

HOW TO WRITE THESES
WITH TWO LINE TITLES

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF AERONAUTICS AND ASTRONAUTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jayant Mukhopadhaya
February 2021

Preface

This thesis tells you all you need to know about...

Acknowledgments

The authors would like to thank The Boeing Company for funding this research under grant number 131303 [IC2017-0559].

Contents

Preface	iv
Acknowledgments	v
1 Multi-Fidelity Gaussian Process Regression	1
1.1 Gaussian Process Regression	2
1.2 Multi-Fidelity Gaussian Process Regression	5
1.3 Application to NASA CRM	8
A A Long Proof	16

List of Tables

1.1 Number of data points of each fidelity that are used to create Figures 1.5-1.7. 9

List of Figures

1.1	GP regression mean and 2σ error estimates when using deterministic and uncertain input data for an underlying analytic function of interest.	4
1.2	Samples of the surrogate models that are candidate functions given the GP regression on the deterministic and the uncertain data.	5
1.3	Comparison of single- and two-fidelity GP regression.	7
1.4	Wall clock time comparison to train and query the multi-fidelity Gaussian Process formulations put forth by Kennedy and O’Hagan [6], and Gratiet [4].	8
1.5	C_L vs α for the NASA CRM, using data from multiple sources of varying fidelity.	10
1.6	C_D vs α for the NASA CRM, using data from multiple sources of varying fidelity.	10
1.7	C_m vs α for the NASA CRM, using data from multiple sources of varying fidelity.	10
1.8	Root Mean Square error when using three-fidelity data vs. using only high-fidelity data points.	11
1.9	Showcasing the superior predictive capability of multi-fidelity data fusion when high-fidelity data is localized in the design space. The left column represents the multi-fidelity GP formulation result, while the middle column shows the results for the single-fidelity formulation. The right column shows the RMSE trends when high-fidelity data is localized. The black circles correspond to the specific case shown in the left and middle columns.	12
1.10	Two-dimensional representation of C_L as a function of α and Mach number.	14
1.11	Root-mean-square-error for two-dimensional functions of Mach and α when using two-fidelity data vs. using only high-fidelity data points.	15

Chapter 1

Multi-Fidelity Gaussian Process Regression

Most computational or experimental analysis techniques provide discrete realizations of a quantity of interest (QoI) at discrete points in the domain of interest. Each analysis, referred to as a function evaluation, has a monetary and/or computational cost associated with it. In practice, higher-fidelity function evaluations are more costly than lower-fidelity ones. If the converse were true, there would be no reason to use a lower-fidelity analysis in place of a higher-fidelity one. Fidelity here refers to how closely an analysis technique mirrors real-life physics.

In theory, the discretization of the domain can be fine enough that it results in a nearly continuous representation of the QoI. Additionally, only the highest-fidelity analysis techniques would be used so that there is very little uncertainty or error in the QoI predictions. In reality, this is not tractable. Often cost minimization is a priority. This equates to QoIs being represented as sparsely as possible, and lower-fidelity analysis techniques replacing higher-fidelity ones wherever possible.

To squeeze the most out of the function evaluations that are available, numerous statistical methods have been developed that use these discrete realizations to create continuous representations of the QoI. These representations are called surrogate models and they assume that the underlying function of interest varies smoothly between the discrete data points. Popular surrogate modeling techniques include radial basis functions [13], Gaussian processes (GP) [7, 9, 14], stochastic collocation [8], and polynomial chaos expansions (PCE) [1, 11].

Of these, the use of multi-fidelity data has been developed for GP [4, 6] and, more recently, for PCE [10, 12]. Both methods handle the multi-fidelity data by using correction terms that are trained on the difference between the low-fidelity and high-fidelity data. GP have a slight advantage in this regard as they use an additive and a multiplicative term in the correction, whereas PCE only use an additive term. PCE have a distinct advantage in sensitivity analyses as Sobol indices can be directly post-processed [2, 17]. On the other

hand, GP can handle uncertain inputs and provide a direct estimation of the error in its modeling. This can be used for adaptive sampling techniques that suggest additional function evaluations to reduce the uncertainty in the model. For this work GP are the surrogate model of choice due to its advanced handling of multi-fidelity data that have associated uncertainties, and the direct estimation of uncertainty in its predictions.

This chapter first introduces the basic equations of GP regression that are used to handle single-fidelity data that can have uncertainties associated with it. It builds upon this by introducing multi-fidelity GP regression that combines data from different sources to build a single, superior surrogate model for the QoI. These multi-fidelity GP equations are then used to create probabilistic aerodynamic databases that represent the performance characteristics of a model aircraft, the NASA CRM. Improvements in using multi-fidelity data vs. single-fidelity data are emphasised as well.

1.1 Gaussian Process Regression

The basic building block of the multi-fidelity framework is Gaussian Process (GP) regression [14]. It is a supervised learning technique used to build a surrogate model for an unknown function $y = f(\mathbf{x})$ given n observed input-output pairs $\mathcal{D} = \{\mathbf{x}_i, y_i\}$ for $i \in \{1, \dots, n\}$. The function can be non-deterministic and have Gaussian noise, σ , associated with its observations. The unknown function can have multi-dimensional inputs, but must have a scalar output. These input-output pairs can be arranged in matrices X , \mathbf{y} and σ . If the function has an m dimensional input then X is an $(n \times m)$ matrix of inputs, and \mathbf{y} and σ are $(n \times 1)$ vectors of outputs and associated uncertainties, respectively.

In the context of this work, while computer simulations are deterministic, they can have modeling errors and uncertainties. These are treated as Gaussian noise in the function of interest with σ that is proportional to the errors and uncertainties. Experimental data is not deterministic as factors such as natural variation in environmental conditions, or sensor limitations, can introduce uncertainties in the data. These are also estimated as Gaussian noise.

Since these observations can be imperfect, each observation is assumed to carry some Gaussian noise associated with it such that $y_i \sim \mathcal{N}(E(f(\mathbf{x}_i)), \sigma_i^2)$. Assuming that all the observations in \mathcal{D} have a joint Gaussian distribution, a GP can be used as a surrogate model for the data. A GP is completely defined by its mean function, $\mu(\mathbf{x})$, and a kernel function $\mathbf{k}(\mathbf{x}, \mathbf{x}'; \theta)$ that is parameterized by some hyperparameters θ . For the purposes of this study, the squared exponential function is used as the kernel function:

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp \left(- \sum_{d=1}^{d=m} \frac{(x_d - x'_d)^2}{2l_d} \right), \quad (1.1)$$

where m is the dimension of the input. The hyperparameters for this kernel function are the signal variance σ_f^2 and the length scales l_d . The kernel function is used to create a kernel matrix $K \in \mathbb{R}^{n \times n}$ where $K_{ij} = \mathbf{k}(\mathbf{x}_i, \mathbf{x}_j)$.

