CSMA 2026
17ème Colloque National en Calcul des Structures
18-22 Mai 2026, Presqu'île de Giens (Var)

Sparse approximation of recursive multi-fidelity Kriging for data fusion in aerodynamics

Mauricio Castaño-Aguirre[1,5], Andrés F. López-Lopera[3], Nathalie Bartoli[1,2], Franck Massa[4,5], Thierry Lefebvre[1,2]

[1] *DTIS, ONERA, Université de Toulouse, 31000, France*
[2] *Fédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, 31000, Toulouse, France*
[3] *IMAG, Université Montpellier, CNRS, 34090, France*
[4] *INSA Hauts-de-France, F-59313 Valenciennes, France*
[5] *Université Polytechnique Hauts-de-France, CNRS, UMR 8201-LAMIH, Mechanical Department, F-59313, Valenciennes, France*

Abstract — Multi-fidelity Kriging (MFK) surrogate modeling combines data of varying accuracy, such as experimental tests and numerical simulations, to improve predictive performance. Autoregressive models are commonly used to capture correlations across fidelity levels, offering interpretability compared to purely data-driven approaches. However, existing MFK frameworks based on autoregressive models often suffer from high computational costs. In this work, we revisit the MFK approach, extending it to account for sparsity in the multi-fidelity setting, maintaining the efficiency of classical sparse approximations while keeping accuracy. We illustrate our framework in a 1D toy example, and assess it on a large-scale multi-fidelity aerodynamic dataset that combines wind tunnel experiments with CFD simulations. In both cases, our framework is capable of handling heteroscedastic noise. Training performance is reduced by approximately 88% while maintaining accuracy compared to standard MFK, highlighting the benefit of incorporating sparse approximations in complex engineering applications.
Mots clés — Gaussian processes, Heteroscedastic noise, Large-scale data fusion, Sparse approximations.

1 Introduction

The design of aerospace vehicles increasingly depends on aerodynamic data from different sources, such as Computational Fluid Dynamics (CFD) simulations, Wind Tunnel (WT) tests, and flight measurements [1, 2]. Although these sources describe the same system, they differ in accuracy, cost, and availability [2]. Wind tunnel tests provide controlled and repeatable experimental conditions that help validate numerical models and study complex aerodynamic effects. On the other hand, numerical simulations, especially CFD, offer great flexibility and lower costs, making them ideal for exploring a wide range of design configurations. Thanks to major methodological and computational advances achieved in recent decades, CFD has become a key tool in the early stages of aerospace design [1].

While these different sources of aerodynamic data were once regarded as competing alternatives, there is now a broad consensus that they are inherently complementary [2]. The contrasts in their characteristics and in the types of information they provide illustrate this complementarity. For instance, CFD simulations can deliver detailed spatial distributions of aerodynamic quantities such as local pressure fields across the aircraft surface, yet their use is typically restricted to a limited number of flow conditions due to the substantial computational cost of each run. Conversely, WT experiments provide only scarce measurements, constrained by the physical placement of sensors on the model, but enable rapid evaluation of numerous flow configurations through controlled variations of the tunnel parameters. Recognizing these distinct yet synergistic capabilities, recent research emphasizes the need for advanced data fusion frameworks capable of integrating heterogeneous aerodynamic datasets into a unified representation [3]. Such methods seek to leverage the complementary strengths of CFD, WT, and flight data, thereby maximizing the overall informational value available for aerospace design and analysis.

MFK provides an efficient framework for combining information from multiple sources of varying fidelity, leveraging Gaussian Process (GP) models and an autoregressive formulation to relate different fidelity levels [3, 4, 5, 6]. While the recursive MFK formulation in [7] enables scalable predictions for moderate datasets, it inherits the standard computational limitations of GP models, which scale cubically

with the number of observations at each fidelity level [8]. This becomes a significant bottleneck when dealing with large-scale aerodynamic datasets.

To overcome this limitation, we adapt a sparse approximation to the MFK framework. The key idea is to approximate the full covariance structure of the lowest fidelity level using a smaller set of inducing points, reducing computational complexity while preserving predictive accuracy. In particular, we adopt the Free Training Conditional (FITC) method in [9], which formulates sparse GP using low rank approximations. By selecting inducing points that maximize a lower bound on the marginal likelihood, FITC provides an approach that minimizes information loss and ensures robust predictions. Integrating FITC into the MFK model leads to a Sparse MFK (SMFK) framework, which maintains the strengths of the autoregressive MFK formulation while enabling efficient and scalable inference on large multi-fidelity datasets that are characterized by having large amounts of data at the lowest levels of fidelity.

The structure of this paper is organized as follows. Section 2 revisits the recursive MFK method proposed by [4]. Section 3 introduces the FITC sparse approximation for the recursive MFK model. The results and advantages of the proposed approach are discussed in Section 4. Finally, Section 5 summarizes the main findings and conclusions of the study.

