymbol{K}_{(l)}$, which is already assembled and inverted (or factorized) during the MF-surrogate construction (see Section 2.1.1). Using (28), therefore, leads to significant computational savings as it avoids the inversion of the extended system of size $n_{(l)} + 1$ for every new point proposed. We emphasize this aspect in Algorithm 1 by making clear the factorizations of the GP matrices.

3.5. Elementary example

Figure 2 illustrates the non-nested MF-EGO on the univariate case with functions $f^{(1)}$ and $f^{(2)}$ of (17). The initial training sets are the ones used in Section 2.2.2. Each plot shows a different iteration EGO iteration $\mathrm{N_{iter}}$ the MF-prediction $\widehat{f_2}(x)$ (black line) and the merit function $\mathrm{M^{NN-MF}}(x, l)$ for $l = 1$ (blue dashed line) and $l = 2$ (red dashed line), respectively. The assumed computing costs are $\mathrm{W}_{(1)} = 1$ and $\mathrm{W}_{(2)} = 10$. At the first NN-MF-EGO iteration ($\mathrm{N_{iter}} = 1$, Fig. 2(a)), the algorithm selects ($x_* \approx 0.6, l_* = 1$), which yield a clear improvement of the prediction $\widehat{f^{(2)}}(x)$ close to the global optimum (see Fig. 2(b)). After the MF-surrogates updates, the merit function drastically reduces for $l = 1$ and increases for $l = 2$ in the neighborhood of the next point, also it is still presenting a global maximum for $l = 1$. However, the procedure does select $l_* = 2$ at the subsequent iterations $\mathrm{N_{iter}} = 3$ and $\mathrm{N_{iter}} = 4$. After $\mathrm{N_{iter}} = 10$ iterations, the merit function $\mathrm{M^{NN-MF}}(x, l)$ is small and $f^{(2)}$ is well approximated by the MF-surrogate $\widehat{f_2}(x)$ at the optimum point (see black and red continuous curves in Fig. 2(f)).

4. Analytical test problems

The objective of this section is to illustrate the effectiveness of the proposed MF-EGO methods and demonstrate the interest of using non-nested training sets. To this end, we apply our non-nested method (NN-MF) on different variants of an analytical problem, and contrast its performances with optimization results obtained with the nested (N-MF) and single fidelity (SF) methods.

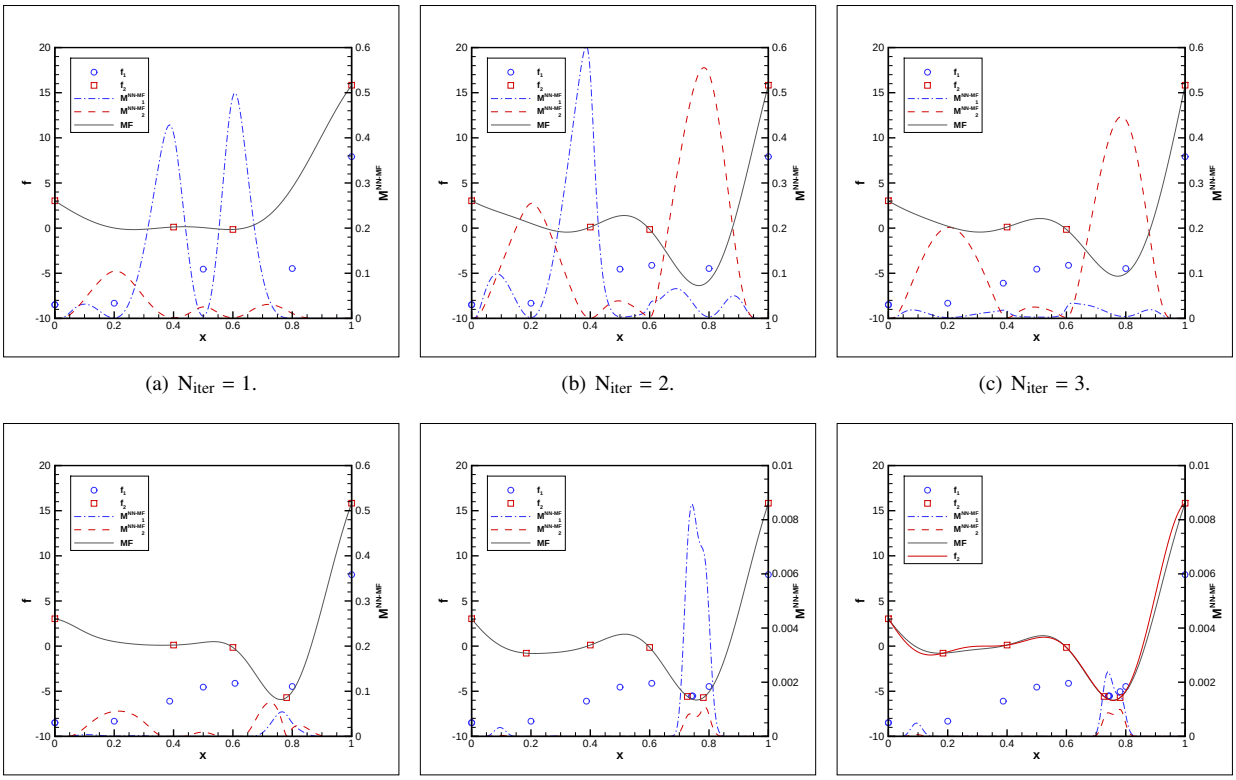
4.1. Problem settings

The analytical test problem is the minimization of the 6-dimensional Hartmann function [21], defined as

$$f(\boldsymbol{x}) = \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}\left(x_j - P_{ij}\right)^2\right), \tag{30}$$

for $\boldsymbol{x} \in \Omega = [0, 1]^6$, and where $\boldsymbol{\alpha} = (1, 1.2, 3, 3.2)^{\mathrm{T}}$ and the matrix $\boldsymbol{A}$ and $\boldsymbol{P}$ are

$$\boldsymbol{A} = \begin{pmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{pmatrix}, \quad \boldsymbol{P} = 10^{-4} \begin{pmatrix} 1312 & 1696 & 5569 & 124 & 8283 & 5886 \\ 2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\ 2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\ 4047 & 8828 & 8732 & 5743 & 1091 & 381 \end{pmatrix}.$$



(a) $N_{iter} = 1$. (b) $N_{iter} = 2$. (c) $N_{iter} = 3$.

(d) $N_{iter} = 4$. (e) $N_{iter} = 8$. (f) $N_{iter} = 10$.

Figure 2: Non-nested EGO iterations for the univariate case of Section 2.2.2. The blue and red symbols indicate the training sets of $f^{(1)}$ and $f^{(2)}$, respectively, while the dashed lines depicts $M^{NN-MF}(x, l)$ for $l = 1$ and 2. The solid black line is the MF-surrogate $\widehat{f^{(2)}}(x)$, while the solid red line in the last plot is the function $f^{(2)}(x)$.

The global optimum is $\boldsymbol{x}_{opt} = [0.20169, 0.150011, 0.476874, 0.275332, 0.311652, 0.6573]$.

We considered the following sequence to define lower fidelity versions of the Hartmann function,

$$U_{k+1}(\boldsymbol{x}) = \frac{1}{2}\left(\frac{f(\boldsymbol{x})^2}{U_k(\boldsymbol{x})} + U_k(\boldsymbol{x})\right), \quad U_0(\boldsymbol{x}) = u_0 \in \mathbb{R}. \tag{31}$$

In the following we initiate the sequence with $u_0 = -5$; the sequence has then $f(\boldsymbol{x})$ as a contractive fixed-point: $U_k \xrightarrow{k\to\infty} f(\boldsymbol{x})$. Using the sequence of U_k, we define the multi-fidelity approximations of $f(\boldsymbol{x})$ as

$$f^{(l)}(\boldsymbol{x}) \doteq U_{k_l}(\boldsymbol{x} + \delta\boldsymbol{x}_{k_l}), \quad l = 1, \ldots, L, \tag{32}$$

for some increasing sequence of indexes $k_l \in \mathbb{N}$, and where the translation vector $\delta\boldsymbol{x}_k$ is given by:

$$(\delta\boldsymbol{x}_k)_i = \delta/k, \quad i = 1, \ldots, 6.$$

The translation by $\delta\boldsymbol{x}_k$ prevents the functions $f^{(l)}(\boldsymbol{x})$ in (32) to have all the same minimum $\boldsymbol{x}_{opt}$; instead, the minimum of $f^{(l)}$ is translated to a distance $\sqrt{6}|\delta|/k_l$ and we shall use δ to control this distance. We remark, however, that the translation does not affect the convergence of U_k to f for finite $\delta \in \mathbb{R}$. Figure 3 illustrates the effects of increasing δ on the multi-fidelity approximations of $f(\boldsymbol{x})$. It is seen that increasing δ reduces the correlation between $U_k(\boldsymbol{x})$ and $U_{k+1}(\boldsymbol{x})$, while the correlation increases with k.

