

Multi-fidelity Gaussian Process for Uncertainty Quantification in Aerodynamic Analysis

Xin Chen, Guangyuan Huang, Ankit Sharma, Atif Riaz, Sergio Jimeno
Cranfield University, Bedford, Bedfordshire, MK43 0AL, United Kingdom

This research contributes to the AIAA FD UQ Discussion Group's Challenge Problem, with an aim to evaluate aerodynamic coefficients and their total uncertainty for an NACA 2412 airfoil. In this research, the problem is tackled with Multi-fidelity Gaussian Process (MFGP), where experimental data and XFOIL predictions are considered as two levels of fidelity. The total uncertainty is quantified as the MFGP's predicted variance. Three training approaches are implemented and tested repeatedly. According to the results, MFGP can effectively capture the nominal values of the aerodynamic coefficients with very sparse experimental data. However, the total uncertainty is underestimated. This is potentially due to the limitation of the maximum likelihood method for estimating the MFGP's hyperparameters.

I. Nomenclature

α	=	angle of attack
δ_f	=	flap deflection angle
ε	=	noise term
σ	=	standard deviation
c_l, c_d, c_m	=	lift, drag, and moment coefficient
$E[\cdot]$	=	expectation
f, ρ, δ	=	functions
$\mathcal{GP}$	=	Gaussian Process
$Mach$	=	Mach Number
$\mathcal{N}$	=	Gaussian Distribution
Re	=	Reynolds Number
$\mathbf{X}, \mathbf{Y}$	=	collection of input vectors
$\{\cdot\}_H$	=	high fidelity model
$\{\cdot\}_L$	=	low fidelity model
$\{\cdot\}^*$	=	new test/prediction

II. Introduction

Aircraft design relies extensively on both physical experiments and computational simulations. In general, experiments and models with higher fidelities can provide more accurate predictions, but at the cost of more experimental/computational resources. As a result, these experiments/models are rarely used for design space exploration and can only be applied to a limited number of design solutions. Conversely, low-fidelity models are relatively inexpensive to evaluate. Therefore, they can be employed to explore a much wider design space at an early stage. However, they may not have sufficient accuracy to support critical decision-making. Within this context, the importance of Uncertainty Quantification and Management (UQ&M) has been highlighted, and a critical challenge is to effectively aggregate different sources of information while maintaining a balance between computational cost and accuracy.

The research work presented in this manuscript contributes to the second Uncertainty Quantification Challenge Problem for Aerodynamics proposed by the AIAA FD UQ Discussion Group [1]. In this challenge, a simplified airfoil testcase is specified and the task is to evaluate the total uncertainty associated with predictions at a set requested points, using data from both experiments and low-fidelity simulations.

To solve this problem, a surrogate-based approach is adopted using the Multi-fidelity Gaussian Process (MFGP). The remaining part of this manuscript is structured as follows: Section III briefly reviews the UQ Challenge Problem. Section IV reviews the background on MFGP and explains its implementation in this research. The results are presented in Section V. Main conclusions and future work are given in Section VI.

Table 1. Experimental data at different angles of attack and flap deflection angles [1]

α [deg]	δ_f [deg]	Repeat Run	c_l	c_d	c_m
0	0	A	0.2142712	0.0120677	-0.0474733
0	0	B	0.2133173	0.0120715	-0.0474752
0	0	C	0.2153800	0.0120713	-0.0474664
0	0	D	0.2141752	0.0120712	-0.0474717
0	0	E	0.2146903	0.0120653	-0.0474728
5	0	A	0.7371244	0.0142539	-0.0426859
5	0	B	0.7380711	0.0142689	-0.0426651
5	0	C	0.7365440	0.0142556	-0.0426878
5	0	D	0.7356163	0.0142534	-0.0426958
5	0	E	0.7372738	0.0142587	-0.0426802
10	0	A	1.2090240	0.0205511	-0.0328444
10	0	B	1.2073470	0.0205052	-0.0329239
10	0	C	1.2083610	0.0205293	-0.0328803
10	0	D	1.2086720	0.0205479	-0.0328533
10	0	E	1.2092780	0.0205541	-0.0328371
0	5	A	0.5357481	0.0131460	-0.0942028
0	5	B	0.5365244	0.0131518	-0.0941909
0	5	C	0.5352645	0.0131419	-0.0942108
0	5	D	0.5362396	0.0131439	-0.0942029
0	5	E	0.5378183	0.0131561	-0.0941781
0	10	A	0.8358800	0.0157307	-0.1366514
0	10	B	0.8350389	0.0157214	-0.1366768
0	10	C	0.8357959	0.0157262	-0.1366615
0	10	D	0.8373402	0.0157373	-0.1366275
0	10	E	0.8364706	0.0157245	-0.1366605
5	5	A	1.0426160	0.0176421	-0.0870176
5	5	B	1.0411650	0.0176252	-0.0870579
5	5	C	1.0403830	0.0176236	-0.0870677
5	5	D	1.0417220	0.0176346	-0.0870377
5	5	E	1.0408130	0.0176238	-0.0870634
-5	-5	A	-0.6449671	0.0147648	-0.0036884
-5	-5	B	-0.6442967	0.0147558	-0.0036745
-5	-5	C	-0.6454917	0.0147620	-0.0036932
-5	-5	D	-0.6445543	0.0147589	-0.0036797
-5	-5	E	-0.6433426	0.0147437	-0.0036550