2 Multi-fidelity Kriging framework

2.1 Autoregressive model

For $\kappa \in \{1,\dots,L\}$ with $L \geq 1$, denote $y_\kappa : \mathcal{D} \to \mathbb{R}$ with $\mathcal{D} \subseteq \mathbb{R}^D$ the function corresponding to the κ-th fidelity level. By convention, we consider $y_0(\cdot)$ and $y_L(\cdot)$ as the lowest and highest fidelity functions, respectively. The autoregressive framework establishes a relationship between two consecutive levels of fidelity given by

$$\begin{aligned} f_\kappa(\boldsymbol{x}) &= \rho_\kappa\, f_{\kappa-1}(\boldsymbol{x}) + \gamma_\kappa(\boldsymbol{x}), && \text{(autoregressive model)} \\ y_\kappa(\boldsymbol{x}) &= f_\kappa(\boldsymbol{x}) + \varepsilon_\kappa(\boldsymbol{x}), && \text{(observation model)} \end{aligned} \tag{1}$$

where $\boldsymbol{x} \in \mathcal{D}$. The model in Eq. (1) considers that $f_\kappa(\cdot)$ is a scaled version of the previous fidelity level $f_{\kappa-1}(\cdot)$ by a scale factor $\rho_\kappa \in \mathbb{R}$. In addition, an additive extra term $\gamma_\kappa : \mathcal{D} \to \mathbb{R}$ is considered to capture the non-linear discrepancies between $f_\kappa(\cdot)$ and $f_{\kappa-1}(\cdot)$. The term $\varepsilon_\kappa(\cdot)$ denotes the additive heteroscedastic noise that accounts for input-dependent variability. Note that the variable ρ_κ is treated here as constant parameter to be estimated, but it can be considered as a function, i.e., $\rho_\kappa : \mathcal{D} \to \mathbb{R}$ [3, 4].

2.2 Recursive MFK framework for nested design of experiments

From a probabilistic point of view, we can place GP priors over the functions $\gamma_0(\cdot),\dots,\gamma_L(\cdot)$. Here, for ease of notation, we denote $\gamma_0(\cdot) := f_0(\cdot)$. In the context of the autoregressive model, this definition arises naturally since the autoregressive formulation begins with a GP at the lowest fidelity level. Consequently, the latent function at this base level, $f_0(\cdot)$, can be directly identified as the GP $\gamma_0(\cdot)$. Here, we consider $\gamma_\kappa(\cdot) \sim \mathcal{GP}(0, k_{\gamma_\kappa})$ with kernel $k_{\gamma_\kappa}(\cdot,\cdot)$ for all $\kappa \in \{1,\dots,L\}$. We assume that each $\gamma_\kappa(\cdot)$ is independent, i.e., $\gamma_\kappa(\cdot) \perp \gamma_{\kappa'}(\cdot)$ for all $\kappa \neq \kappa'$, with $\kappa, \kappa' \in \{1,\dots,L\}$.

As shown in [10], the autoregressive model described in Eq. (1) results from the Markov assumption $\operatorname{cov}(f_\kappa(\boldsymbol{x}), f_{\kappa-1}(\boldsymbol{x}') | f_{\kappa-1}(\boldsymbol{x})) = 0$ for all $\boldsymbol{x} \neq \boldsymbol{x}'$ and for all $\kappa \in \{1,\dots,L\}$. This property means that if $f_{\kappa-1}(\boldsymbol{x})$ is known, then nothing more can be learned about $f_\kappa(\boldsymbol{x})$ from any observation of the lower fidelity $f_{\kappa-1}(\boldsymbol{x}')$ for $\boldsymbol{x}' \neq \boldsymbol{x}$ [4]. In addition, when conditioning the model to observation data, it is assumed that the experimental designs $\boldsymbol{X}_0,\dots,\boldsymbol{X}_L$ are nested, i.e., $\boldsymbol{X}_\kappa \subseteq \boldsymbol{X}_{\kappa-1}$. Here, $\boldsymbol{X}_\kappa = \left(\boldsymbol{x}_{\kappa,i}\right)_{1 \leq i \leq N_\kappa}$ is the DoE of the κ-th fidelity level containing the $N_\kappa \in \mathbb{N}$ input values. This nested assumption can be dropped, but allows obtaining closed-form expressions [4].

As suggested by [7], it is possible to establish a recursive implementation of the GP $y_\kappa(\cdot)$ by using the posterior of the GP $f_{\kappa-1}(\cdot)$ that adjusts the dataset $(\boldsymbol{X}_{\kappa-1}, \boldsymbol{y}_{\kappa-1})$ with the observation vector $\boldsymbol{y}_{\kappa-1} = y_{\kappa-1}(\boldsymbol{X}_{\kappa-1})$. Denote as $\widehat{f}_{\kappa-1}(\cdot) \sim \mathcal{GP}(m_{c,\kappa-1}, k_{c,\kappa-1})$ the posterior of the fidelity level $\kappa-1$. Then, we have

$$y_\kappa(\boldsymbol{x}) = \rho_\kappa\, \widehat{f}_{\kappa-1}(\boldsymbol{x}) + \gamma_\kappa(\boldsymbol{x}) + \varepsilon_\kappa(\boldsymbol{x}).$$

We recall that $\gamma_\kappa(\cdot) \sim \mathcal{GP}(0, k_{\gamma_\kappa})$ with kernel $k_{\gamma_\kappa}(\cdot,\cdot)$ for all $\kappa \in \{0,\dots,L\}$. We can then establish the next expressions for $\kappa \in \{1,\dots,L\}$.