To enable the GP to estimate functions with a non-zero mean, the mean of $f(\mathbf{x})$ is represented using

p fixed basis functions, $\mathbf{h}(\mathbf{x})$, and learned regression coefficients β . At a minimum, these basis functions include a constant term, but can have multiple polynomial terms. With these in mind, the surrogate model Z evaluated at some location of interest, \mathbf{x}_* , can be represented as some mean value plus a zero-mean GP:

$$Z(\mathbf{x}_*) = \mathbf{h}(\mathbf{x}_*)^T \beta + \mathcal{GP}(0, K(\mathbf{x}_*, \mathbf{x}'_*; \theta)). \quad (1.2)$$

The n_* sample locations and the basis functions can also be arranged in matrices $X_* \in \mathbb{R}^{n_* \times m}$ and $H \in \mathbb{R}^{p \times n_*}$ such that each row of X_* is a m -dimensional sample location and each column of H_* is a p -dimensional result of the basis functions at the locations in X_* .

Combining the GP regression equations for noisy observations with those incorporating explicit basis functions, and writing in the matrix notation, the surrogate model is defined as

$$Z(X_*) \sim \mathcal{GP}(\mu(X_*), \sigma^2(X_*, X_*)), \quad (1.3)$$

$$\mu(X_*) = H_*^T \hat{\beta} + K(X_*, X)[K(X, X) + \text{diag}(\sigma_i)]^{-1}(y - H^T \hat{\beta}), \quad (1.4)$$

$$\sigma^2(X_*, X_*) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \text{diag}(\sigma_i)]^{-1}K(X, X_*), \quad (1.5)$$

where $\hat{\beta} = (H^T V^{-1} H)^{-1} H^T V^{-1} y$ is the best linear estimator for the regression coefficients and $V = K(X, X) + \text{diag}(\sigma_i)$ represents the kernel matrix at the observed points ($K(X, X)$) and includes the Gaussian noise that is associated with each observation (σ_i). The prediction from the surrogate model $Z(X_*)$ is defined by the mean $\mu(X_*)$ and the uncertainty associated with these predictions is represented by the diagonal of the $\sigma^2(X_*, X_*)$ function. To fully define the GP, the hyperparameters of the kernel function need to be learned from the data. The hyperparameters are chosen by maximizing the marginal log-likelihood of the model,

$$\log p(y|x; \theta) = -\frac{1}{2} \log |V| - \frac{1}{2} \alpha^T V^{-1} \alpha - \frac{n}{2} \log 2\pi, \quad (1.6)$$

where $\alpha = (y - H^T \hat{\beta})$.

For consistency across sections, the following low- and high-fidelity analytic functions will be used to show the functioning of the GP regression process:

$$f_{LF}(x) = 0.5 (6x - 2)^2 \sin(12x - 4) + 10(x - 0.5) - 5. \quad (1.7)$$

$$f_{HF}(x) = 2f_{LF}(x) - 20x + 20 + \sin(10 \cos(5x)). \quad (1.8)$$

The high-fidelity function differs from the low-fidelity function by a few polynomial terms and, has high frequency variation that is absent in the low fidelity approximation. The function f_{HF} is plotted in Figure 1.1 as the solid red line.

To show a basic example of how single-fidelity GP regression uses discrete function evaluations to create a continuous representation of the QoI, only the high fidelity function is used. Figure 1.1 shows the GP

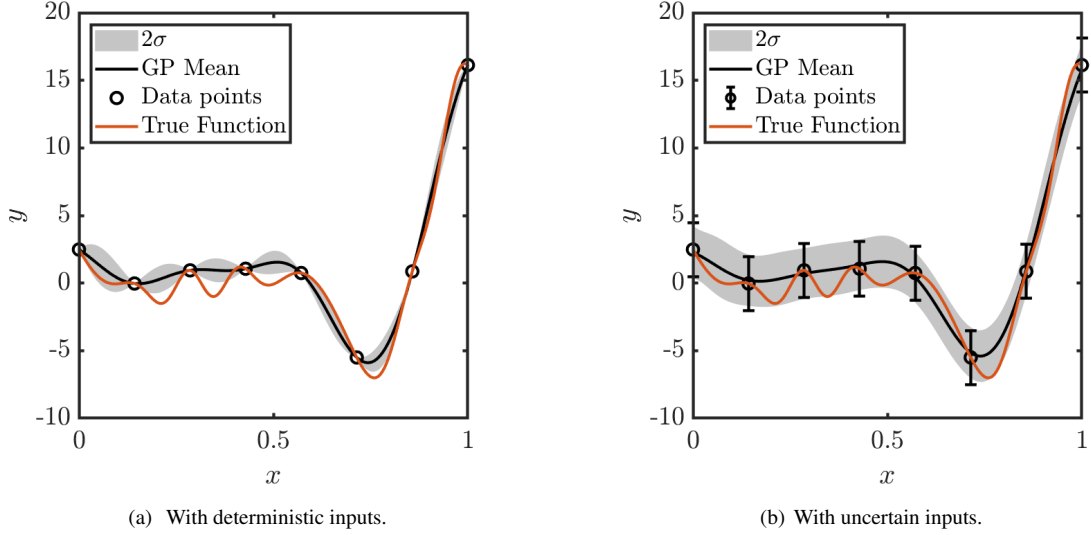


Figure 1.1: GP regression mean and 2σ error estimates when using deterministic and uncertain input data for an underlying analytic function of interest.

regression results when using 8 discrete data points to estimate the function defined by Equation 1.8. This function is represented by the solid red line. The data points, shown as black circles, are uniformly distributed between 0 and 1. The GP regression mean prediction is the solid black line and the 2σ error estimate is represented by the gray area. In the case of Figure 1.1(a), the function evaluations are considered exact, with no associated uncertainty ($\sigma_i = 0$). The error estimate from the GP regression goes to zero near these data points where the uncertainty in the value of the underlying function is zero. The error increases between the data points where the uncertainty in the modeling parameters introduces uncertainty in the prediction of the surrogate model in those areas. Figure 1.1(b) assumes that there is uncertainty associated with the function evaluations ($\sigma_i = 1.0$). The larger, near constant size, error estimate shows that GP regression respects these uncertain data inputs and the GP incorporates it into the prediction of the underlying function.