III. Problem Statement

The UQ Challenge Problem has been presented in [1] and is briefly reviewed here for essential information to conduct this research. Considering an NACA 2412 airfoil in incompressible flow at a chord Reynolds number of 700,000 and with a flap defined at 30% of its chordwise location, experimental data are provided for the airfoil's coefficients of lift (c_l), drag (c_d), and quarter-chord pitching moment (c_m) at different angles of attack (α) and flap deflection angles (δ_f) as listed in Table 1. For each combination of α and δ_f , five repeated measurements are given,

and the difference between these repeated measurements are caused by precision of $\pm 0.02^\circ$ in the angle of attack, $\pm 0.1^\circ$ in the flap angle, $\pm 0.5\%$ in nominal Reynolds number, and measurement uncertainty (0.008 for c_l and c_m , 0.0008 for c_d).

Apart from the experimental data, a special version (v6.99uq) of XFOIL [2] has been provided. The target of this UQ challenge is to use the experimental data and XFOIL for prediction of the airfoil's aerodynamic coefficients and the associated total uncertainty (which includes the model, experiment, numerical uncertainty, etc.) at four requested conditions as listed in Table 1. Among these conditions, the last three are new conditions which are not included in the experimental data. For XFOIL execution, 100 panels on the airfoil shall be used to predict the results, while fully turbulent conditions are simulated by employing a low critical amplification ratio of 0.01.

Table 2. Requested conditions

α [deg]	δ_f [deg]
0.0	0.0
2.0	2.0
7.0	15.0
-3.0	-3.0

IV. Methodology

In this research, the problem is tackled with a surrogate-based approach using the Multi-fidelity Gaussian Process (MFGP), because of its abilities to aggregate information from multiple sources and to provide the variance for new predictions. Similar approach can also be found in existing research work [3]. In the remaining part of this section, the basics of GP and MFGP are introduced in Sections IV.A and IV.B, respectively. The specific implementation of MFGP for this research is explained in Sections IV.C and IV.D.

A. Gaussian Process

A Gaussian Process (GP) can be considered as a distribution over functions. For any collection of input vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, the corresponding function values $f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)$ has a joint multivariate Gaussian distribution, which can be noted as:

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

where $m(\mathbf{x})$ is a mean function and $k(\mathbf{x}, \mathbf{x}')$ is a covariance function.

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

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

where $E[\cdot]$ refers to the expectation.

In practice, the covariance function is specified, and a typical choice is the square exponential function:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_d \frac{(x_d - x'_d)^2}{l_d^2}\right) \quad (4)$$

where x_d and x'_d are the d_{th} component of two D -dimensional input vectors ($\mathbf{x}$ and $\mathbf{x}'$), respectively. σ_f and l_d are hyperparameters of the covariance function.

Given observations $\{\mathbf{X}, \mathbf{Y}\}$ at n locations:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \quad (5)$$

$$\mathbf{Y} = [y_1, y_2, \dots, y_n] \quad (6)$$

where

$$y_i = f(\mathbf{x}_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (7)$$

ε_i is a noise term which is assumed to follow an independent Gaussian distribution with variance σ_i^2 (which should be distinguished from σ_f in Eq (4) and can vary across the observations). In practice, σ_i , σ_f and l_d can be estimated with a maximum likelihood approach.

At any new test points $\mathbf{X}^*$, the conditional distribution of the predicted output is also Gaussian:

$$f(\mathbf{X}^*)|\mathbf{X}, \mathbf{Y} \sim \mathcal{N}(\bar{\mathbf{Y}}^*, \text{cov}(\mathbf{Y}^*)) \quad (8)$$

where $\bar{\mathbf{Y}}^*$ and $\text{cov}(\mathbf{Y}^*)$ are the mean and covariance of the predicted outputs, respectively

$$\bar{\mathbf{Y}}^* = k(\mathbf{X}^*, \mathbf{X})[k(\mathbf{X}, \mathbf{X}) + \Delta]^{-1}\mathbf{Y} \quad (9)$$

$$\text{cov}(\mathbf{Y}^*) = k(\mathbf{X}^*, \mathbf{X}^*) - k(\mathbf{X}^*, \mathbf{X})[k(\mathbf{X}, \mathbf{X}) + \Delta]^{-1}k(\mathbf{X}, \mathbf{X}^*) \quad (10)$$

Δ is a diagonal matrix consisting of the variances σ_i^2 of the observation noises ε_i as defined in Eq (7):

$$\Delta = \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix} \quad (11)$$

In this formulation, $\text{cov}(\mathbf{Y}^*)$ can be considered as a measure of uncertainty associated with the new prediction $\mathbf{Y}^*$. The reader is referred to [4] for more information on GP.