In all the section, we fixed $L = 3$ and use the multi-fidelity sequence $f^{(1,2,3)}$ defined by $k_1 = 1$, $k_2 = 3$ and $k_3 = k_L = \infty$; in other words $f^{(L)}(\boldsymbol{x}) = f(\boldsymbol{x})$. Regarding the initialization of the EGO procedures, all the tests use the

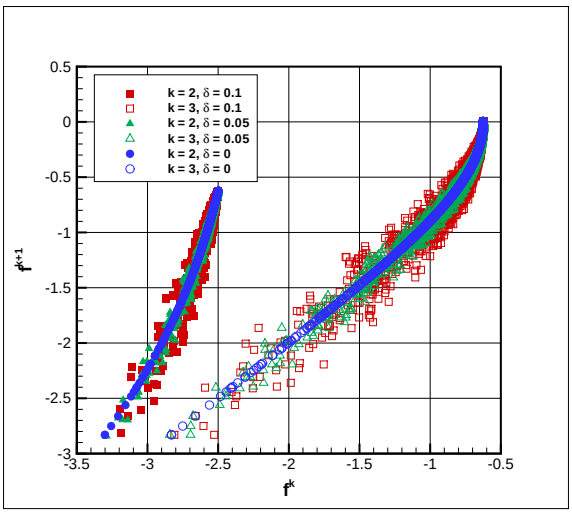


Figure 3: Random evaluations of 5000 samples of the multi-fidelity analytical test problem (32).

same initial training sets. For the MF-methods, the initial lower fidelity set $\mathcal{X}^{(1)}$ consists of $n_{(1)} = 20$ training points drawn randomly in Ω using Latin Hypercube Sampling (LHS). The next training sets $\mathcal{X}^{(2)}$ and $\mathcal{X}^{(3)}$ are obtained by selecting recursively, and at random, $n_{(2)} = 15$ and $n_{(3)} = 10$ elements of their respective predecessor sets ($\mathcal{X}^{(1)}$ and $\mathcal{X}^{(2)}$ respectively). The initial training sets are nested, even for the NN-MF-EGO. For the SF-EGO, the initial training set is selected as $\mathcal{X} = \mathcal{X}^{(1)}$. Finally, unless stated otherwise, we assume $\mathrm{W}_{(1)} = 1$, $\mathrm{W}_{(2)} = 100$ and $\mathrm{W}_{(\mathrm{L})} = 1,000$ for the computational costs of the different fidelity models, and set $\mathrm{N}_{\mathrm{iter}} = 400$ for the number of EGO-iterations in the all tests. Note that the total computational cost $\mathrm{W}_{\mathrm{tot}}$

$$\mathrm{W}_{\mathrm{tot}} = \sum_{l=1}^{\mathrm{L}} \mathrm{W}_{(l)} n_{(l)}, \tag{33}$$

increases with $\mathrm{N}_{\mathrm{iter}}$, but changes from a method to another depending on the level of the new training points.

The results reported in the following consist in the distance between the surrogate-based optimum and the true optimum $\|\boldsymbol{x}_{\mathrm{opt}} - \hat{\boldsymbol{x}}_{\mathrm{opt}}\|$, or the error on the optimal objective value $|f(\boldsymbol{x}_{\mathrm{opt}}) - \widehat{f^{(\mathrm{L})}}(\hat{\boldsymbol{x}}_{\mathrm{opt}})|$, as functions of the cost $\mathrm{W}_{\mathrm{tot}}$. Therefore, we directly compare the efficiencies of the different methods. We consider the the effects of the translation parameter δ in Section 4.2, the impact of observation noise in the middle-fidelity model $f^{(2)}$ in Section 4.3, and, finally, the costs per level in Section 4.4.

4.2. Impact of the shift

We start by solving the problem for $\delta = 0$ in the definition of the multi-fidelity functions. We report in Fig. 4 the convergence of the three EGO-methods: SF, N-MF, and NN-MF. We plot in the top row of Fig. 4 the distances to the optimum point and global minimum of f as functions of the computational cost $\mathrm{W}_{\mathrm{tot}}$. We see that the MF and SF methods have a fast decay of the errors initially (note the log-scale in the distances), although there is a slight shift in the SF method whose cost is higher. For the MF-methods, the fast decay continues until reaching the number of allowed EGO iterations. In contrast, the initial fast convergence stage of the SF-method is followed by a long plateau. After the plateau, the errors reduce again but at a much lower rate than initially. Eventually, all methods exhibit solutions with comparable accuracy, but the two MF-methods perform much better than the SF-method: the computational effort of the MF-methods is roughly 10 times less to reach a distance of 10^{-2} on the optimum. Also, the performances of the two MF-methods are very close; the N-MF-method appears to yield a slightly better estimate of the optimum for a slightly lower cost.

These differences in the convergence trends can be understood by inspecting the evolutions with $\mathrm{W}_{\mathrm{tot}}$ of the training set sizes $n_{(l)}$ of the MF-methods, reported in the bottom row Fig. 4. The plot for the N-MF-method indicates

that the first and second fidelity levels are more frequently selected. In contrast, NN-MF-method never selects and evaluates the first level until a cost of about 4 10^4, after which it becomes the most frequent level evaluated. As the paths of $n_{(2)}$ and $n_{(3)}$ are similar for the two MF-methods, one could expect a higher efficiency for the NN-MF-method, owing to saving evaluations of $f^{(1)}$, but in fact the distance and error for N-MF-method are slightly better at a given cost. This difference is explained by a slightly improved prediction of the nested multi-fidelity model. However, we underline that the two MF-methods have close efficiencies, and both do significantly better than pure SF-method.

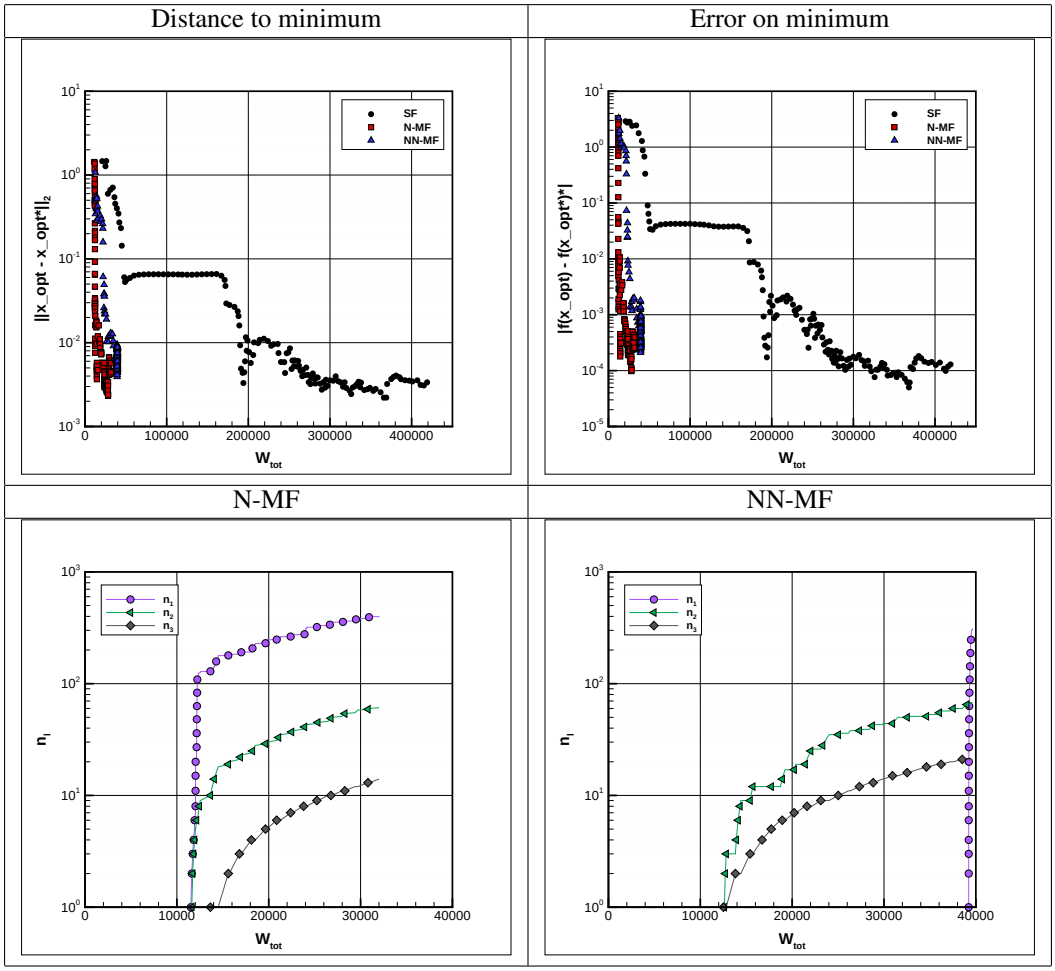


Figure 4: Distances to optimum and errors on optimum value as functions of the computational cost for the NN-MF, N-MF and SF-EGO methods (top line), and training sample sizes of the MF-methods (bottom line). Case of $\delta = 0$.

We now turn to the case where the definition of the multi-fidelity functions has a non-zero δ. Figure 5 reports the results for the previous situation but with parameter values $\delta = 0.05$ (left) and $\delta = 0.1$ (right). The plots depict the distances to $\boldsymbol{x}_{\text{opt}}$ and errors on f_{opt} for the different methods with $\delta = 0.05$ (left) and $\delta = 0.1$ (right). We see that for the two δ tested, the NN-MF-method (triangles) converges with a rate comparable to the case of $\delta = 0$ (see Fig. 4). This result proves the robustness of the NN-MF-method when the successive multi-fidelity functions are not much correlated, with important changes in the locations of their respective minimums. In contrast, the N-MF-method (squares) exhibits behavior that depends on δ. When $\delta = 0.05$, the convergence with the cost is similar to the NN-MF-method; but for $\delta = 0.1$, the errors on $\boldsymbol{x}_{\text{opt}}$ and f_{opt} are stagnating to quite high values, about two orders of magnitude higher than for the NN-MF-method at the same cost. A closer analysis of the N-MF results indicates that the optimization procedure converges to a local minimum of the Hartmann function, and the number of EGO iterations allowed in this example is not enough to learn that this minimum is not the global one.