An important feature of GP regression that is used extensively in this work is the ability to create samples of the GP that are potential representations of the underlying function and take into account the potential errors in the GP. A sample mean, $\mu_S(X_*)$ at X_* locations is generated as

$$\mu_S(X_*) = \mu(X_*) + \sigma^2(X_*, X_*)U, \quad (1.9)$$

where U is a $n_* \times 1$ vector of random variables drawn from a standard normal distribution $\mathcal{N}(0, 1)$. Figure 1.2 show samples drawn from the two GP regressions performed on 1.8 using deterministic and uncertain data. Each colored line is a separate sample and represents a potential candidate for the underlying function that is being estimated, based on the limited data that is provided.

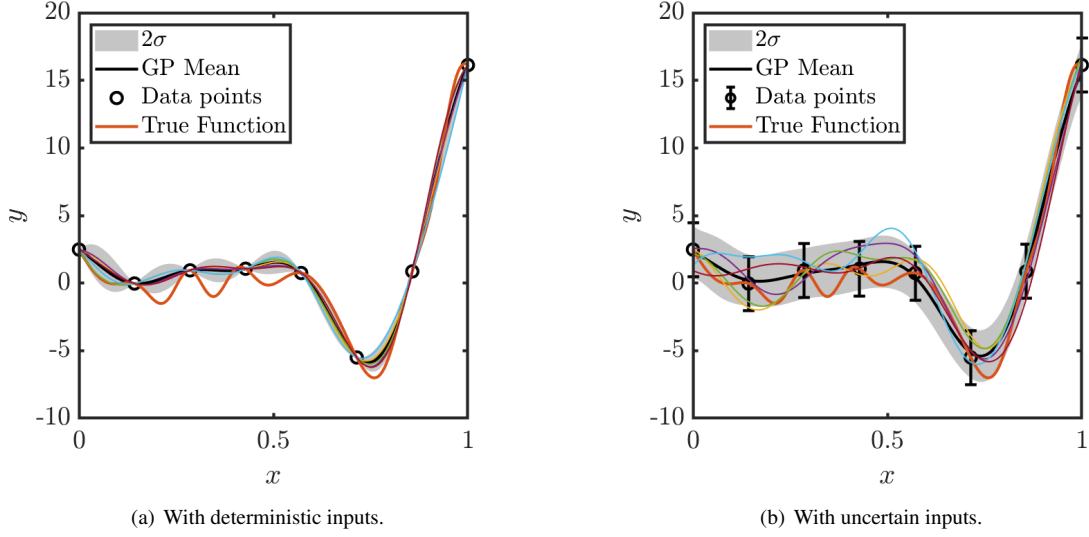


Figure 1.2: Samples of the surrogate models that are candidate functions given the GP regression on the deterministic and the uncertain data.

1.2 Multi-Fidelity Gaussian Process Regression

It is often the case that simulations or experiments of a sufficiently-high fidelity are too expensive to perform over the entire domain of interest for a modeled problem. In many cases, there are lower-fidelity approximations available that can be evaluated quickly to perform parameter studies. The aim of the multi-fidelity GP is to use data from different fidelity levels to create a surrogate model that can best approximate the highest-fidelity function and its uncertainty, while reducing the required number of high-fidelity function evaluations.

Assume there are s information sources $f_t(\mathbf{x})$, where $t \in \{1, 2, \dots, s\}$, and the function at the highest fidelity level, $f_s(\mathbf{x})$, is being approximated using a Gaussian Processes, $Z_s(\mathbf{x}) \sim \mathcal{N}(\mu_s(\mathbf{x}), \sigma_s^2(\mathbf{x}))$. An auto-regressive formulation of the multi-fidelity framework is used. This was first put forward in [6] and was improved upon by [4] to reduce computational cost and improve predictions. The GP approximation at the t fidelity level is modeled as

$$Z_t(\mathbf{x}) = \rho_{t-1}(\mathbf{x})Z_{t-1}(\mathbf{x}) + \delta_t(\mathbf{x}), \quad (1.10)$$

$$\rho_{t-1}(\mathbf{x}) = \mathbf{g}_{t-1}^T(\mathbf{x})\beta_{\rho_{t-1}}, \quad (1.11)$$

where $\mathbf{g}_{t-1}(\mathbf{x})$ is a set of q basis functions, similar to $\mathbf{h}(\mathbf{x})$ in the previous section, $\beta_{\rho_{t-1}}$ is the learned regression coefficients, and $\delta_t(\mathbf{x})$ is modeled using a GP. A way to interpret these terms is to consider $\delta_t(\mathbf{x})$ the additive bias and $\rho_{t-1}(\mathbf{x})$ the multiplicative bias between fidelity levels t and $t - 1$. To account for the

different fidelity levels and their corresponding data, the subscript t is added to the notation introduced in Section 1.1. For example, X_t refers to all the input data at level t . Additionally, the term $\Sigma_t = \text{diag}(\sigma_{i,t}^2)$ is introduced, which refers to the noise in the outputs \mathbf{y}_t .

In Appendix B of [4], Gratiet presents the predictive equations for the case when the design sets are not nested ($\mathcal{D}_t \not\subseteq \mathcal{D}_{t-1}$) and the data has no process noise, such that Σ_t is a null matrix. This work extends those equations to include process noise $\Sigma_t \neq \emptyset$, which produces the following representations for the mean and covariance equations for fidelity level $t \neq 1$ as

$$\begin{aligned} \mu_t(X_*) &= \rho_{t-1}(X_*) \mu_{t-1}(X_*) + H_*^T \beta_t \\ &+ \left[\left(\rho_{t-1}(X_*) \rho_{t-1}(X_t)^T \right) \odot \sigma_{t-1}^2(X_*, X_t) + K_t(X_*, X_t) \right] \\ &\times \left[\left(\rho_{t-1}(X_t) \rho_{t-1}(X_t)^T \right) \odot \sigma_{t-1}^2(X_t, X_t) + V_t \right]^{-1} \\ &\times (\mathbf{y}_t - \rho_{t-1}(X_t) \odot \mu_{t-1}(X_t) - F_t^T \beta_t), \end{aligned} \quad (1.12)$$