B. Multi-fidelity Gaussian Process

The MFGP employs multiple GPs in a recursive manner to integrate data from different sources. The main philosophy is to use the dense and inexpensive low-fidelity data for capturing the general trend and to use the expensive and sparse high-fidelity data for calibration. Consider two simulation models (f_L and f_H) with low and high fidelities, respectively. Using the autoregressive formulation [5], the high-fidelity model can be expressed as:

$$f_H(\mathbf{x}) = \rho(\mathbf{x})f_L(\mathbf{x}) + \delta(\mathbf{x}) \quad (12)$$

where $\rho(\mathbf{x})$ is a scaling parameter that captures the correlation between fidelities, $\delta(\mathbf{x})$ is a discrepancy function that represents the difference between the two models beyond scaling. In the MFGP approach, $f_L(\mathbf{x})$ and $\delta(\mathbf{x})$ are modelled by two independent Gaussian Processes.

$$f_L(\mathbf{x}) \sim \mathcal{GP}(0, k_l(\mathbf{x}, \mathbf{x}')) \quad (13)$$

$$\delta(\mathbf{x}) \sim \mathcal{GP}(0, k_\delta(\mathbf{x}, \mathbf{x}')) \quad (14)$$

where $k_l(\mathbf{x}, \mathbf{x}')$ and $k_\delta(\mathbf{x}, \mathbf{x}')$ are the corresponding covariance functions. Given training data from both models

$$y_{L,i} = f_L(\mathbf{x}_{L,i}) + \varepsilon_{L,i}, i = 1, 2, \dots, n \quad (15)$$

$$y_{H,j} = f_H(\mathbf{x}_{H,j}) + \varepsilon_{H,j}, j = 1, 2, \dots, m \quad (16)$$

outputs at new test points can be predicted in a similar manner as explained in Section IV.A. The reader is referred to [5, 6] for more information on MFGP.

C. MKGP for Integration of Experiment and XFOIL

In this research, three independent MFGPs are constructed in the input space of $\alpha \in [-6.0, 11.0] \times \delta_f \in [-6.0, 16.0]$ for c_l , c_d , and c_m , respectively.

The experiment is considered as a high-fidelity “simulation” f_H as defined by Eq (12), while the data as shown in Table 1 are considered as high-fidelity observations as defined by Eq (16). Uncertainty in the experimental data (including measurement, Re , α , and δ_f) is reflected in the variance of the repeated measures and considered as a lumped noise term at each experimental point, respectively. These noises are formulated as $\varepsilon_{H,j}$ in Eq. (16) and are estimated using maximum likelihood optimization.

XFOIL is considered as the low-fidelity function f_L , while its uncertainty (in model form, parameter, and numeric error) is also lumped and captured by the scalar term $\rho(\mathbf{x})$ and discrepancy term $\delta(\mathbf{x})$ in Eq. (12). As the XFOIL's predictions are deterministic, the noise term $\varepsilon_{L,i}$ should be zero.

D. Training Process

Three training approaches are applied to evaluate the accuracy and computational cost of MFGP. In all the approaches, the 7 experimental points (with 5 repeated measurements at each point) are utilized as the high-fidelity training set $\{\mathbf{X}_H, \mathbf{Y}_H\}$.

The low-fidelity training set $\{\mathbf{X}_L, \mathbf{Y}_L\}$ in the first approach is produced by execution of XFOIL at 1000 locations in the input space of α and δ_f all at once, using Latin Hypercube Sampling (LHS). In the second approach, the low-fidelity training set is produced in an iterative manner, as illustrated in Fig 1. The initial samples in the input space are the same as the experimental points ($\mathbf{X}_L^{(0)} = \mathbf{X}_H$), but with $\mathbf{Y}_L^{(0)}$ predicted by XFOIL instead of the experimental measurements. This is because the low-fidelity samples need to be included in the high-fidelity ones, as required by the MFGP formulation [6].

In each iteration, three independent MFGPs are updated for c_l , c_d , and c_m , respectively. Infilling points $\mathbf{X}^{(i)}$ are allocated at the locations with the highest predicted variance of each output (σ_Y^2) using a nested optimization process. If an infilling point is already inside the current set, LHS is used to replace the redundant point. XFOIL is then executed at these infilling points, and the outputs $\mathbf{Y}_L^{(i)}$ are used to update all three MFGPs. The training loop will stop if the total number of XFOIL executions reaches 1000. The third approach is similar to the second one, but with 50 additional points in the initial training set and 10 additional points for infilling in each iteration. All the additional points are produced with LHS. The setups of the three approaches are summarized in Table 3.