- For $\kappa = 0$, as $f_0(\cdot) =: \gamma_0(\cdot) \sim \mathcal{N}(0, k_{\gamma_0})$, then $\widehat{f}_0(\cdot)$ is GP-distributed with classical conditional mean and conditional variance given by

$$\begin{aligned} m_{c,0}(\boldsymbol{x}) &= \boldsymbol{k}_0^\top(\boldsymbol{x})(\boldsymbol{K}_0 + \boldsymbol{T}_0)^{-1}\boldsymbol{y}, \\ v_{c,0}(\boldsymbol{x}) &= k_{\gamma_0}(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{k}_0^\top(\boldsymbol{x})(\boldsymbol{K}_0 + \boldsymbol{T}_0)^{-1}\boldsymbol{k}_0(\boldsymbol{x}), \end{aligned} \tag{2}$$

  where $\boldsymbol{K}_0 = k_{\gamma_0}(\boldsymbol{X}_0, \boldsymbol{X}_0)$ and $\boldsymbol{k}_0(\boldsymbol{x}) = k_{\gamma_0}(\boldsymbol{X}_0, \boldsymbol{x})$. We denote the parameters of $k_{\gamma_0}(\cdot,\cdot)$ as $\boldsymbol{\theta}_0$ and $\boldsymbol{T}_0 = \mathrm{diag}(\mathrm{var}(\varepsilon_0(\boldsymbol{x}_{0,1})),\dots,\mathrm{var}(\varepsilon_0(\boldsymbol{x}_{0,N_0}))) = \mathrm{diag}(\tau^2_{0,1},\dots,\tau^2_{0,N_0})$ corresponding to the noise variances of the lowest fidelity level.

- For $1 \le \kappa \le L$, we have that $f_\kappa(\cdot) \sim \mathcal{N}(\rho_\kappa m_{c,\kappa-1}, \rho_\kappa^2 k_{c,\kappa-1} + k_{\gamma_\kappa})$ and $\widehat{f}_\kappa(\cdot)$ is GP-distributed with conditional mean and conditional variance given by

$$\begin{aligned} m_{c,\kappa}(\boldsymbol{x}) &= \rho_\kappa m_{c,\kappa-1}(\boldsymbol{x}) + \boldsymbol{k}_\kappa^\top(\boldsymbol{x})(\boldsymbol{K}_\kappa + \boldsymbol{T}_\kappa)^{-1}(\boldsymbol{y} - \rho_\kappa m_{c,\kappa-1}(\boldsymbol{X}_\kappa)), \\ v_{c,\kappa}(\boldsymbol{x}) &= \rho_\kappa^2 v_{c,\kappa-1}(\boldsymbol{x}) + k_{\gamma_\kappa}(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{k}_\kappa^\top(\boldsymbol{x})(\boldsymbol{K}_\kappa + \boldsymbol{T}_\kappa)^{-1}\boldsymbol{k}_\kappa(\boldsymbol{x}), \end{aligned} \tag{3}$$

  where $\boldsymbol{K}_\kappa = k_{\gamma_\kappa}(\boldsymbol{X}_\kappa, \boldsymbol{X}_\kappa)$ and $\boldsymbol{k}_\kappa(\boldsymbol{x}) = k_{\gamma_\kappa}(\boldsymbol{X}_\kappa, \boldsymbol{x})$. Here, $\boldsymbol{\theta}_\kappa$ denotes the parameters of $k_{\gamma_\kappa}(\cdot,\cdot)$, and $\boldsymbol{T}_\kappa = \mathrm{diag}(\tau^2_{\kappa,1},\dots,\tau^2_{\kappa,N_\kappa})$ denotes the noise variances at each fidelity level κ. In this case, one must jointly estimate $(\boldsymbol{\theta}_\kappa, \rho_\kappa)$.

We clarify that we limited the developments to $\widehat{f}_\kappa(\cdot)$ since $\widehat{y}_\kappa(\cdot)$ cannot be evaluated as the variance $\mathrm{var}(\varepsilon_\kappa(\boldsymbol{x}))$ is unknown for any unobserved point $y_\kappa(\boldsymbol{x})$. We also note that both parameter sets, $(\boldsymbol{\theta}_\kappa)_{0\le\kappa\le L}$ and $(\rho_\kappa)_{1\le\kappa\le L}$, can be estimated via maximum likelihood [8]. From Eq. (2) and Eq. (3), we observe that the recursive MFK suffers from scalability issues for large datasets, as the computational complexity per fidelity level is $O(N_\kappa^3)$. This motivates the use of sparse approximations [11, 12] at low fidelity levels where N_κ is large. This is typically the case at the lowest fidelity level (i.e., $\kappa = 0$), where data are obtained from fast and inexpensive computer simulations [5, 6].