and

$$\begin{aligned} \sigma_t^2(X, \tilde{X}) &= \left(\rho_{t-1}(X) \rho_{t-1}(\tilde{X})^T \right) \odot \sigma_{t-1}^2(X, \tilde{X}) + K_t(X, \tilde{X}) - \\ &\left[\left(\rho_{t-1}(X) \rho_{t-1}(X_t)^T \right) \odot \sigma_{t-1}^2(X, X_t) + K_t(X, X_t) \right] \\ &\left[\left(\rho_{t-1}(X_t) \rho_{t-1}(X_t)^T \right) \odot \sigma_{t-1}^2(X_t, X_t) + V_t \right]^{-1} \\ &\left[\left(\rho_{t-1}(X_t) \rho_{t-1}(\tilde{X})^T \right) \odot \sigma_{t-1}^2(X_t, \tilde{X}) + K_t(X_t, \tilde{X}) \right], \end{aligned} \quad (1.13)$$

where (X, \tilde{X}) are generic input arguments, $V_t = K_t(X_t, X_t) + \Sigma_t$, and $\rho_{t-1}(X) = G_{t-1}(X)^T \beta_{\rho_{t-1}}$. $G_{t-1}(X)$ is a $q \times n$ matrix where each column is a q -dimensional result of the basis functions for the corresponding m -dimensional row of input samples in X , and $\beta_{\rho_{t-1}}$ are learned regression coefficients. For the lowest fidelity level, $t = 1$, the regular GP regression equations, Equations 1.4 and 1.5, are used. For a set of sample locations X_* , the mean predictions at fidelity level t is given by $\mu_t(X_*)$ and the variance in the predictions is given by the diagonal of $\sigma_t^2(X_*, X_*)$.

To fully define the GP of each fidelity level, the regression coefficients ($\beta_{\rho_{t-1}}$ and β_t) and the hyperparameters of the kernel functions of each fidelity level need to be learned from the data. The parameter estimation equations from [4] are extended for noisy observations:

$$\begin{bmatrix} \beta_t & \beta_{\rho_{t-1}} \end{bmatrix} = \begin{bmatrix} J_t^T (K_t(X_t, X_t) + \Sigma_t)^{-1} J_t \end{bmatrix}^{-1} \begin{bmatrix} J_t^T (K_t(X_t, X_t) + \Sigma_t)^{-1} \mathbf{y}_t \end{bmatrix}, \quad (1.14)$$

with $J_1 = H_1$ and for $t > 1$, $J_t = \begin{bmatrix} G_{t-1} \odot (\mu_{t-1}(X_t) \mathbf{1}_{q_{t-1}}) & F_t \end{bmatrix}$. $\mathbf{1}_{q_{t-1}} \in \mathbb{R}^{q_{t-1} \times n_t}$ is a matrix of ones. The hyperparameters of the kernel functions are learned by minimizing the negative marginal log-likelihood of each fidelity level.

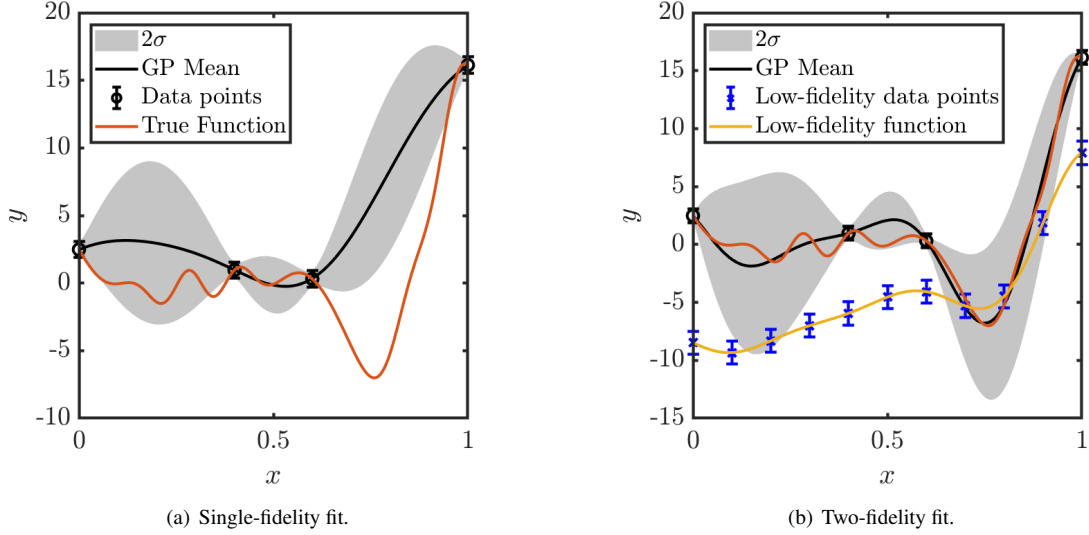


Figure 1.3: Comparison of single- and two-fidelity GP regression.

$$\log p(\mathbf{y}_t | X; \theta) = - \left(\frac{1}{2} \log |V_t| + \frac{1}{2} \alpha^T V_t^{-1} \alpha + \frac{n_t}{2} \log 2\pi \right), \quad (1.15)$$

where $\alpha = (\mathbf{y}_t - \rho_{t-1} \beta_{\rho_{t-1}} - F_t \beta_t)$.

To show the potential advantage of using multi-fidelity data to estimate a function of interest, the analytic functions defined by Equations 1.7 and 1.8 are used as the low-fidelity approximation and the high-fidelity function of interest respectively. For this case, the number of high-fidelity data points is restricted to 4. Figure 1.3(a) shows the mean and error estimates for a one-fidelity GP trained on just the 4 high-fidelity data points. $\sigma_i^{HF} = 0.3$ in this case. There aren't enough data points for the GP regression to learn all the nuanced trends in the high-fidelity function of interest. For Figure 1.3(b), the low-fidelity function (yellow line) is introduced by using 11 equally spaced function evaluations (blue squares). The low-fidelity data doesn't have some of the higher-frequency information that is present in the high-fidelity data, but it approximates the general trends fairly well. In this case, the low-fidelity data bolsters the scant 4 high-fidelity data points, and the multi-fidelity GP that combines both sets of data, is able to provide a more accurate representation of the underlying function of interest. It is also important to notice the large error estimates in GP prediction. More high-fidelity data would be required to reduce the uncertainty in the GP modeling.