Table 3. Summary of Three Approaches for Training

Approach	Initial Samples $\mathbf{X}_L^{(0)}$	Iterative Infilling Samples $\mathbf{X}_L^{(i)}$
1	7 ($\mathbf{X}_H$) + 1000 (LHS)	0
2	7 ($\mathbf{X}_H$)	3 (Max σ_Y^2)
3	7 ($\mathbf{X}_H$) + 50 (LHS)	3 (Max σ_Y^2) + 10 (LHS)

Each approach is repeated 5 times to test their stability. The entire process is implemented in Python using an open-sourced package: Surrogate Modeling Toolbox (SMT) [7].

V. Results

Fig 2 illustrates the convergence histories of c_l , c_d , and c_m predictions against the number of XFOIL executions, using one of the five repeated tests of each approach. The input values of α and δ_f (as listed in Table 2) are distinguished by the colors, while the error bars indicate the standard deviations (total uncertainty) of the predicted value. The first, second, and third approaches are distinguished by the solid straight lines, semi-transparent polylines, and solid polylines, respectively.

It can be seen that most predictions stabilized quickly after the first 200 executions of XFOIL, except for the point at $\alpha = 7.0, \delta_f = 15.0$, especially in c_d predictions. This is expected because this point is outside the envelope of the provided experimental points in the input space, which requires extrapolation of the data. In addition, it also represents a highly non-linear condition with relatively large flow separation.

Fig 3 is a zoom-in view of the convergence histories at point $\alpha = 7.0, \delta_f = 15.0$. It can be seemed that both the second and third approaches are still oscillating even at the end of training (with 1000 XFOIL executions), while the third approach (solid polylines) is slightly more stable than the second one (semi-transparent polylines).

Fig 5 and Fig 6 show the convergence histories of the predicted means and standard deviations, respectively. The columns are corresponding to c_l , c_d , and c_m , while the rows are corresponding to different approaches. In each subplot, all the five repeated convergence histories are plotted superimposed. According to the figures, predictions across the repeated tests are very consistent except for the point $\alpha = 7.0, \delta_f = 15.0$. Similarly, the third approach is slightly more consistent than the second one, across the repeated tests. It can also be noted from Fig 6 that the predicted standard deviations for point $\alpha = 0.0, \delta_f = 0.0$ is the lowest. This is also expected as this point is included in the high-fidelity data for training. However, this value is even below the standard deviation of the raw data, which implies that the uncertainty might have been underestimated.

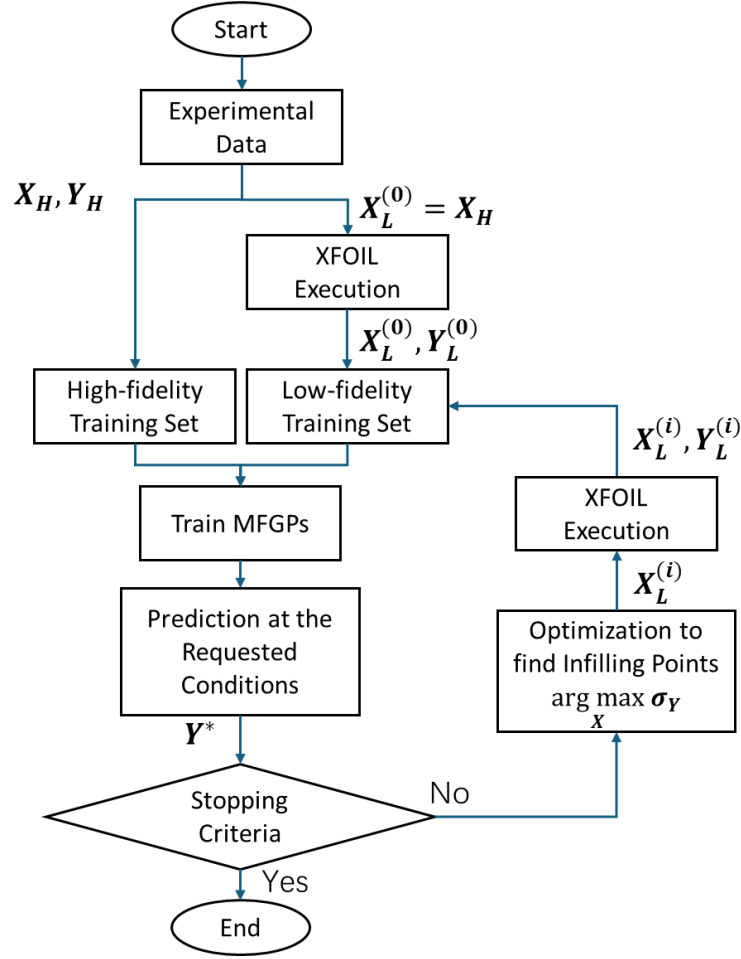


Fig 1. Iterative Training Process of MFGP.