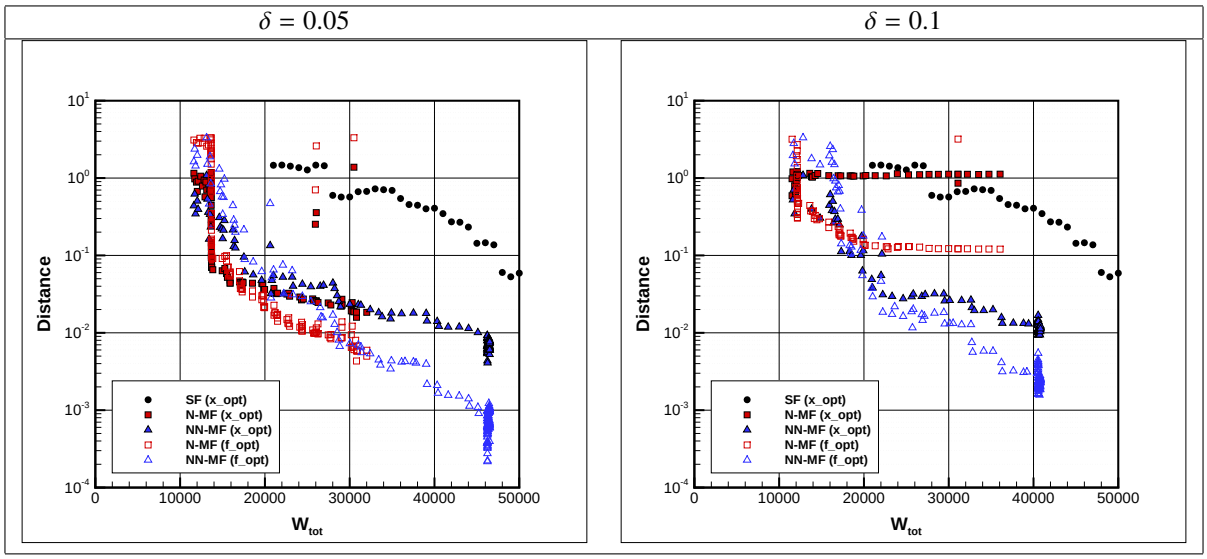


Figure 5: Distances to optimal solution as functions of the cost W_{tot} for the SF, N-MF and NN-MF methods. Cases of $\delta = 0.05$ (left) and $\delta = 0.1$ (right).

The evolutions of the training-sets size with the computational cost, depicted in Fig. 6, help to better understand the differences between the behavior of the two MF methods. It is seen that δ has not a significant impact on the evolutions of $n_{(l)}$ for the two MF-method: the N-MF-method uses more low-level evaluations in the construction of the MF-surrogate, while the NN-MF-method considers lower level evaluations only at the end of the EGO iterations when the global optimum is already identified. These differences explain the higher robustness to the shift of the NN-MF-method since one can expect an overall improvement of the MF-surrogate when the estimation involves more correlated functions. Still, the NN-MF-EGO performs much better than the SF-EGO, by substituting several evaluations of $f^{(\mathrm{L})}$ with $f^{(2)}$ without impacting much the predictive quality of the model.

4.3. Noisy estimations

In the next tests, we investigate the impact of evaluation noise in the intermediate multi-fidelity functions on the MF-EGO and show that the NN-MF-method presents an advantage over the N-MF-method in that case. This situation corresponds, for instance, to simulations using low-resolution meshes or very permissive convergence criteria in non-linear solvers, in order to obtain rough predictions at a low cost. One could also face this situation when using probabilistic predictors $f^{(l)}$. We then define the noisy version $f^{(2)}$ as

$$f_\epsilon^{(2)}(\boldsymbol{x}) = f^{(2)}(\boldsymbol{x})\left[1 + \eta(\boldsymbol{x})\right], \tag{34}$$

where $\eta(\boldsymbol{x})$ is a white noise random field with uniform distribution in $[0, 0.1]$.

Figure 7 shows the distances and errors to the optimum as functions of the computational cost when solving problem (4.3) with $\delta = 0$ (top-row) and $\delta = 0.1$ (bottom-row). The NN-MF-method clearly outperforms the other approaches, determining the optimum at a cost that increases with δ. It is essential to notice that the SF-method is not affected by the noise since it affects the fidelity level $l = 2 < \mathrm{L}$. In contrast, for the considered number of EGO-iterations, the N-MF-method is not improving over the SF-method, and it is seen to stagnate at a local minimum. Further, because of the nested property, which forces to compute the levels 1 and 2 when level 3 is selected, noisy evaluations from the level 2 are always added in the N-MF surrogate construction. It explains why the N-MF-EGO has difficulties in finding the true global optimum; it would need many more evaluations to reach the same prediction accuracy. In comparison the NN-MF-method evaluates $f^{(2)}$ infrequently compare to $f^{(\mathrm{L})}$. The number of evaluations can be appreciated from the evolutions of the training sets size reported in Fig, 8, which are very similar to the case without noise.

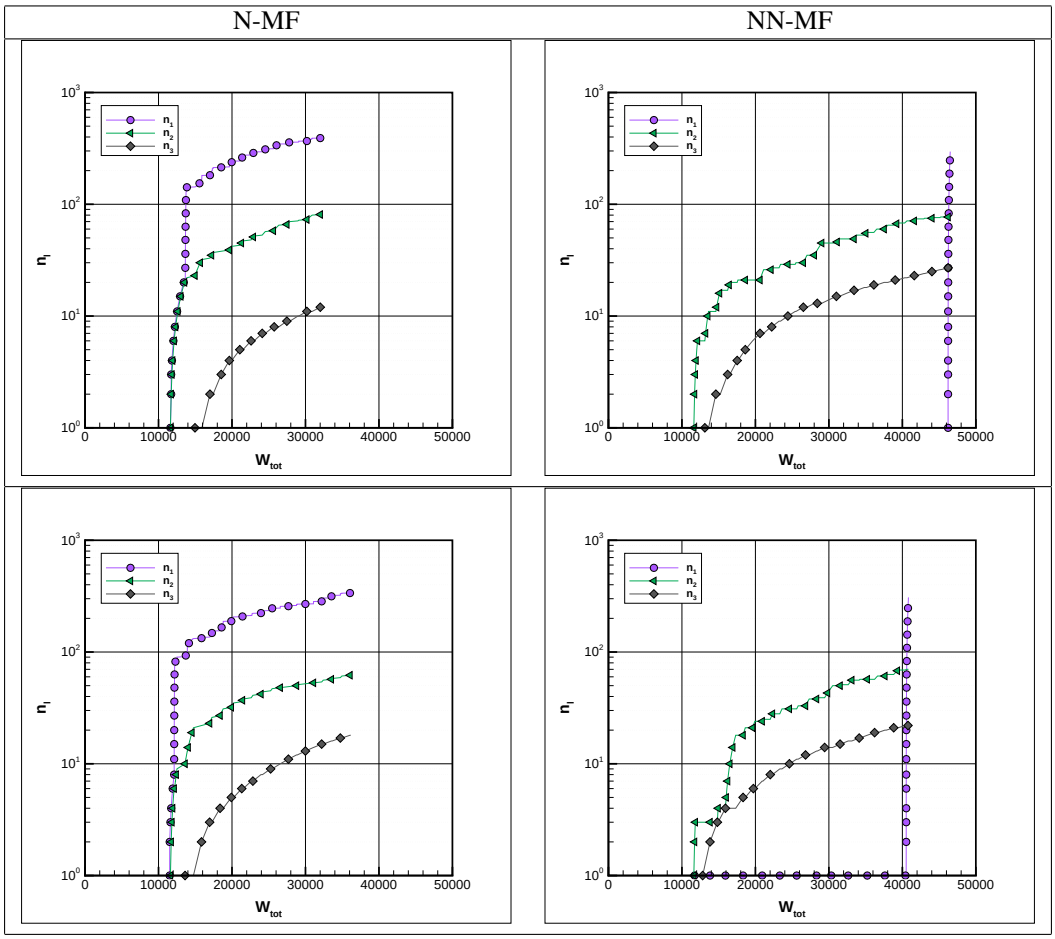


Figure 6: Training sets size as functions of the cost W_{tot} for the N-MF and NN-MF methods. Cases of $\delta = 0.05$ (top-row) and $\delta = 0.1$ (bottom-row).

4.4. Impact of costs

The behaviors of the MF-methods are dependent on the costs associated with the multi-fidelity levels. To illustrate this dependence, we increase the cost associated with the second fidelity model and set in this section $W_{(2)} = 800$. This choice increases sharply the contrast between the costs associated with the evaluation of models $l = 1$ and $l = 2$ while reducing the computational gap significantly with the highest fidelity model.

Figure 9 shows, as functions of W_{tot} and for $\delta = 0$ (left) and $\delta = 0.1$ (right), the distances and errors to the optimum. Once more, it is seen that the NN-MF-method outperforms the N-MF and SF-methods and remains effective for the two values of δ tested. In fact, for the NN-MF-method, the comparison with the respective plots in Figures 4 and 5 shows no qualitative changes when $W_{(2)}$ rises, but an increase of the overall cost of the optimization. For the N-MF-method, more erratic behavior is reported for $\delta = 0$, while the method is again seen to plateau when $\delta = 0.1$, with no convergence to the solution in the prescribed number of EGO iterations.