As mentioned earlier in this section, the recursive formulation put forth by Gratiet [5] improves on the work originally done by Kennedy and O'Hagan [6] by reducing the computational complexity of the training and sampling steps of the multi-fidelity GP. This is achieved by splitting the dataset into each individual fidelity level instead of agglomerating the data from all levels into one set of equations. This results in having to invert smaller matrices, which greatly improves the computational cost of the process. Figure 1.4 shows the

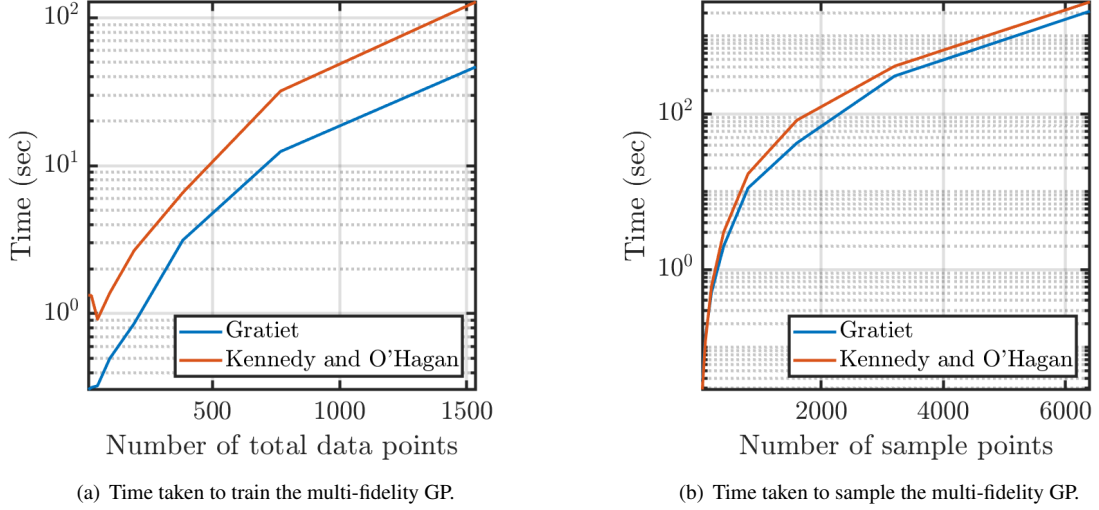


Figure 1.4: Wall clock time comparison to train and query the multi-fidelity Gaussian Process formulations put forth by Kennedy and O'Hagan [6], and Gratiet [4].

comparative times for the training and sampling steps. The same analytic functions used thus far, Equations 1.7 and 1.8, are used to train the GP regressions. In this case, the number of high-fidelity data points (n_{HF}) and low-fidelity data points (n_{LF}) had a constant ratio: $\frac{n_{HF}}{n_{LF}} = 0.2$. These savings in computational time increase with more fidelity levels and higher-dimensional functions.

1.3 Application to NASA CRM

To provide more quantitative applications of the RANS UQ methodology introduced in Chapter ??, we demonstrate its inclusion in the multi-fidelity GP framework. To showcase the predictive capabilities of the multi-fidelity GP framework, the CFD data is augmented with both low-fidelity simulations, using the Athena Vortex Lattice (AVL) code [3], and high-fidelity experimental data, from the wind tunnel campaigns [15, 16] used in the preceding section. For the vortex-lattice simulations, the uncertainty information is provided by subject matter experts (industry users and academics), while for the experimental data, the uncertainty intervals used are those described in the wind-tunnel campaign reports.

For the ease of illustration, C_L , C_D , and C_m are initially considered one-dimensional functions of α ($m = 1$). The methodology described earlier in Section 1.2 is generally applicable to functions of many variables. Aerodynamic databases are often multi-dimensional functions, often with a maximum of 5 input variables (angle of attack, side-slip angle, Mach number, altitude, and dynamic pressure). Multi-dimensional, multi-fidelity databases are explored towards the end of this section.

The results of the multi-fidelity modeling with one input variable (α) are presented in Figures 1.5 – 1.7.

Table 1.1: Number of data points of each fidelity that are used to create Figures 1.5-1.7.

Data Source	Data Points
Low Fidelity (AVL)	23
Medium Fidelity (CFD)	11
High Fidelity (Wind Tunnel)	5

For each figure, the solid black line represents the mean predicted by the GP and the grey area represents the $\pm 2\sigma$ error estimate as predicted by the GP. The left column of figures shows purely the AVL data and a single-fidelity GP fit on that data. In the middle we introduce the SU2 RANS CFD data with uncertainty bounds informed by the RANS UQ methodology, and show the two-fidelity GP fit. On the right we introduce a limited set of wind tunnel data points to inform the highest fidelity, and show the resulting three-fidelity fit. For each QoI the build-up of the database is shown, and the distribution of data points across fidelity levels is shown in Table 1.1.

In the three-fidelity fits of Figures 1.5(c) – 1.7(c) the wind tunnel data points are evenly spread across the domain of interest, $-2^\circ \leq \alpha \leq 12^\circ$. These data points also have uncertainties associated with them, but these are very small and cannot be seen on this scale in the figures. For the second and third column of figures, the mean and standard deviation predictions are made by the multi-fidelity GP methodology from Section 1.2.