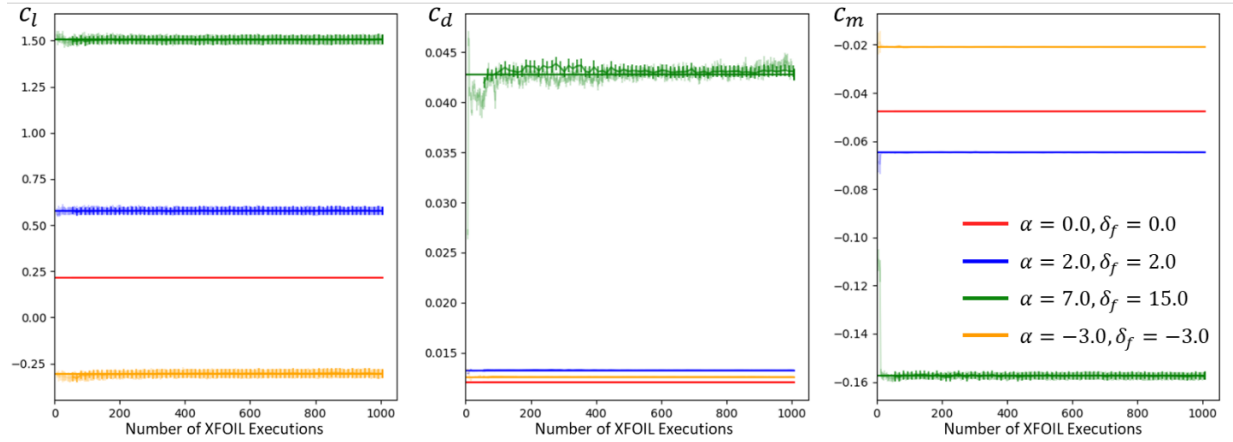


Fig 2. Convergence histories of aerodynamic coefficients at the requested conditions in each iteration