As expected, more significant changes occur in the size of the training sample sets when $W_{(2)}$ increases, see Fig. 10. A priori, reducing the gap between the costs of levels 2 and 3 will favor the selection of level 3 in the NN-MF-method because the computational saving at level 2 is less important. The effect in the N-MF-method is not as clear because selecting the level 3 implies to compute also levels 1 and 2. Overall, the plots of Fig. 10 indicate that for the N-MF-method, increasing $W_{(2)}$ favors a more frequent selection of $l = 3$, since the difference between n_2 and n_3 reduces compared to the previous cases $W_{(2)} = 100$. Regarding the first level, it is still infrequently selected during the initial phase of the N-MF-method, but it becomes the most selected level selected subsequently. The N-MF-method starts to plateau when $l = 1$ becomes the dominant level selected. In contrast, the differences between n_2 and n_3

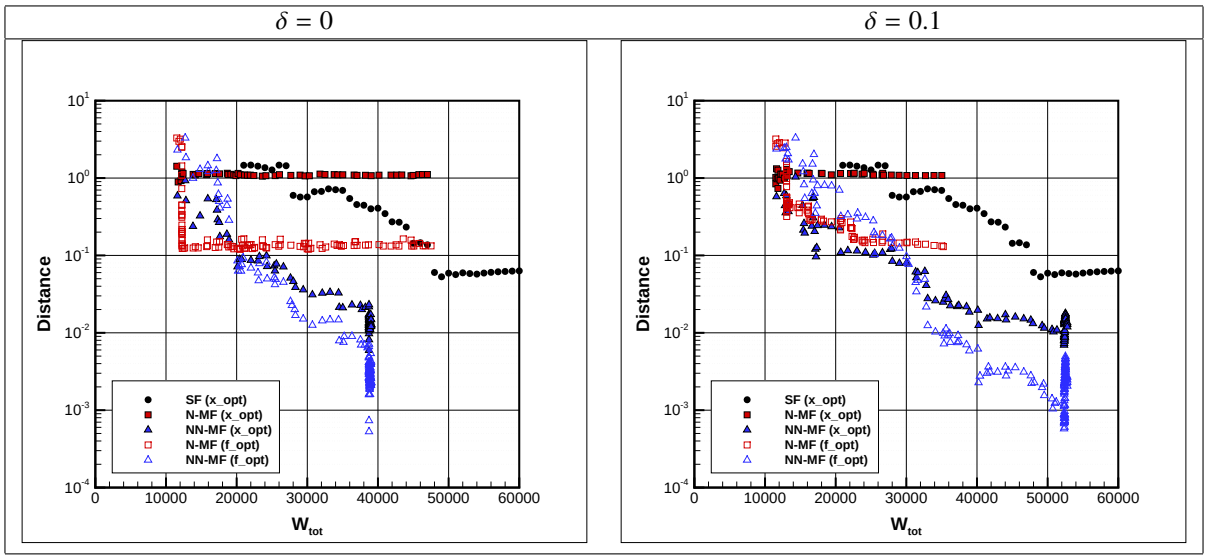


Figure 7: Distances to optimal solution as functions of the cost W_{tot} for the SF, N-MF and NN-MF methods. Cases of noisy estimations of $f^{(2)}$ (see (34)) with $\delta = 0$ (left) and $\delta = 0.1$ (right).

increases when $W_{(2)}$ is raised in the NN-MF-method. In the case with $\delta = 0.1$, the algorithm systematically selects the last level $l = 3$ until a fairly converged solution is obtained. In other words, the NN-MF-method returns to the SF method (with lower error because it can exploit the initial training sets at lower levels).

4.5. Summary

From the tests presented in this section, we can conclude that the efficiency of the N-MF-EGO method deteriorates as the correlation between the successive levels is decreasing. The nested property is also detrimental in the presence of evaluation noise or when the contrast in the cost of the highest models is not progressive enough, to the point where the N-MF-method can be less effective than its direct SF counterpart. The non-nested MF-method, on the contrary, is robust and much less sensitive to low correlations between levels. Besides, it can entirely focus the effort on the last level if the lower levels are not useful to improve the prediction or when the contrast in computational costs is not significant enough. This behavior indicates that the proposed NN-MF merit function provides a robust and flexible strategy able to disregard useless fidelity models. This feature is particularly attractive in situations where the fidelity levels may not possess a clear hierarchy.

5. Application to nonlinear FSI problem

We now apply the MF-EGO methods on a complex problem corresponding to the optimization of the trimmings of a realistic system of sails. The physics of sail systems involve very complex phenomena, such as nonlinear Fluid-Structure Interaction (FSI) and aero-elastic instabilities. Moreover, the modeling of real sailing conditions is still an open challenge because of the significant uncertainties in the prediction of wind and sea states. To our knowledge, the sails optimization has thus been limited so far to somehow idealized situations. For instance, the authors in [25] performed sail-shape optimizations without accounting for the full FSI problem. Sail trimmings optimization based on the FSI problem was considered in [4], but for two-dimensional "sails" only. Regarding three-dimensional FSI problems, the authors of [23] mention optimization of the trimming of sails, but within an inviscid flow approximation, and the authors provide very few details on the optimization procedure they used. In our opinion, our work in [27] was the first one to consider the trimming optimization of a realistic sail system, and validate the numerical optimization using a comparison with the optimization of an experimental sail system in a wind tunnel. The optimization procedure used in [27] is the SF-EGO method.

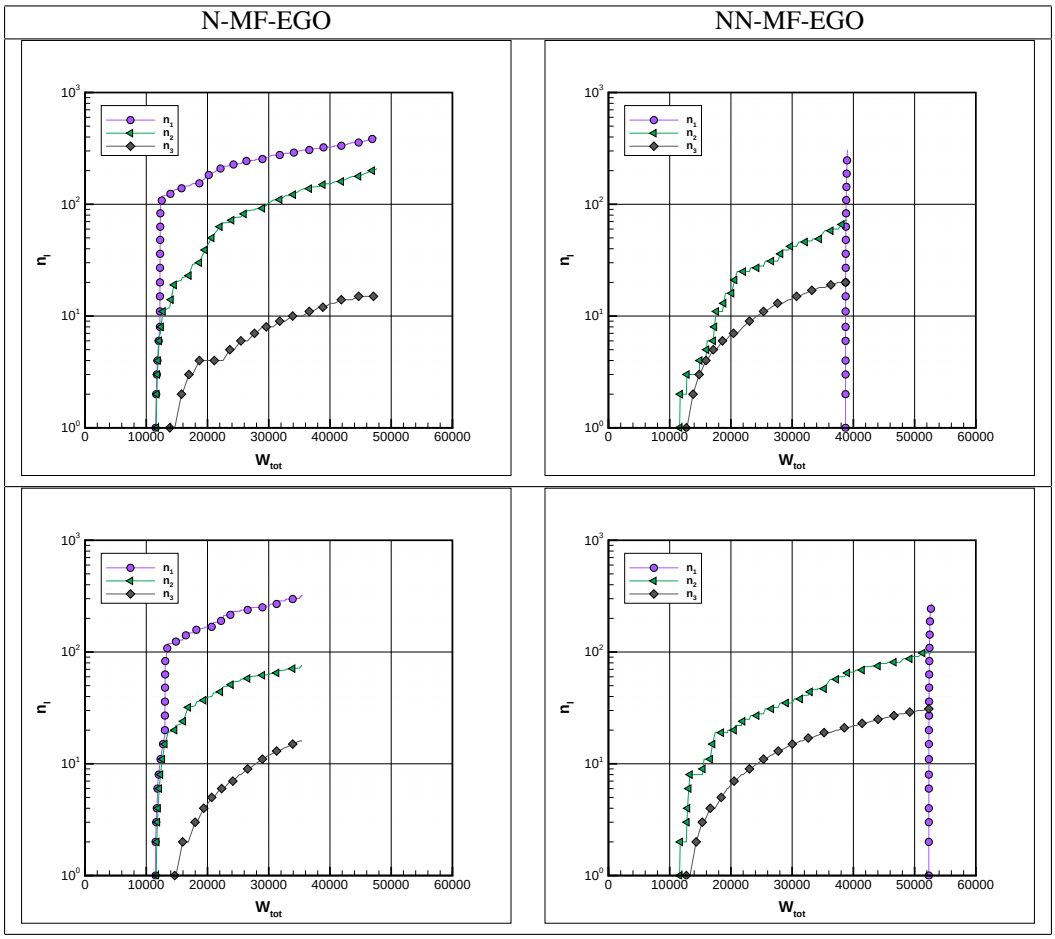


Figure 8: Training sets size as functions of the cost W_{tot} for the N-MF (left) and NN-MF (right). Cases of noisy estimations of $f^{(2)}$ (see (34)) with $\delta = 0$ (top) and $\delta = 0.1$ (bottom).

The present section aims at demonstrating the interest of the NN-MF-EGO approaches for the optimization of complex sail systems with realistic physical models (e.g., nonlinear FSI and turbulent flows). In particular, we want to evidence the reduction of the optimization cost of the NN-MF method. Multi-fidelity models are natural for the optimization of flexible sails systems, where the models for the flow (fluid problem) and the sails deformation (structure problem) can both involve different levels of simplification with significantly differing computational costs. As a result, one can combine these different structures and fluid models to obtain different fidelity models of the FSI problem. To focus on the multi-fidelity aspects, we restrict ourselves to a problem with only two trimming parameters (the mainsail and jib sheet length) but for a costly FSI modeling including detailed nonlinear structural model and three-dimensional turbulent flow simulations (the highest fidelity model).