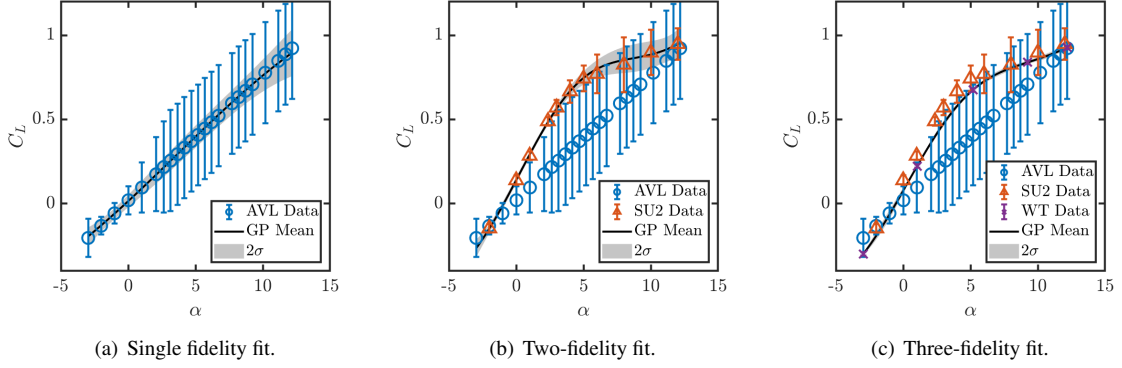
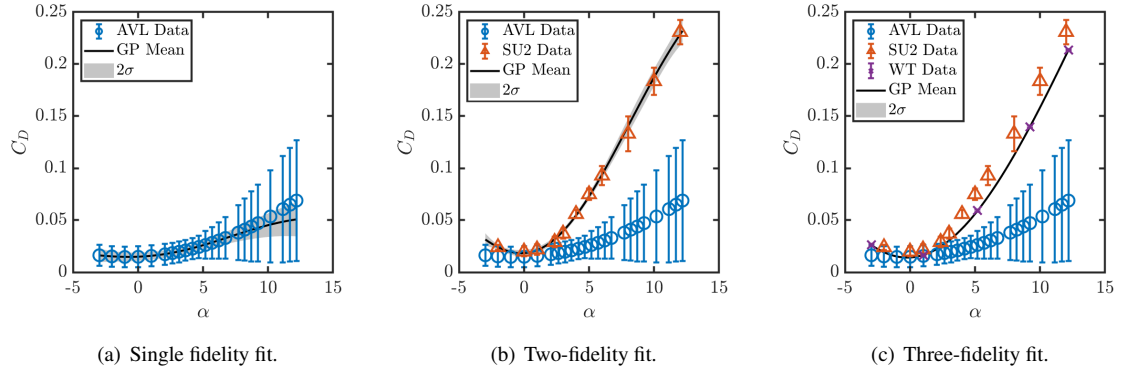
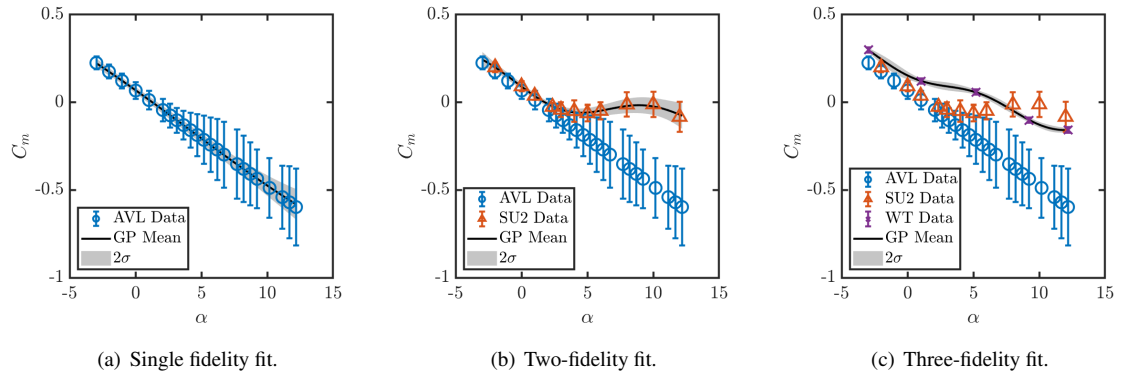
The multi-fidelity GP is able to learn the biases between the different fidelity levels and provide predictions that fit very well with the highest fidelity. To show the benefit of using multi-fidelity data vs. using only high-fidelity data points, the root-mean-square error (RMSE) for both cases and for each QoI is presented in Figure 1.8. The error is calculated using the N highest-fidelity data points that aren't used in training the model as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\mu_{s,i} - y_{s,i})^2}{N}}, \quad (1.16)$$

where $y_{s,i}$ is the i -th data point of the highest (s) fidelity, and $\mu_{s,i}$ is the highest-fidelity prediction from the GP at the same input conditions.

Since the QoIs are simple functions of α , not many high-fidelity data points are required to accurately capture the functional dependence. The differences between the prediction accuracy for a single- vs. multi-fidelity fit is not significant. Nonetheless, the trends are what would be expected. When high-fidelity data is scarce, the multi-fidelity predictions perform better for all QoIs since the low-fidelity data helps provide general trends that are learned by the multi-fidelity GP. As the number of high-fidelity data points increases, the RMSE converges since there is enough data for both fits to accurately reproduce the functional dependence. When the domain is very well covered by high-fidelity data, the single-fidelity fits can do marginally better since the low-fidelity data (in the multi-fidelity fits) doesn't provide any useful information and can serve to introduce noise in the predictions.

Another strength of this multi-fidelity GP methodology is apparent when the high-fidelity data is localized

Figure 1.5: C_L vs α for the NASA CRM, using data from multiple sources of varying fidelity.Figure 1.6: C_D vs α for the NASA CRM, using data from multiple sources of varying fidelity.Figure 1.7: C_m vs α for the NASA CRM, using data from multiple sources of varying fidelity.

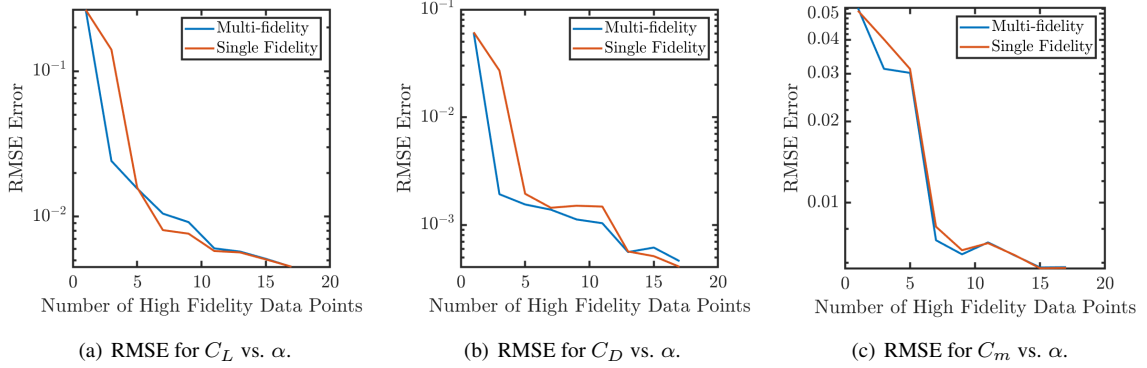


Figure 1.8: Root Mean Square error when using three-fidelity data vs. using only high-fidelity data points.

to a certain part of the domain. Such a situation might arise if resources are limited and it is not feasible to perform high-fidelity evaluations over the entire domain of interest. It might also be the case that the lower-fidelity simulations are fairly accurate in a certain part of the domain and, consequently, introduce smaller uncertainties in these regions of the domain. For the NASA CRM it is mentioned in Section ?? that at low angles of attack, where the flow remains attached to the aircraft, RANS CFD simulations are quite successful at predicting performance metrics. This is evidenced by the smaller uncertainty bounds predicted by the RANS UQ methodology in Figure ?? at $\alpha < 5^\circ$. In this case, an engineer might conclude that highest-fidelity evaluations are not necessary at $\alpha < 5^\circ$ and that sufficient accuracy can be achieved with just the lower-fidelity sources.