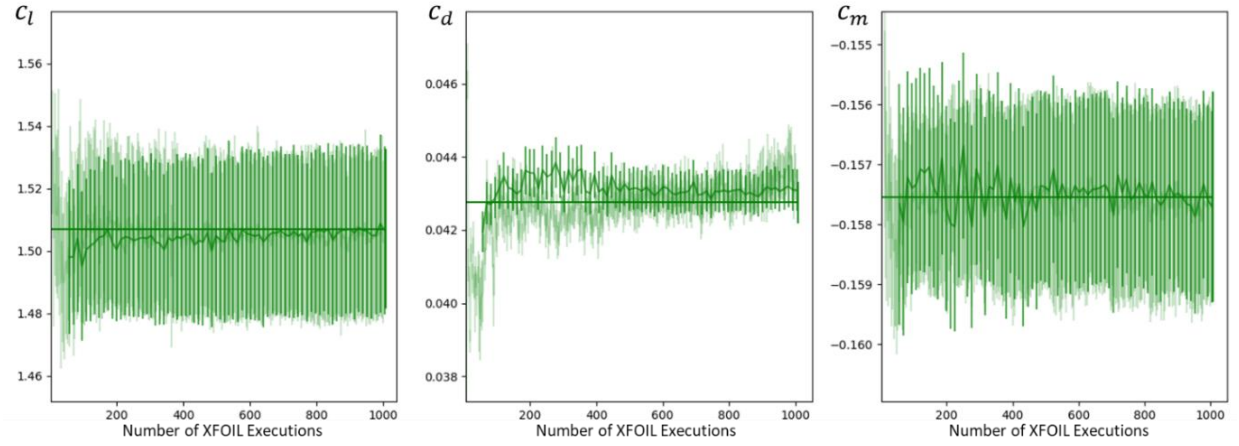


Fig 3. Zoom-in view of convergence histories for $\alpha = 7.0, \delta_f = 15.0$

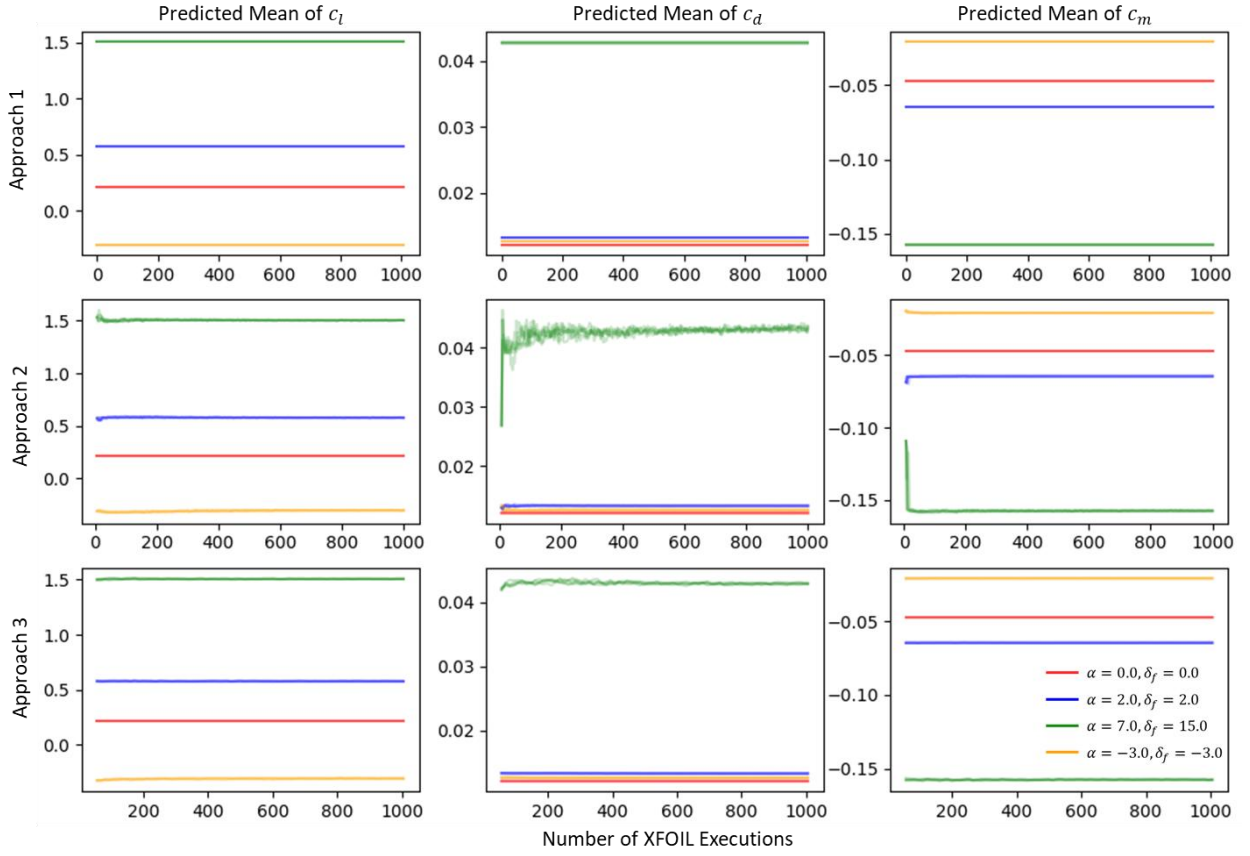


Fig 4. Convergence histories of predicted means against XFOIL execution

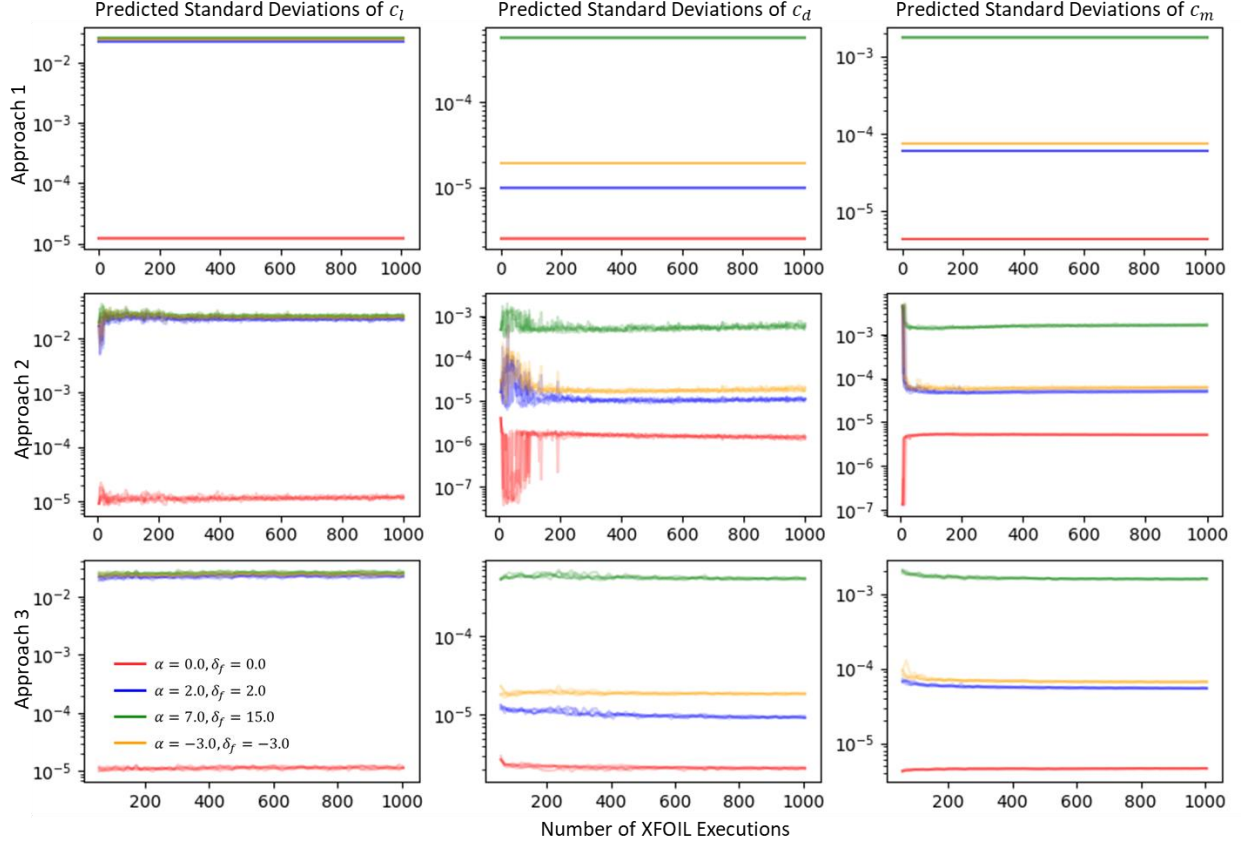


Fig 5. Convergence histories of predicted standard deviations against XFOIL execution