Indeed, the numerical models of sail systems are composed of several fidelity levels, that may concern both the structural and fluid models. The following optimization problem includes five fidelity levels (in structure and fluid) on a two-parameter objective function in order to allow easy surrogate visualizations. The application is also solved with the highest fidelity level alone to perform SF and MF efficiency comparisons. The optimization problem and the numerical solvers are presented in Section 5.1. Results are then given in Section 5.2.

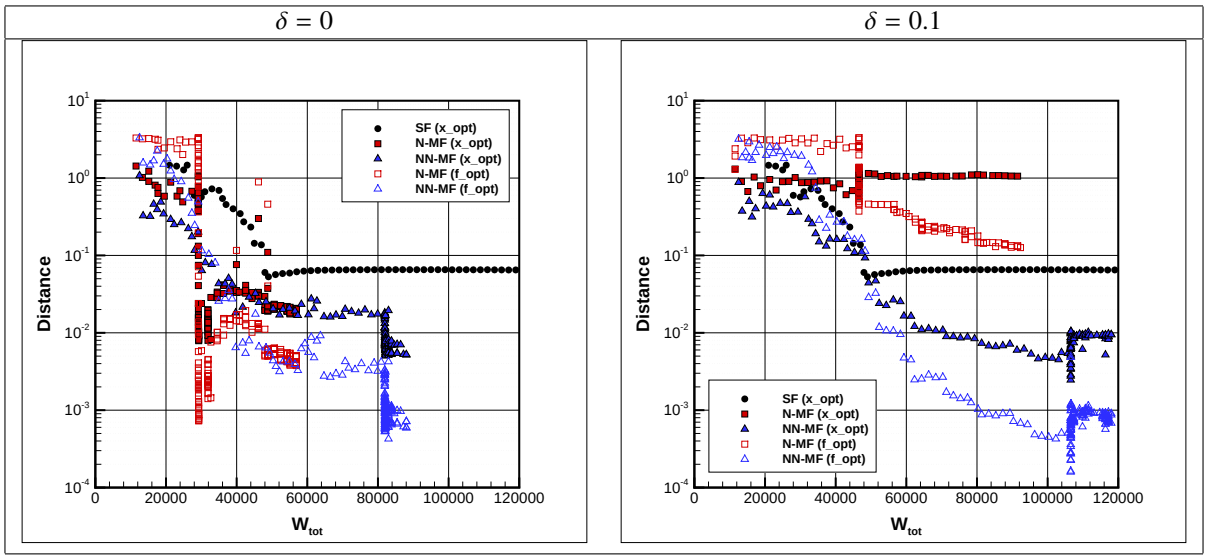


Figure 9: Distances to optimal solution as functions of the cost W_{tot} for the SF, N-MF and NN-MF-methods. Cases of $\delta = 0$ (left) and $\delta = 0.1$ (right), with cost $W_{(2)} = 800$.

5.1. Optimization of sails trimming

5.1.1. Problem definition

We consider the trimmings optimization of a *Figaro* mainsail and jib system. The *Incidence Sails* company provided the geometric and elastic characteristics of the two sails. The boat is assumed to cruise at a constant speed of 5.6 knots, in a constant wind with a true wind-speed of 8 knots (at 10 meters height) and a true wind-angle of 50 degrees. The trim, heel, and heading angles of the boat are fixed to 0, 10, and 0 (degrees), respectively. The optimization concerns two trimming parameters: the mainsail sheet (L_{main}) and jib sheet (L_{jib}) lengths (see Fig. 11).

The objective is to minimizing a composite objective function $f(\boldsymbol{x})$, consisting is a linear combination of the propulsion F_{prop} and the side F_{side} forces:

$$\boldsymbol{x}_{\text{opt}} = \arg\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}) = -\left(F_{\text{prop}}(\boldsymbol{x}) + 0.05F_{\text{side}}(\boldsymbol{x})\right). \tag{35}$$

Note that the propulsion force is counted negatively, for it is minimized, while the coefficient on F_{side} is set to penalize effects incurring from the side force, such as added hydrodynamic drag to counter the leeway-drift. The design domain is $\Omega = [1.5, 1.7] \times [2.45, 2.65]$, which correspond to a trimming range of 20 cm for the two sheets.

The MF framework uses L = 5 levels as reported in Table 1. The different levels combine different models of structure and fluid flow, each with increasing predictive capabilities. Concerning the structural differences, the sails and the rig will be considered rigid or flexible, depending on the level. Similarly, the flow models encompass inviscid and viscous assumptions, with different mesh refinements and boundary treatments in the latter case (levels $l = 4$ and 5). Specifically, the flow model at level $l = 5$ involves roughly 50% more cells, with resolved wall boundary conditions, compared to the preceding level, $l = 4$, which relies on a wall function. Table 1 also reports estimates of typical computational costs associated with the resolution of the different models. We see that the models based on the inviscid flow assumption are much less expensive to solve than for viscous flow models. Also, the mesh-refinement in the viscous flow models between levels 4 and 5 translates into a roughly ten times increase of the computational time.

5.1.2. Numerical solvers

We now briefly describe the different solvers involved in the optimization.

Structural solvers. We rely on the ARA software developed by K-Epsilon for the structural model of the elastic sails and rig. The definition of the geometric and elastic characteristics of the sails, as designed by Incidence Sails, are

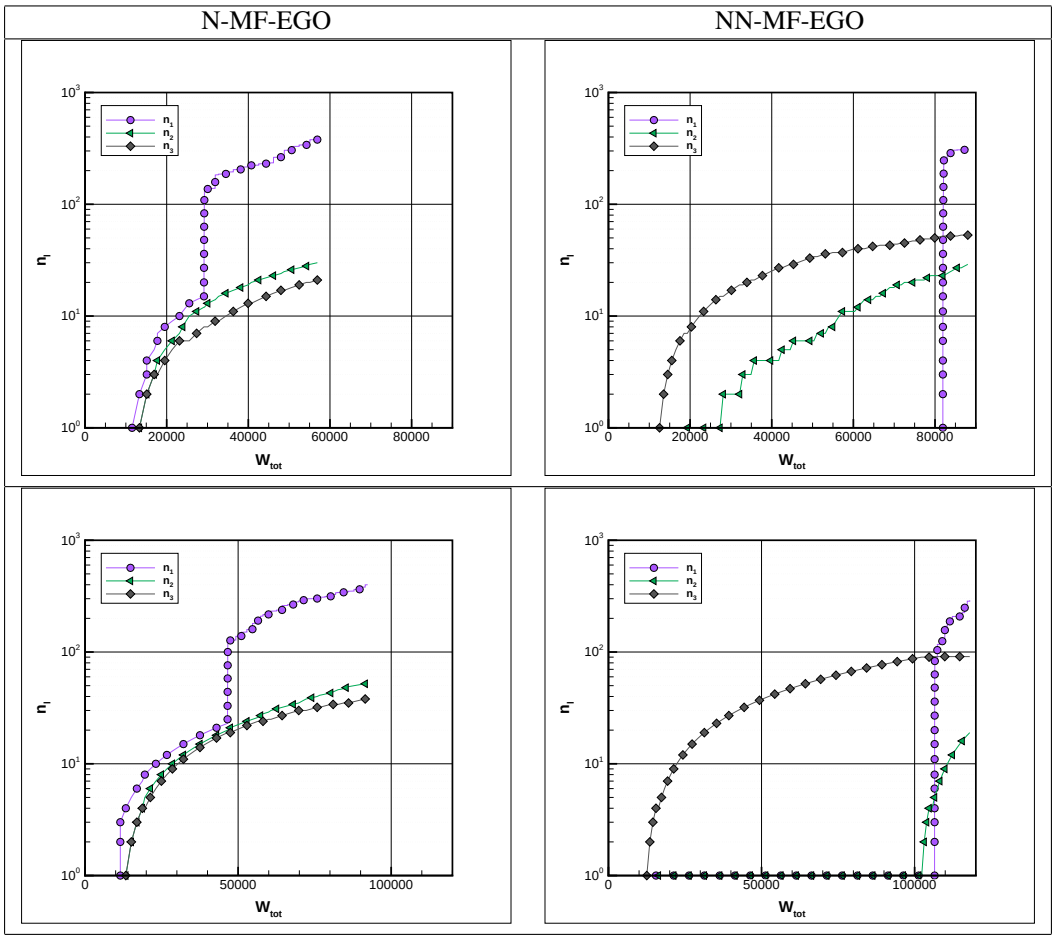


Figure 10: Training sets size as functions of the cost W_{tot} for the N-MF (left) and NN-MF (right) methods. Cases of $\delta = 0$ (top-row) and $\delta = 0.1$ (bottom-row), with $W_{(2)} = 800$.