3 Sparse approximations for the recursive multi-fidelity Kriging

Sparse approximations in GPs, such as the Fully Independent Training Conditional (FITC) [11], may allow scalable inference and effective information transfer between fidelity levels, thereby reducing computational costs while preserving the essential correlations across fidelities [10]. The objective of this section is to consider an FITC approximation at the lowest fidelity level. Revisiting the model of Section 2, we can observe that the lowest fidelity level can be modeled as a sparse GP. By Eq. (2) we can see that the conditional mean depends on the training points $\boldsymbol{X}_0$. To improve scalability, a sparse approximation can be introduced via a corrected Nyström method [9], which approximates the full covariance matrix while preserving its diagonal structure. Let $\boldsymbol{Z}_0 = (\boldsymbol{z}_{0,i})_{1\le i\le M_0}$ be the set of inducing points for the lowest fidelity, such that $M_0 \ll N_0$. We define the inducing variables as $\boldsymbol{u}_0 = f_0(\boldsymbol{Z}_0)$. As $\boldsymbol{u}_0 \in \mathbb{R}^{M_0}$ is a vector composed of evaluations of $f_0(\cdot)$ at inducing points $\boldsymbol{Z}_0$, then it is Gaussian distributed. The corrected Nyström approximation of the covariance matrix is defined as:

$$\widetilde{Q}_0 = Q_0 + \mathrm{diag}(\boldsymbol{K}_0 - Q_0), \tag{4}$$

where $Q_0 = \boldsymbol{K}_{f_0,u_0}\boldsymbol{K}_{u_0,u_0}^{-1}\boldsymbol{K}_{u_0,f_0}$ represents the low-rank (Nyström) approximation. Note that here we write the indices in the covariance matrices to distinguish whether the covariance is computed between the inputs $\boldsymbol{X}_0$ and the inducing points $\boldsymbol{Z}_0$ used for the approximation. More precisely, $\boldsymbol{K}_{f_0,u_0} = k_{\gamma_0}(\boldsymbol{X}_0, \boldsymbol{Z}_0)$, $\boldsymbol{K}_{u_0,u_0} = k_{\gamma_0}(\boldsymbol{Z}_0, \boldsymbol{Z}_0)$ and $\boldsymbol{K}_{u_0,f_0} = k_{\gamma_0}(\boldsymbol{Z}_0, \boldsymbol{X}_0)$.

Covariance parameter estimation. Now, we want to estimate the lowest fidelity level by using the approximation given by $\widetilde{Q}_0$. As a preliminary step to estimate the covariance parameters, we formulate the joint distribution of the latent process $\boldsymbol{f}_0$ (Eq. (1)) representing, in this case, the approximation

applied over the lowest fidelity level, using the corrected Nyström approximation, which serves as a computationally tractable surrogate for the full covariance structure [9]:

$$p(\boldsymbol{f}_0) \sim \mathcal{N}(\mathbf{0}, \widetilde{Q}_0), \tag{5}$$

then the posterior $p(\boldsymbol{y}_0|\boldsymbol{f}_0)$ is computed like in standard GPs, as $p(\boldsymbol{y}_0|\boldsymbol{f}_0) = \mathcal{N}(\boldsymbol{f}_0, \boldsymbol{T}_0)$ where $\boldsymbol{T}_0$ indexes the noise variances. To compute the marginal likelihood, we integrate out $\boldsymbol{f}_0$:

$$p(\boldsymbol{y}_0) = \int p(\boldsymbol{y}_0|\boldsymbol{f}_0)\, p(\boldsymbol{f}_0)\, d\boldsymbol{f}_0. \tag{6}$$

Since both $p(\boldsymbol{y}_0|\boldsymbol{f}_0)$ and $p(\boldsymbol{f}_0)$ are Gaussian, then the marginal likelihood is also Gaussian:

$$p(\boldsymbol{y}_0) = \mathcal{N}(\mathbf{0}, \widetilde{Q}_0 + \boldsymbol{T}_0). \tag{7}$$

By taking the logarithm of the marginal distribution, we obtain the marginal log-likelihood for the lowest fidelity level $y_0(\cdot)$ under the FITC framework:

$$\mathcal{L}(\boldsymbol{\theta}_0) = -\frac{1}{2}\log|\widetilde{Q}_{\boldsymbol{\theta}_0} + \boldsymbol{T}_{\boldsymbol{\theta}_0}| - \frac{1}{2}\boldsymbol{y}_0^\top \left(\widetilde{Q}_{\boldsymbol{\theta}_0} + \boldsymbol{T}_{\boldsymbol{\theta}_0}\right)^{-1} \boldsymbol{y}_0 - \frac{N}{2}\log(2\pi). \tag{8}$$

We can observe that $\mathcal{L}(\boldsymbol{\theta}_0)$ now depends on the inversion of $\widetilde{Q}$,[1] which can be computed more efficiently using the matrix inversion lemma (Woodbury matrix identity) [12]:

$$\begin{aligned}
(\widetilde{Q} + \boldsymbol{T})^{-1} &= \left(\boldsymbol{K}_{\boldsymbol{u}_0,\boldsymbol{f}_0} \boldsymbol{K}_{\boldsymbol{u}_0,\boldsymbol{u}_0}^{-1} \boldsymbol{K}_{\boldsymbol{u}_0,\boldsymbol{f}_0} + \underbrace{\operatorname{diag}(\boldsymbol{K}_{\boldsymbol{f}_0,\boldsymbol{f}_0} - Q_0 + \boldsymbol{T}_0)}_{\Lambda} \right)^{-1} \\
&= (\boldsymbol{K}_{\boldsymbol{f}_0,\boldsymbol{u}_0} \boldsymbol{K}_{\boldsymbol{u}_0,\boldsymbol{u}_0}^{-1} \boldsymbol{K}_{\boldsymbol{u}_0,\boldsymbol{f}_0} + \boldsymbol{\Lambda})^{-1} \\
&= \boldsymbol{\Lambda}^{-1} - \boldsymbol{\Lambda}^{-1} \boldsymbol{K}_{\boldsymbol{f}_0,\boldsymbol{u}_0} \left(\boldsymbol{K}_{\boldsymbol{u}_0,\boldsymbol{u}_0} + \boldsymbol{K}_{\boldsymbol{u}_0,\boldsymbol{f}_0} \boldsymbol{\Lambda}^{-1} \boldsymbol{K}_{\boldsymbol{f}_0,\boldsymbol{u}_0}\right)^{-1} \boldsymbol{K}_{\boldsymbol{u}_0,\boldsymbol{f}_0} \boldsymbol{\Lambda}^{-1}.
\end{aligned} \tag{9}$$

This reduces the computation complexity as inverting this new approximated covariance matrix scales as $O\left(N_0 M_0^2\right)$. Here, $\mathcal{L}(\boldsymbol{\theta}_0)$ is obtained by modifying the GP prior, and so the optimization does not reliably approximate the exact GP model as we replace the exact GP with a Nyström approximation. However, this method has shown accurate performance in diverse regression tasks [11].

Conditioning to noisy data. Recall that the prior considered in FITC replaces the real covariance matrix $\boldsymbol{K}_{\boldsymbol{f}_0,\boldsymbol{f}_0}$ using the corrected Nyström approximation in Eq. (4). The same approximation is used to replace the cross-covariance between training points and test points. We consider prediction at a new input point $\boldsymbol{x} \in \mathcal{D}$. The joint distribution of $(\boldsymbol{f}_0, f_0(\boldsymbol{x}))$ is then given by

$$p(\boldsymbol{f}_0, f_0(\boldsymbol{x})) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \widetilde{Q}_0 & \boldsymbol{q}_0(\boldsymbol{x}) \\ \boldsymbol{q}_0^\top(\boldsymbol{x}) & k_0(\boldsymbol{x},\boldsymbol{x}) \end{bmatrix}\right), \tag{10}$$

where $\boldsymbol{q}_0(\boldsymbol{x}) = k_{\gamma_0}(\boldsymbol{x}, \boldsymbol{Z}_0) k_{\gamma_0}^{-1}(\boldsymbol{Z}_0, \boldsymbol{Z}_0) k_{\gamma_0}(\boldsymbol{Z}_0, \boldsymbol{X}_0)$. Note that $q_0(\boldsymbol{x}, \boldsymbol{X})$ serves as a proxy of $k_{\gamma_0}(\boldsymbol{x}, \boldsymbol{X})$.

When predicting $f_0(\cdot)$, we compute the conditional distribution $p(f_0(\boldsymbol{x})|\boldsymbol{y}_0)$, which is Gaussian with the following mean and variance:

$$\begin{aligned}
m_{c,0}(\boldsymbol{x}) &= \boldsymbol{q}_0^\top(\boldsymbol{x})(\widetilde{Q}_0 + \boldsymbol{T}_0)^{-1} \boldsymbol{y}_0, \\
v_{c,0}(\boldsymbol{x}) &= k_{\gamma_0}(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{q}_0(\boldsymbol{x})^\top (\widetilde{Q}_0 + \boldsymbol{T}_0)^{-1} \boldsymbol{q}_0(\boldsymbol{x}).
\end{aligned} \tag{11}$$

Observe that the predictive mean is obtained by projecting the observations through the inducing variables, while the predictive variance retains the exact prior variance $k_{\gamma_0}(\boldsymbol{x},\boldsymbol{x})$ but subtracts the explained component through the inducing space. Finally, the expressions in Eq. (11) can be plugged into the recursive formulas in Eq. (3) for $\kappa = 1$, yielding an MFK framework that incorporates sparsity at the lowest fidelity level.

[1] The change in notations for $\boldsymbol{T} \to \boldsymbol{T}_\theta$ and, $\widetilde{Q} \to \widetilde{Q}_{\boldsymbol{\theta}}$ are used to show the dependency on the hyper-parameters $\boldsymbol{\theta}$. For simplicity of notation, the subscript is omitted in subsequent developments.

Remark. A similar procedure can be applied to induce sparsity at higher fidelity levels (i.e., for $\kappa \geq 1$). In this work, we focus only on the lowest fidelity level, as the challenge of handling large datasets typically comes from the lowest-fidelity computer code, as illustrated in Section 4.2. We also note that other sparse methods based on low-rank approximations, e.g., the subset of regressors approximation and the deterministic training conditional [9], can also be adapted to the multi-fidelity setting in a similar fashion. However, in this work, we focus on FITC, as it preserves the variances of the exact conditional while maintaining independence between training points, which is essential for computational tractability [12].