The MFGP predictions of c_l , c_d , and c_m in the entire design space (α and δ_f) with 1000 XFOIL executions are presented in Fig 6, Fig 7, and Fig 8. The subplot on the left compares the MFGP (in blue) with a reference single fidelity GP (in red) based only on the experimental data. The difference is not evident for c_l predictions due to its linearity. However, the MFGP and the single fidelity GP for c_d are dramatically different, where the latter fail to capture the change of c_d at large α and δ_f , due to lack of data.

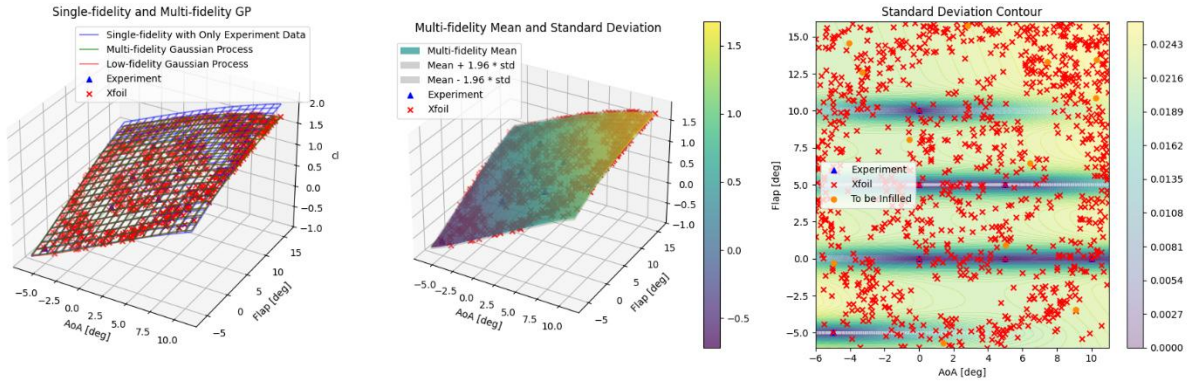


Fig 6. MFGP for lift coefficient with 1000 XFOIL executions: (a) comparison between MFGP and a single GP based only on the experimental data, (b) MFGP prediction with $\pm 1.96 \sigma$, (c) contour of predicted standard deviation in the entire design space

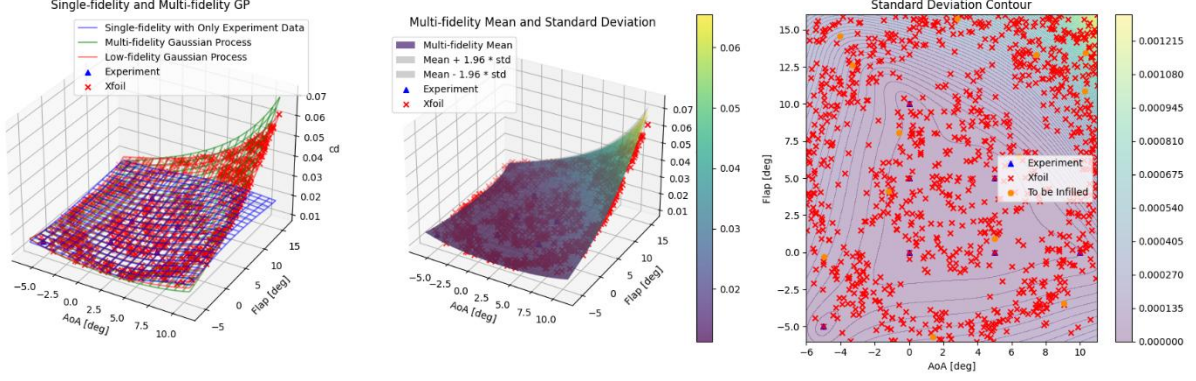


Fig 7. MFGP for drag coefficient with 1000 XFOIL executions: (a) comparison between MFGP and a single GP based only on the experimental data, (b) MFGP prediction with $\pm 1.96 \sigma$, (c) contour of predicted standard deviation in the entire design space