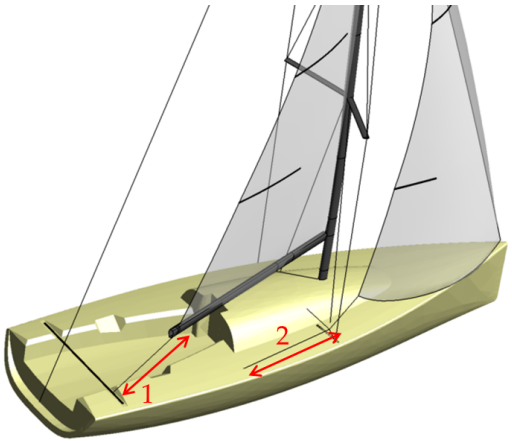


Figure 11: Design variables: mainsail sheet (1) and jib sheet (2) lengths.

imported using the SAILPACK software developed by BSG Developments. The elastic model involves Timoshenko beams for the rig (mast, boom) and sail (battens) elements, cables (sail sheets), and Constant Strain Triangles (CST)

Level	Sails model	Rig model	Fluid model	Fluid mesh size	Cost $c_{(l)}/c_{(1)}$
$l = 1$	Rigid	Rigid	Inviscid	–	1.0
$l = 2$	Flexible	Rigid	Inviscid	–	3.3
$l = 3$	Flexible	Flexible	Inviscid	–	4.2
$l = 4$	Flexible	Flexible	Viscous & Wall-law	2.3M	148.0
$l = 5$	Flexible	Flexible	Viscous & Resolved wall	3.6M	1430.4

Table 1: Structural and fluid model characteristics for the 5 fidelity levels considered.

membrane elements (cloth) of various types in large displacement formulation [2]. A wrinkle model [14] is added to capture local sail deformations with a reasonable number of membrane elements. See [5] or [20] for further details on CST elements and the wrinkle model. The nonlinear solutions (equilibrium under the fluid loads) are computed employing a Newton method with Aitken relaxation.

Flow solvers. Two different flow solvers implement the inviscid and viscous flow models. In the inviscid case, we use the software AVANTI [26]. This solver combines a Vortex Lattice Method (VLM) [14] at the sails surface, with a particle method for the dynamics of their wakes. In this approach, the sails surfaces are discretized in rectangular panels supporting doublets distributions, while particles carrying vorticity are released at the sails trailing edge to enforce the Kutta condition. The long-time limit of the unsteady simulation is considered as the steady inviscid solution. The inviscid model (see Fig. 12(a)) assumes slip conditions over all the sail surface and therefore cannot account for detached flow situations which can occur if the trimming parameters induce large angle of attacks or highly cambered sail profiles. However, the inviscid solver can accommodate substantial changes in the sails geometry since it requires only the discretization of the sail surfaces.

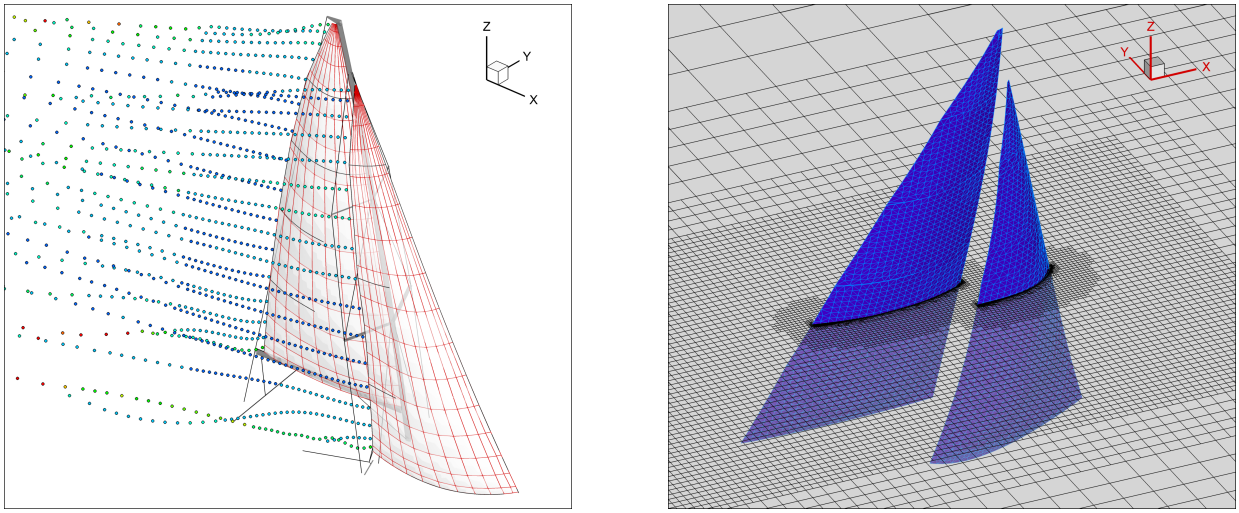
In the viscous case, we used ISIS-CFD software (of FINE$^{\text{TM}}$/Marine), which solves the incompressible Unsteady Reynolds-Average Navier-Stokes Equations (URANS). The turbulence model we used in this work is the two equations shear stress transport (SST) $k - \omega$ model [19]. This solver is based on a parallel finite-volume method, accommodating meshes with both structured and unstructured cells. Geometrical changes are less trivial to treat than for the inviscid case, and computationally more demanding, as the viscous solver requires a volume mesh fitted to the deformed sail geometries. In this work, we used the semi-automated mesh-generator HEXPRESS$^{\text{TM}}$ to construct meshes on reference geometries, and the Mesh Deformation Propagation (MDP) method [6] to adapt the mesh as the sails deform.

Coupled problem. We assume that the FSI problem has a steady solution for all values of the trimming parameters within the optimization range, and use a quasi-steady approach. Specifically, a quasi-monolithic algorithm [5] is employed to couple the structural and flow solvers, and produce steady FSI solutions. This algorithm relies on an implicit coupling strategy suited to partitioned solvers. Briefly, the resolution of the structural model is nested inside the iterations of the flow solver. This approach preserves the convergence and stability properties of the monolithic approach. More details on the coupling algorithms can be found in [5].

5.2. Optimization results

The NN-MF-method is initialized from nested training sets $\mathcal{X}^{(l)}$ with $n_l = 1 + 2(\text{L} - l)$ elements. Therefore, the initialization uses a single observation of $f^{(\text{L})}$, the model for which we wish to optimize the trimming parameters.

We start by presenting in Fig. 13 the MF-surrogates $\widehat{f^{(l)}}$ obtained at the end of the optimization procedure. We first remark a clear separation between the predictions of the surrogates based on the inviscid flow models ($l = 1, 2, 3$), and the predictions of the surrogates incorporating viscous flow estimations ($l = 4, 5$). These differences are particularly important as the trimming variables depart from the optimum point, with a clear under-estimation of the objective value in the inviscid cases: $\widehat{f^{(l=1,2,3)}}(\boldsymbol{x}) \ll \widehat{f^{(l=4,5)}}(\boldsymbol{x})$. This behavior is due to the inability of the inviscid models to account for the separation of the flow from the sails, as discussed below. Perhaps more surprising, $\widehat{f^{(1)}}$ is the inviscid MF-surrogate with predictions the closest to the viscous ones, albeit with the least dependences on the trimming parameters compared to the other inviscid surrogates $\widehat{f^{(2,3)}}$, which exhibit more noticeable dependences on the optimization variables. In contrast, the viscous MF-surrogates $\widehat{f^{(4)}}$ and $\widehat{f^{(5)}}$ have predictions in close agreement to each other over the whole domain Ω. This proximity is not surprising since, compared to $\widehat{f^{(4)}}$, the construction of



(a) Inviscid model. (b) Viscous model.

Figure 12: Illustration of the fluid meshes corresponding to the inviscid and viscous solvers.

$\widehat{f^{(5)}}$ incorporates just a few evaluations of the objective function $f^{(5)}$, based on a more refined mesh and a different boundary flow conditions, but the same physical models otherwise.