4 Results

In this section, we present the results of the models proposed in Sections 2 and 3, which are referred to as MFK and SMFK, respectively. Our objective is to evaluate how these models perform across different dimensional settings with two-fidelity scenarios ($L = 2$). In particular, we assess the proposed model in a complex three-dimensional (3D) Large Reference Model (LRM), leveraging both WT and CFD data. This experiment includes a two-fidelity configuration based on a nested Design of Experiments (DoE) with 250 WT experiments and 1949 CFD simulations introduced in [5].

In the following, we denote (by convention) the lowest fidelity (LF) and highest fidelity (HF) levels as $y_{LF}(\cdot) := y_0(\cdot)$ and $y_{HF}(\cdot) := y_1(\cdot)$, respectively. Thus, N_{LF} and N_{HF} correspond to the number of observations at the LF and HF levels. Here, M_{LF} denotes the inducing points of SMFK, whose locations are estimated using the k-means algorithm [13]. In our experiments, Python codes related to MFK and SMFK have been developed and are available on SMT [14]. The source codes are in the GitHub repository: `https://github.com/SMTorg/smt`.

4.1 Illustrative example using Forrester functions

In aerodynamic experiments such as wind tunnel tests, measurement noise is rarely constant across all flow conditions. Factors such as varying Reynolds number, turbulence intensity, sensor resolution, or structural vibrations introduce heteroscedastic noise, that is, noise whose variance depends on the operating conditions. To illustrate this phenomenon in a simplified setting, we consider a one-dimensional experiment based on the well-known Forrester function [15].

Let $f(x)$ denote the deterministic functions. We define the two fidelity levels as:

$$f_{LF}(x) = 0.5\,(6x-2)^2 \sin(12x-4) + 10(x-0.5) - 5, \qquad \text{(LF)}$$

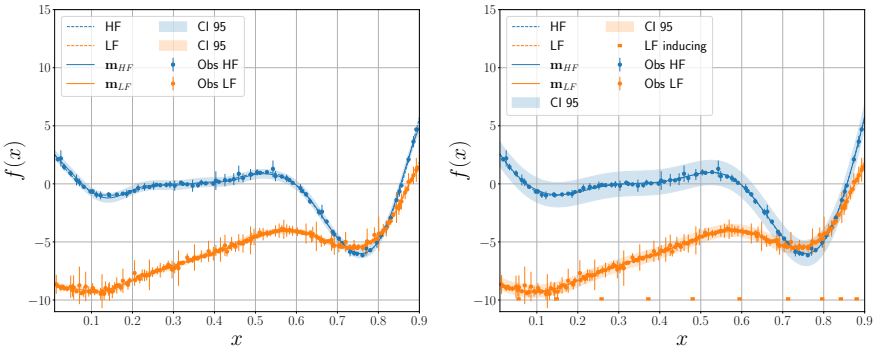
$$f_{HF}(x) = (6x-2)^2 \sin(12x-4), \qquad \text{(HF)}$$

which serves as a proxy for a generic aerodynamic response (e.g., lift coefficient as a function of angle of attack). We simulate noisy observations for the latent functions f, adding the noise term,

$$y_i = f(x_i) + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}\big(0, \tau_i^2\big), \tag{12}$$

where the noise variance τ_i^2 varies with x_i, introducing heteroscedasticity. Hence, the signal-to-noise ratio is not constant across the domain, which reflects the nonuniform uncertainty patterns typically observed in aerodynamic measurements [16]. Figure 1 shows a comparison between MFK and SMFK in a two-fidelity-level scenario assuming heteroscedasticity in both fidelity levels. There, an experiment with 80 HF observations and 160 LF observations has been suggested. For the SMFK model, we only consider 10 inducing points for the LF. We can notice from Figure 1 narrower confidence intervals in MFK. These trends are in agreement with the comparative analysis of sparse approximations since fewer inducing points provide a coarser approximation of the posterior, leading to higher predictive variance [11].

Finally, Table 1 presents the results obtained over 10 independent simulations. As observed, the MFK model appears more sensitive to the heteroscedastic noise in the data. In contrast, the sparse GP approximation demonstrates greater robustness to noisy observations. This behavior arises because the reduced representational capacity and the formulation of sparse models act as implicit regularizers, helping to prevent overfitting [9]. In comparison, full GP models can overfit the data when noise levels are high or non-Gaussian, as the marginal likelihood tends to fit individual observations too closely.



(a) MFK prediction (b) SMFK predictions

Figure 1: Example predictions for MFK (a) and SMFK (b). The panel shows the LF (orange) and HF (blue) predictions, with observations for each fidelity level indicated as dots. Vertical lines denote ± 3 standard deviations, and the shaded regions represent 95% confidence intervals.

Table 1: Root Mean square Error (RMSE) and training time for MFK and SMFK after 10 runs in a two-fidelity Forrester function with heteroscedastic noise.