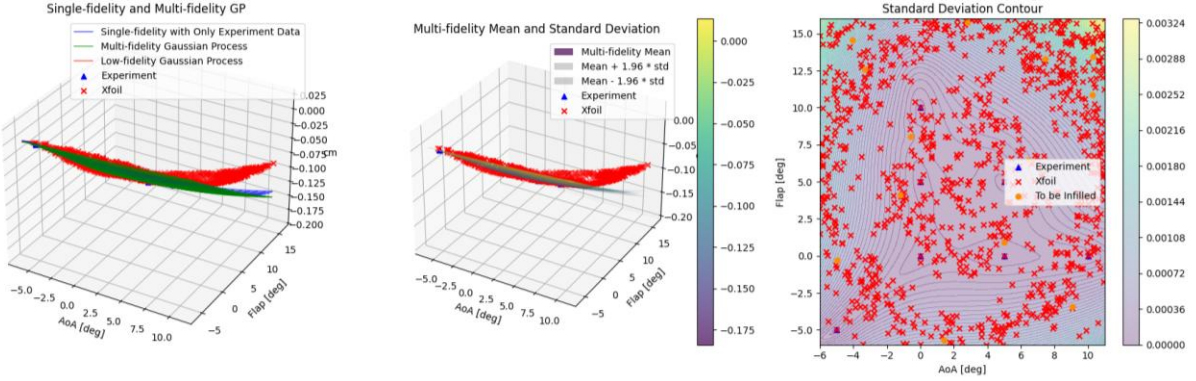


Fig 8. MFGP for moment coefficient with 1000 XFOIL executions: (a) comparison between MFGP and a single GP based only on the experimental data, (b) MFGP prediction with $\pm 1.96 \sigma$, (c) contour of predicted standard deviation in the entire design space

VI. Conclusion

This research contributes to an open UQ Challenge Problem which aims to quantify uncertainty in model predictions by integrating experimental data and low-fidelity simulations. The problem is tackled with a surrogate-based approach using the Multi-fidelity Gaussian Process, and the total uncertainty is quantified as the MFGP's predicted variance. Three training approaches are implemented and tested repeatedly to explore the balance between accuracy and computational cost.

According to the results, MFGP can effectively capture the nominal values of the aerodynamic coefficients with very sparse experimental data. Stable and consistent predictions can be produced with 200~300 XFOIL evaluations in an adaptive and iterative training approach. However, the total uncertainty (indicated by the predicated variances) seems to be underestimated. This is potentially due to the limitation of the maximum likelihood method for estimating the hyperparameters, where the uncertainty of the latter is not captured in the current setup. Future work will focus on the Bayesian approach for estimating the hyperparameters and to improve the stability of the training process.

Acknowledgments

The research leading to these results has received funding from the Innovate UK, Aerospace Technology Institute (ATI) in the UK, under the Out of Cycle NEXt generation highly efficient air transport (ONEheart) project (Ref no. 10003388).

References

- [1] AIAA FD UQ Discussion Group, “Second Uncertainty Quantification Challenge Problem for Aerodynamics”, URL: <https://engage.aiaa.org/viewdocument/challenge-problem-2-final-draft?CommunityKey=9807d37f-1a4e-4fd3-ab28-916b6756b09d&tab=librarydocuments&hlmlt=VT>
- [2] Drela, M., "XFOIL: An analysis and design system for low Reynolds number airfoils," In *Low Reynolds number aerodynamics: proceedings of the conference notre dame, Indiana, USA, 5–7 June 1989*, pp. 1-12. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989.
- [3] Meliani, M., Bartoli, N., Lefebvre, T., Bouhlel, M.A., Martins, J.R. and Morlier, J., “Multi-fidelity efficient global optimization: Methodology and application to airfoil shape design,” In *AIAA Aviation Forum*, AIAA Paper 2019-3236, June 2019.
- [4] Rasmussen, C. E., and Williams, C. K. I., *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.
- [5] Kennedy, M. C., & O'Hagan, A., “Bayesian calibration of computer models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 63, No. 3, 2001, pp. 425-464.
- [6] Le Gratiet, L., "Multi-fidelity Gaussian process regression for computer experiments," Ph.D. Dissertation, Université Paris-Diderot-Paris VII, Paris, 2013.
- [7] Saves, P., Lafage, R., Bartoli, N., Diouane, Y., Bussemaker, J., Lefebvre, T., Hwang, J.T., Morlier, J. and Martins, J.R., “SMT 2.0: A Surrogate Modeling Toolbox with a focus on hierarchical and mixed variables Gaussian processes,” *Advances in Engineering Software*, Vol. 188, 2024, p.103571.



DSPACE

<https://dspace.org/>

Multi-fidelity gaussian process for uncertainty quantification in aerodynamic analysis

Chen, Xin; Huang, Guangyuan; Sharma, Ankit; Riaz, Atif; Jimeno Altelarrea, Sergio

2026-01-12

American Institute of Aeronautics and Astronautics (AIAA)

<https://doi.org/10.2514/6.2026-0297>

Chen X, Huang G, Sharma A, et al., (2026) Multi-fidelity gaussian process for uncertainty quantification in aerodynamic analysis. In: AIAA SCITECH 2026 Forum, 12-16 January 2026, Orlando, USA, Article number 2026-0297