To simulate such a situation, a multi-fidelity GP is created that uses AVL and SU2 data that spans the entire domain of interest, but uses wind tunnel evaluations only at high angles of attack ($\alpha > 5^\circ$). This is a manufactured situation where we choose to ignore some of the wind tunnel data to illustrate the ability of the multi-fidelity GP framework to perform reliably without high-fidelity information that spans the domain of interest. Figure 1.9 presents the predictions made by a three-fidelity GP trained on AVL, SU2, and wind tunnel data (left column), and a single-fidelity GP trained on just the wind tunnel data (middle column). The wind tunnel data used to train the models is restricted to high angles of attack, but the unused wind tunnel data is also included in the plots to discern the quality of the predictions. The right column presents the RMSE of the predictions made by the single- and multi-fidelity GP when using localized high-fidelity data. The case that is shown in the left and middle columns, is highlighted in this error comparison. From the RMSE comparison, it is clear that having accurate low-fidelity data at low angles of attack informs the GP prediction in that region, and allows it to follow the trend of the physical phenomena more accurately than when only the localized high-fidelity data is used.

To explore the performance of the multi-fidelity predictive capability in multiple dimensions, the same aerodynamic coefficients from before (C_L , C_D , and C_m) are considered functions of both α and Mach number. Two sources of information, AVL simulations and wind tunnel data, are used to create two-fidelity,

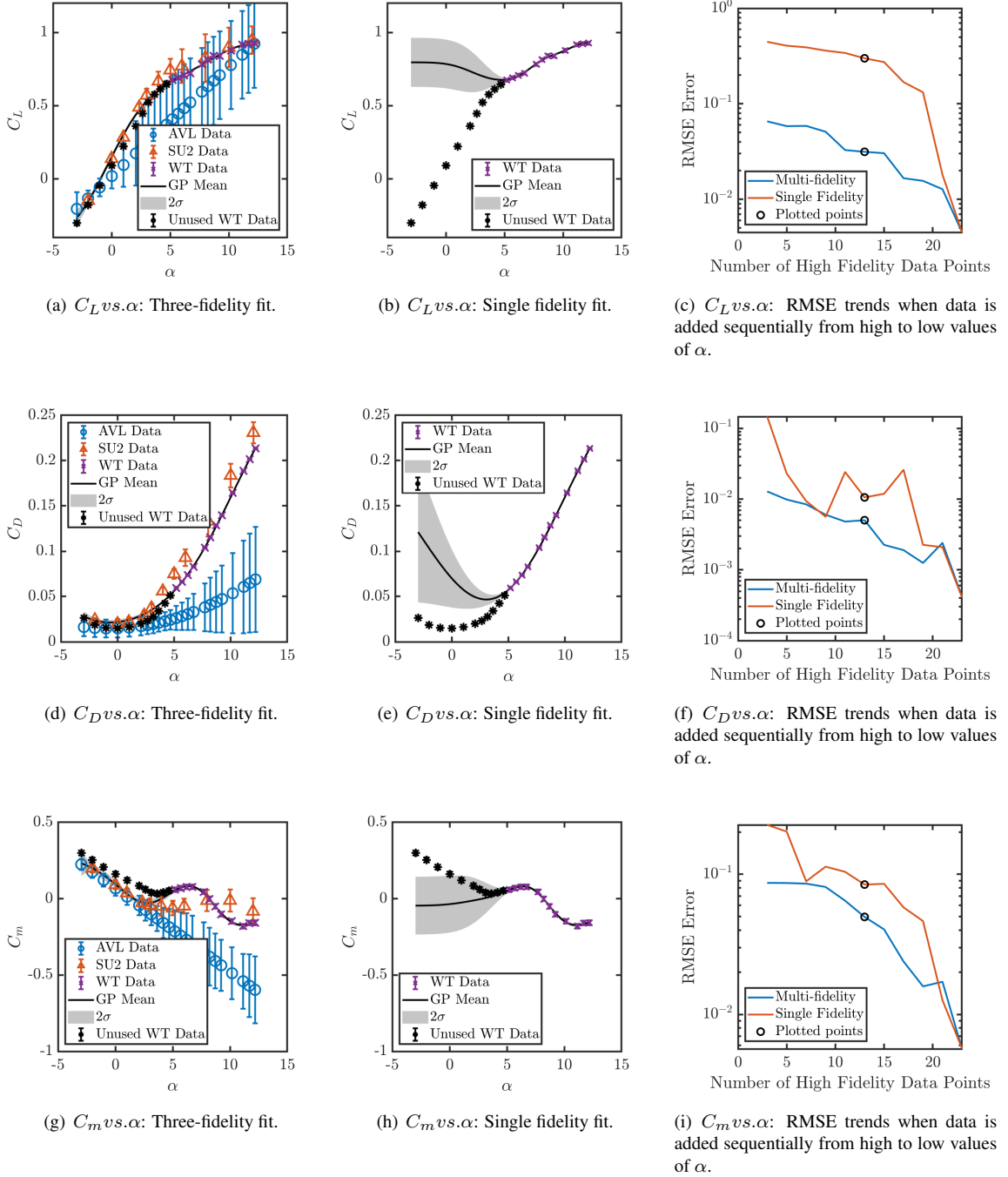
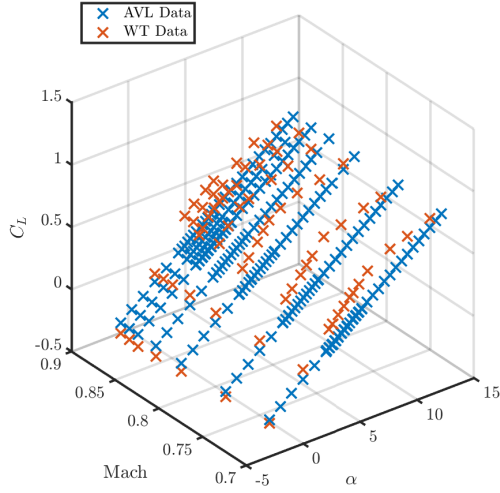


Figure 1.9: Showcasing the superior predictive capability of multi-fidelity data fusion when high-fidelity data is localized in the design space. The left column represents the multi-fidelity GP formulation result, while the middle column shows the results for the single-fidelity formulation. The right column shows the RMSE trends when high-fidelity data is localized. The black circles correspond to the specific case shown in the left and middle columns.