MFK		SMFK	
RMSE	Training time (s)	RMSE	Training time (s)
0.128±0.011	0.439±0.121	0.057±0.006	0.243±0.042

4.2 Multi-fidelity LRM database

The dataset used in this study consists of WT experiments conducted at ONERA WT department, and CFD simulations [17]. Tests for the WT are performed in two different ONERA WT centers: the F1-WT, focusing on high Reynolds numbers under low subsonic flow, and the S1-WT, focuses on transonic speeds. The CRM shape is used to build two Large Reference Models (LRMs): the first one for S1-WT with a wing span of 3.5m (scale 1/16.835) and the second one for F1-WT with a wing span of 3m (scale 1/19.5). The CFD data is acquired using Reynolds-averaged Navier-Stokes equations. Here, the equation Spalart-Allmaras turbulence model is used. This CFD data uses a mesh named L1, composed of 63.9×10^3 cells. The CFD data have been taken in a parametric space larger than that of WT data to stretch out the study to the entire flight aircraft envelope. This highlights how the two datasets complement each other: one is more realistic but limited in scope, while the other expands that knowledge. This feature is evident in the reduced aerodynamic dataset from [5].

In this experiment, the multi-fidelity dataset consists of three input variables: the Mach number, M, defined as the ratio between the flow velocity and the local speed of sound, which characterizes compressibility effects in aerodynamics; the Reynolds number, Re, a dimensionless quantity representing the ratio of inertial to viscous forces in the flow, thereby governing the transition between laminar and turbulent regimes; and the angle of attack, α, which denotes the inclination of the airfoil or body relative to the incoming flow direction and strongly influences lift and drag generation and three aerodynamic coefficients are considered as outputs: the drag coefficient, C_x, which quantifies the nondimensionalized resistance force acting opposite to the flow direction; the lift coefficient, C_z, which measures the nondimensionalized force component perpendicular to the flow; and the pitching moment coefficient, C_M, which represents the nondimensionalized moment about a reference point, typically the aerodynamic center, and characterizes the longitudinal stability of the configuration. These input–output variables collectively define the aerodynamic response surface to be explored and modeled.

Since the nested configuration of the dataset is compatible with the recursive MFK model described in Section 2, we then propose comparing the performance of both MFK and SMFK. For SMFK, FITC approximations are considered. Model evaluation follows the procedure outlined in [5], using a WT validation set consisting of 1000 test samples. In this experiment, the influence of the number of inducing points is explored until reaching convergence [18]. Heteroscedastic noises are estimated for each fidelity level as detailed in [5].

Table 2: Analysis of the RMSE and training times for MFK and SMFK over 10 different random DoEs, $N_{HF} = 250$. Note that MFK does not use inducing points M. Lower RMSE and training time values are highlighted in bold.

Output	M_{LF}	RMSE		Training time (s)	
		MFK	SMFK	MFK	SMFK
C_x	5	0.006±0.004	0.086±0.014	279.187±10.254	**2.687±0.011**
	50		0.006±0.002		10.312±1.245
	100		**0.005±0.002**		37.343±2.132
C_z	5	0.013±0.006	0.045±0.026	475.421±12.355	**2.828±0.154**
	50		0.013±0.003		12.921±1.874
	100		**0.012±0.002**		48.234±3.451
C_M	5	**0.008±0.003**	0.014±0.004	273.406±12.147	**2.593±0.818**
	50		0.009±0.003		12.937±1.694
	100		0.009±0.002		31.843±3.141

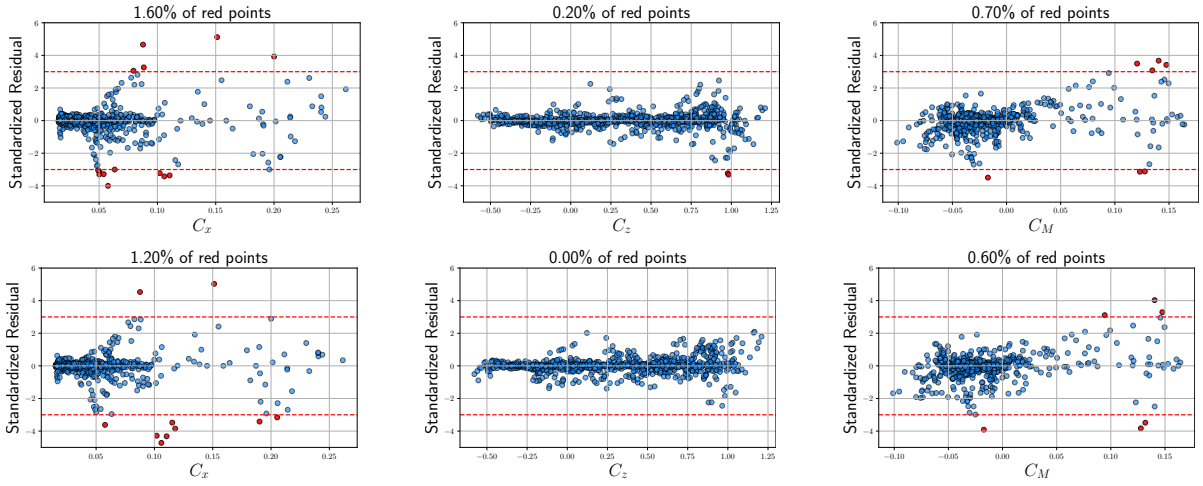


Figure 2: Standardized residuals for C_x, C_z and C_M coefficients (first, middle and last column, respectively) over 1000 test WT points. Results are shown for (top) the MFK model, and (bottom) the SMFK model with $M_{LF} = 100$. Residuals falling outside the $[-3,3]$ range are highlighted in red.