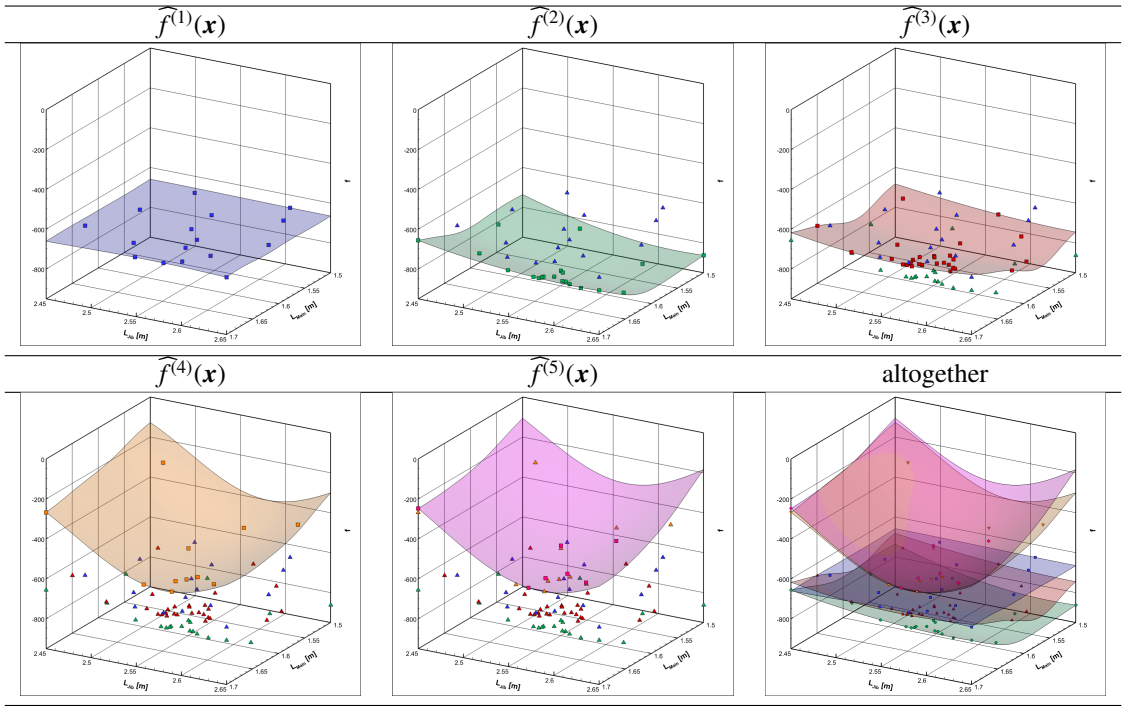
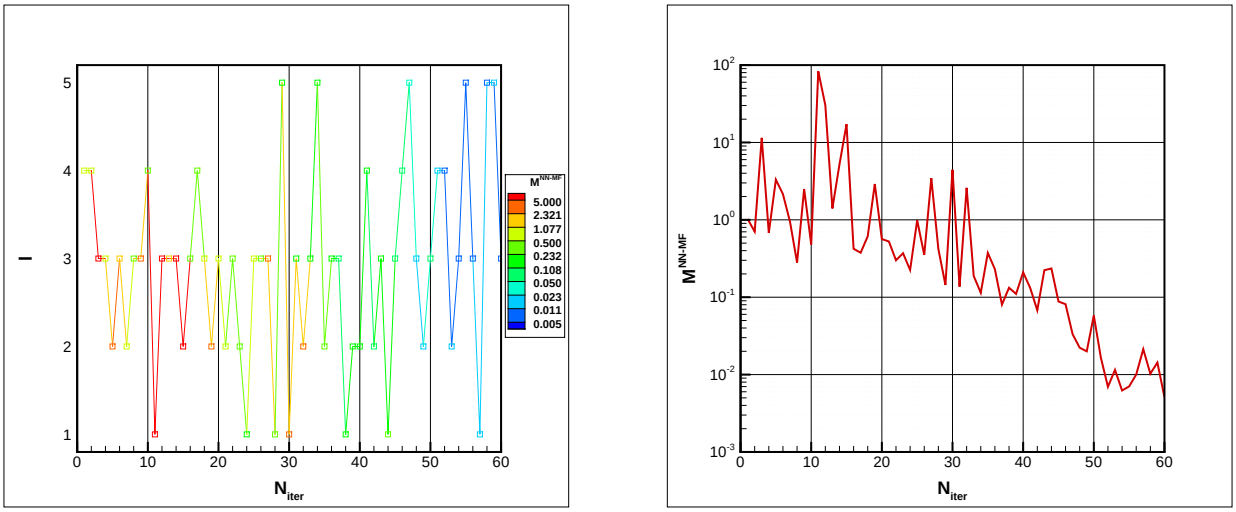


Figure 13: MF surrogates after $N_{iter} = 60$ iterations of the NN-MF optimization.

Eventhough, the MF-models $\widehat{f^{(l)}}(\boldsymbol{x})$ reported above are only approximations of the MF-models $f^{(l)}(\boldsymbol{x})$, they provide a fair idea of the situation. In particular, they highlight potential difficulty of defining a clear hierarchy of fidelity level, when considering various physical approximations. Such situations clearly call for robust and agile MF-construction

and infilling strategies, in order to minimize as much as possible the impact of *a priori* regarding the predictive capabilities of the models.

Figure 14 depicts the convergence of the optimization process up to $N_{iter} = 60$ iterations of the NN-MF approach. The value of M^{NN-MF} is used to color the curve, and it indicates the expected improvement in the computed levels as a function of N_{iter}. Thus, we observe a decrease of the merit function and the procedure converges in $N_{iter} \approx 60$ iterations. We also qualitatively note that levels 2 and 3 are the ones that have been evaluated most of the time during the optimization.



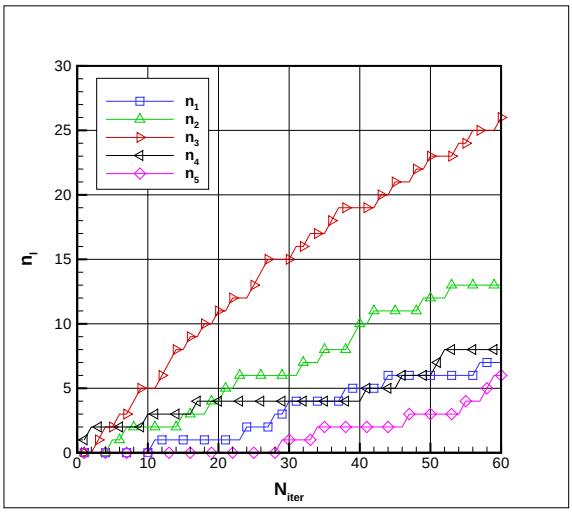
(a) (b)

Figure 14: MF-EGO convergence when solving (35).

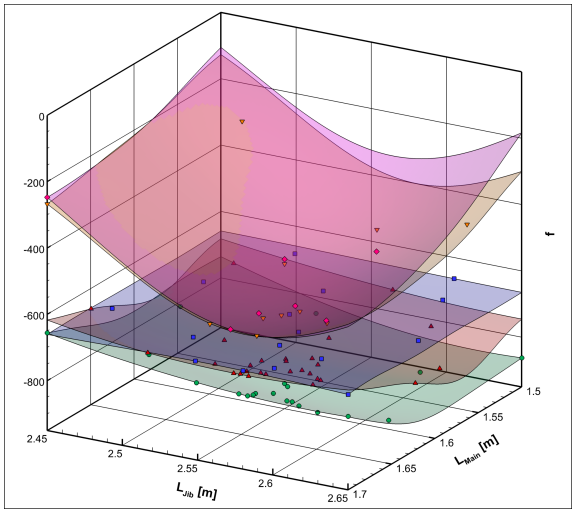
Figure 15(a) reports the number of calls per fidelity level as a function of N_{iter}. The level $l = 3$ was selected most of the time, while level $l = 1$ has approximately the same number of calls than level 4 and 5. We also note that during the first 20 iterations, the level $l = 4$ is the second fidelity level of choice. These trends can be explained using Fig. 15(b), which depicts the mean prediction of the objective function $f_l(\boldsymbol{x})$ as a function of the two trimming parameters for all fidelity levels. The surface colors correspond to the colors used to differentiate the levels in Fig. 15(a). We note that the prediction f_1 is closer to the highest fidelity level f_5, compared to the predictions f_2 and f_3. However, the prediction of the objective function by the level 1 model has very low variabilities compared to the other levels. Consequently, although the computational cost of this level is low, the non-nested EGO approach does not compute a large number of observations at level $l = 1$ because the model is easy to learn, and additional points do not provide a significant reduction of the prediction variance at the highest fidelity level $l = 5$. Models at levels 2 and 3 overestimate the objective function, but they contain variabilities that correlate well with f_5. The level 4 is very close to the highest fidelity level in terms of mean response and variabilities. This proximity is not surprising, since both models use the same structural and flow solvers, and differences are only on the sails boundary conditions and mesh refinements. Overall, the non-nested approach explores Ω using low-cost levels and avoids useless fidelity levels by eventually refining promising areas with the two highest-fidelity levels.

Figure 16 provides a (sailor) view of the optimally trimmed sails that are given by each fidelity level. The sails and rigs colors correspond to fidelity levels colors already used in Fig. 15(a). Those views allow us to understand the differences in objective responses of each level shown in Fig. 15(b). The two viscous models 4 and 5 are very close in terms of the jib and mainsail shapes. Differences are much more significant with the fidelity levels based on the inviscid model. The mainsail drafts, in particular, are more pronounced for the viscous models. This difference comes from the inviscid flow model, which assumes an attached flow over the sails, leading to overestimations of aerodynamic forces and sail shapes with large drafts as a result.

We also solve the trimming optimization problem (35) using the SF-EGO approach one the highest fidelity level 5, using an initial LHS plan of 5 observations. Figure 17 shows the convergences of the SF and MF optima as functions

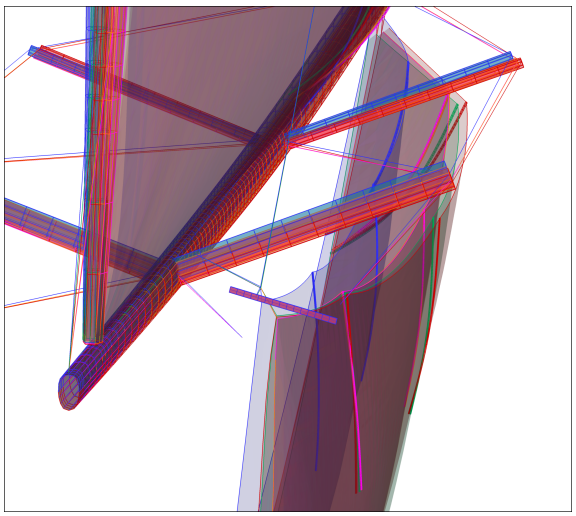


(a) number of calls per level as a function of N_{iter}.

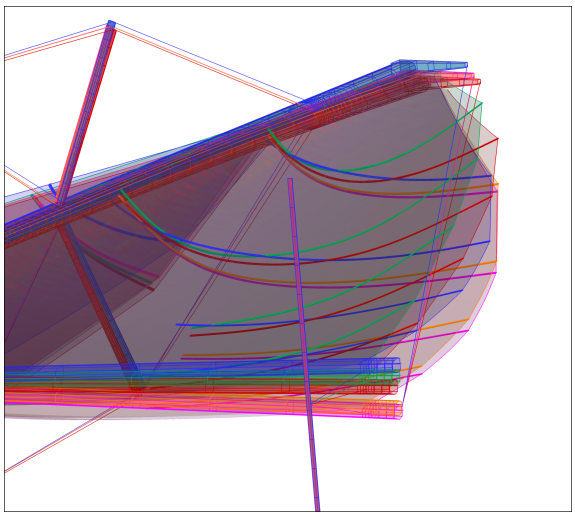


(b) MF surrogates at $N_{iter} = 60$. The surface colors correspond to the ones in the legend of Fig. 15(a).

Figure 15: Number of call per level and MF surrogates at convergence when solving (35).



(a) Jib view.



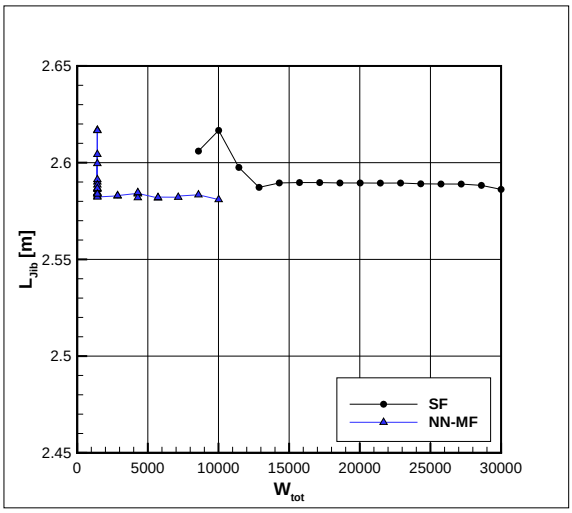
(b) Main view.

Figure 16: Optimal trimming at each fidelity level when solving (35).

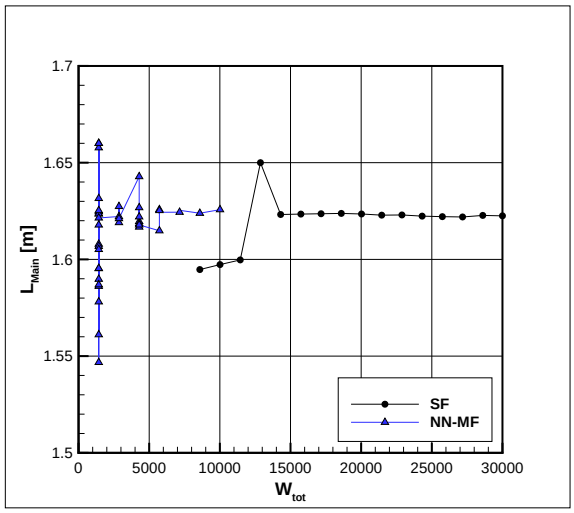
of the cost. The NN-MF has a computational time ≈ 3 in comparison to the computation time of the SF method. At the end of the optimizations, the mainsail and jib lengths parameters are in very close agreement with differences that are ≈ 0.25 cm for the mainsail and ≈ 0.55 cm for the jib.

6. Conclusion

We have presented in this paper a new infilling formulation for multi-fidelity surrogate-based optimizations. The multi-fidelity approach allows using Efficient Global Optimization (EGO) with surrogate models based on observations at different fidelity levels. We proposed and compared two distinct approaches differing by the use of nested or non-nested training-sets. In the two cases, we proposed suitable merit functions, based on the concept of Expected



(a) Jib length variable as a function of the cost.



(b) Main length variable as function of the cost.

Figure 17: Optimal trimmings of SF and MF surrogates when solving (35).

Improvement, to select the new point and level in the EGO procedure. Specifically, the sequential approach finds the most promising design point and the fidelity level for the evaluation of the objective function at the new point. The selection criteria balance the computational cost at each level and the resulting uncertainty reduction in the multi-fidelity estimation of the objective function.

The proposed MF optimization approaches were validated and tested on several analytical problems. The results show that, compared to the nested formulation, the non-nested MF approach is more efficient in terms of numerical cost, but also much more robust against noisy evaluations of intermediate fidelity levels and low correlations between levels. It is also less sensitive to the computational costs associated with the levels.

Finally, the method was applied to the optimization of a fully nonlinear numerical fluid-structure interaction problem. Here we aimed at optimizing the sail trimmings of Figaro racing yacht. The problem has two parameters, and we considered five fidelity levels related both to structure and fluid models. In this example, the non-nested multi-fidelity EGO was compared to the single-fidelity EGO only. This complex application highlighted the interest of using a non-nested multi-fidelity approach to reduce the computational cost significantly.

Future works will concern the formulation of multi-fidelity surrogate-based optimization in the parallel computing framework. Indeed, the computational costs and the multi-fidelity responses can be used to define a formulation of merit function, which allows several evaluations at each iteration of the EGO. Such a parallel extension would help of reducing, even more, the global cost of the optimization.

References

[1] N. Aghajari and M. Schäfer. Efficient shape optimization for fluid–structure interaction problems. *Journal of Fluids and Structures*, 57:298 – 313, 2015.

[2] B. Augier. *Etudes expérimentales de l'interaction fluide-structure sur surface souple: application aux voiles de bateaux.* PhD Thesis, Université de Bretagne Occidentale, 2012.

[3] N. Bartoli, M. Meliani, J. Morlier, T. Lefebvre, M.-A. Bouhlel, and J. Martins. Multi-fidelity efficient global optimization: Methodology and application to airfoil shape design. In *AIAA Aviation 2019 Forum*, page 3236, 2019.

[4] V. G. Chapin, R. Neyhousser, G. Dulliand, and P. Chassaing. Design optimization of interacting sails through viscous CFD. In *INNOVSail, Innovation in high performance sailing Yacht*, Lorient, 2008.

[5] M. Durand. *Interaction fluide-structure souple et légère, application aux voiliers*. PhD Thesis, Ecole Centrale de Nantes, 2012.

[6] M. Durand, A. Leroyer, C. Lothodé, F. Hauville, M. Visonneau, R. Floch, and L. Guillaume. FSI investigation on stability of downwind sails with an automatic dynamic trimming. *Ocean Engineering*, 90:129–139, 2014.

[7] K. Elsayed. Optimization of the cyclone separator geometry for minimum pressure drop using co-kriging. *Powder Technology*, 269 (Supplement C):409 – 424, 2015. ISSN 0032-5910.

[8] A. I. Forrester, A. Sóbester, and A. J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 463(2088):3251–3269, 2007. ISSN 1364-5021.

[9] B. Glaz, P. P. Friedmann, and L. Liu. Helicopter vibration reduction throughout the entire flight envelope using surrogate-based optimization. *Journal of the American Helicopter Society*, 54(1):12007, 2009.

[10] N. Hansen. The CMA Evolution Strategy: A Comparing Review. In *Towards a new evolutionary computation*, pages 75–102. Springer, 2006.

[11] D. Huang, T. T. Allen, W. I. Notz, and N. Zheng. Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models. *Journal of global optimization*, 34(3):441–466, 2006.

[12] S. Jeong, M. Murayama, and K. Yamamoto. Efficient optimization design method using kriging model. *Journal of aircraft*, 42(2):413–420, 2005.

[13] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global optimization*, 13(4):455–492, 1998.

[14] J. Katz and A. Plotkin. *Low-Speed Aerodynamics*. Cambridge University Press, second edition, 2001.

[15] M. C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000. ISSN 00063444.

[16] J. P. Kleijnen. Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, 192(3):707 – 716, 2009. ISSN 0377-2217.

[17] L. Le Gratiet and C. Cannamela. Kriging-based sequential design strategies using fast cross-validation techniques with extensions to multi-fidelity computer codes. working paper or preprint, Oct. 2012. URL `https://hal.archives-ouvertes.fr/hal-00744432`.

[18] L. Le Gratiet and J. Garnier. Recursive co-kriging model for Design of Computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5):365–386, 2014. URL `https://hal.archives-ouvertes.fr/hal-01108813`.

[19] F. R. Menter, M. Kuntz, and R. Langtry. Ten Years of Industrial Experience with the SST Turbulence Model. *Turbulence, heat and mass transfer*, 4:625–632, 2003.

[20] K. Nakashino and M. C. Natori. Efficient modification scheme of stress-strain tensor for wrinkled membranes. *AIAA journal*, 43(1):206–215, 2005.

[21] V. Picheny, T. Wagner, and D. Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626, Sep 2013. ISSN 1615-1488.

[22] M. Plutowski, S. Sakata, and H. White. Cross-validation estimates imse. In *Advances in neural information processing systems*, pages 391–398, 1994.

[23] R. Ranzenbach, D. Armitage, and A. Carrau. Mainsail Planform Optimization for IRC 52 Using Fluid Structure Interaction. In *The 21st Chesapeake Sailing Yacht Symposium, SNAME*, 2013.

[24] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[25] N. Rousselon. Optimization for Sail Design. In *modeFRONTIER Conference*, 2008.

[26] Y. Roux, M. Durand, A. Leroyer, P. Queutey, M. Visonneau, J. Raymond, J. M. Finot, F. Hauville, and A. Purwanto. Strongly coupled VPP and CFD RANSE code for sailing yacht performance prediction. In *3rd High Performance Yacht Design Conference, 2-4 December*, pages 215–225, Auckland, 2008.

[27] M. Sacher, F. Hauville, R. Duvigneau, O. Le Maître, N. Aubin, and M. Durand. Efficient optimization procedure in non-linear fluid-structure interaction problem: Application to mainsail trimming in upwind conditions. *Journal of Fluids and Structures*, 69:209 – 231, 2017. ISSN 0889-9746.

[28] T. Simpson, J. Poplinski, N. P. Koch, and J. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, 17(2):129–150, 2001. ISSN 1435-5663.