Table 2 compares the RMSE and training times of MFK and SMFK for the three outputs available in the LRM database over 10 random DoEs with $N_{HF} = 250$. MFK achieves slightly lower RMSE than SMFK at small M_{LF}, but SMFK approaches MFK accuracy as M_{LF} increases. Importantly, SMFK is significantly faster across all cases, highlighting its computational efficiency while maintaining comparable accuracy for larger low-fidelity datasets. Finally, Figure 2 shows the comparison of the outputs C_x, C_z, C_M, showing a significant improvement in terms of outliers in the standardized residuals for SMFK.

5 Conclusion

This study demonstrates that incorporating sparse approximations into the recursive multi-fidelity model of Section 2 provides a substantial enhancement in both computational efficiency and predictive robustness, particularly in the presence of heteroscedastic noise. The use of inducing variables not only reduces the cubic computational complexity traditionally associated with GP but also introduces a natural smoothing effect on the conditional mappings between fidelity levels. This smoothness is especially beneficial when the observational noise varies across the input domain, as it mitigates local overfitting and ensures more stable uncertainty propagation through the recursive hierarchy.

By integrating sparse approximations in the lowest fidelity level, the model achieves improved noise

regularization and a more coherent representation of cross-fidelity correlations. The results obtained in our experiments indicate a consistent improvement as reflected by comparable RMSE values and lower training times compared to the standard recursive MFK formulation.

Overall, the proposed sparse recursive formulation provides a principled and scalable alternative to the classical recursive MFK. It enables efficient learning from large and noisy datasets while preserving the interpretability and probabilistic rigor of the original recursive formulation. Future work will extend this framework to generalize sparse approximations to all fidelity levels but for non-nested DoEs, further expanding its applicability to complex real-world systems.

Acknowledgement. This research is part of the ECOS Nord project C25M01. Part of the work was conducted while AFLL was affiliated with CERAMATHS, UPHF. We thank R. Lafage (ONERA) for his support in integrating the Python implementation of the SMFK model into the toolbox SMT.

References

[1] F. Johnson, E. Tinoco, and N. Yu. 30 years of development and application of CFD at boeing commercial airplanes, seattle. *Computers & Fluids*, 34:1115–1151, 2005.

[2] M. Malik and D. Bushnell. Role of computational fluid dynamics and wind tunnels in aeronautics. Technical report, NASA Technical Report TP-2012-217602, 2012.

[3] M. Meliani, N. Bartoli, T. Lefebvre, M.-A. Bouhlel, J. Martins, and J. Morlier. Multi-fidelity efficient global optimization: Methodology and application to airfoil shape design. In *AIAA Aviation 2019 Forum*, pages 1–18, Dallas, United States, June 2019.

[4] L. Le Gratiet. *Multi-fidelity Gaussian Process Regression for Computer Experiments*. PhD thesis, Université Paris-Diderot-Paris VII, 2013.

[5] R. Conde Arenzana, A. F. López-Lopera, S. Mouton, N. Bartoli, and T. Lefebvre. Multi-fidelity Gaussian process model for CFD and wind tunnel data fusion. In *Aerobest*, 2021.

[6] D. J. J. Toal. Applications of multi-fidelity multi-output Kriging to engineering design optimization. *Structural and Multidisciplinary Optimization*, 2023.

[7] 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.

[8] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. The MIT Press, 2005.

[9] J. Q. Quinonero-Candela and C. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, 6:1939–1959, 2005.

[10] 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.

[11] E. Snelson and Z. Ghahramani. Sparse Gaussian processes using pseudo-inputs. In *Advances in Neural Information Processing Systems 18*, pages 1257–1264, 2005.

[12] M. van der Wilk. *Sparse Gaussian Process Approximations and Applications*. PhD thesis, University of Cambridge, 2019.

[13] H. Valayer, N. Bartoli, M. Castaño-Aguirre, R. Lafage, T. Lefebvre, A. F. López-Lopera, and S. Mouton. A python toolbox for data-driven aerodynamic modeling using sparse Gaussian processes. *Aerospace*, 11(4), 2024.

[14] P. Saves, R. Lafage, N. Bartoli, Y. Diouane, J. Bussemaker, T. Lefebvre, J. T. Hwang, J. Morlier, and J. R. R. A. Martins. SMT 2.0: A surrogate modeling toolbox with a focus on hierarchical and mixed variables Gaussian processes. *Advances in Engineering Software*, 188:103571, 2024.

[15] A. I. J. Forrester, A. Sobester, and A. J. Keane. *Engineering Design via Surrogates*. Wiley, 2007.

[16] S. Narayan and R. Bose. The uncertainty of the experimentally-measured momentum coefficient. *Experiments in Fluids*, 2024.

[17] L. Cambier, S. Heib, and S. Plot. The ONERA elsA CFD software: Input from research and feedback from industry. *Mechanics & Industry*, 14:159–174, 2013.

[18] D. R. Burt, C. E. Rasmussen, and M. van der Wilk. Rates of convergence for sparse variational Gaussian process regression. In *International Conference on Machine Learning*, volume 97, pages 862–871, 2019.