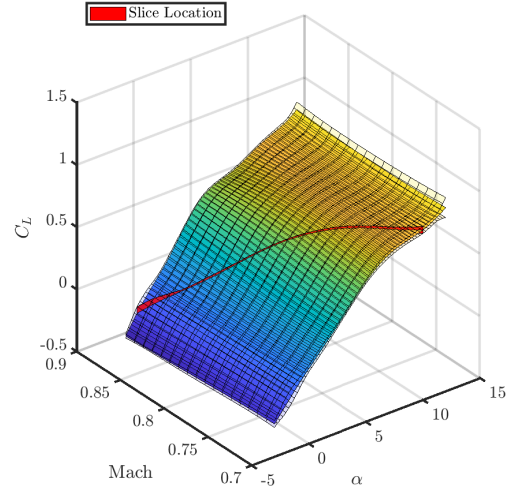
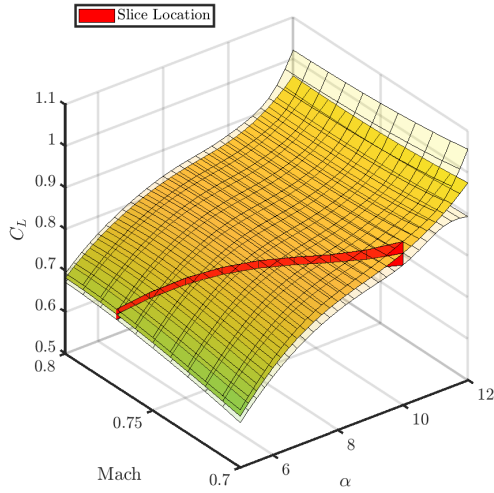
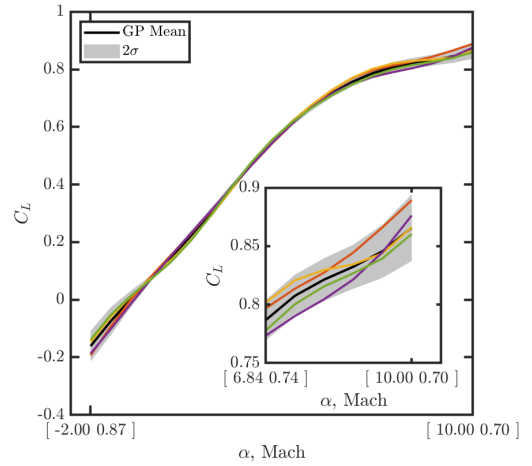
two-dimensional GP. Visualizing the GP predictions in two dimensions is difficult but the results for C_L are shown in Figure 1.10. Figure 1.10(a) shows the two-fidelity data sets that are used to train the multi-fidelity GP. Figure 1.10(b) shows the surfaces that represent the mean (as a contoured solid surface), and 2σ interval estimates (as translucent surfaces sandwiching the mean prediction). The difference between the mean and intervals is hard to discern at that scale. Figure 1.10(c) focuses in on the high-angle of attack (and low Mach number) area when the uncertainties are larger, to show the different surfaces clearly. To show a one-dimensional representation of the data, a slice spanning the two dimensions is taken (shown in red). The mean and 2σ intervals along that slice are plotted in Figure 1.10(d), along with some deterministic samples of the GP (multi-colored lines) that show how they might vary spatially. Again since uncertainties are small, an inset focuses on the high-angle of attack (and low Mach number) area of the domain where the individual samples can be recognized.

These samples of the GP represent candidates for possible functions that would be explained by the data that is provided. The slight variations between the samples are dictated by the error estimate from the GP, which in turn depends on the uncertainty in the input data and GP parameters that are learned. This GP sampling procedure is used extensively in later chapters to create multiple representations of the same aircraft. These samples can be put through the same flight simulations to analyze how the uncertainties in data affect predicted flight performance.

Predictions from the two-fidelity GP are compared to those made from single-fidelity GP that use only the wind tunnel data. The change in the root-mean-square-error (Equation (1.16)) is shown in Fig 1.11. When trying to represent two-dimensional functions, the multi-fidelity fit retains its advantage for longer, with the single-fidelity fit taking ≈ 50 high-fidelity data points to achieve similar accuracy. If the number of high-fidelity points is increased beyond that, the two fits behave identically. For these results, the high-fidelity data was spread evenly across the domain of interest: $-2^\circ \leq \alpha \leq 12^\circ$ and $0.7 \leq \text{Mach} \leq 0.87$. As the number of input dimensions is increased, more data points would be required to capture the functional trends. Leveraging the multi-fidelity improvement in these high-dimensional spaces would be beneficial in reducing time spent collecting high-fidelity data where a lower-fidelity might suffice.



(a) AVL and wind tunnel data points.

(b) Surfaces representing mean and 2σ predictions from GP. Red slice shows one-dimensional location for sampling the GP which is shown in (d).(c) Zoomed in view of surfaces in (b) at high angles of attack (and lower Mach number). The mean and 2σ predictions interval surfaces are more discernible at this scale.(d) One-dimensional representation of mean and 2σ estimates at slice location shown in (b). Multiple samples (colored lines) of the GP are also overlaid to show examples of deterministic sampling. Inset plot focuses in on high angles of attack.Figure 1.10: Two-dimensional representation of C_L as a function of α and Mach number.

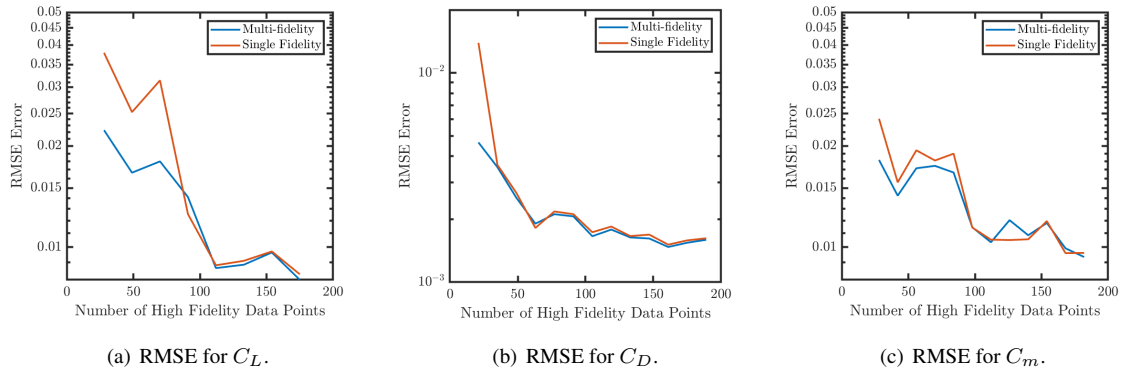


Figure 1.11: Root-mean-square-error for two-dimensional functions of Mach and α when using two-fidelity data vs. using only high-fidelity data points.

Appendix A

A Long Proof

...

Bibliography

- [1] Géraud Blatman and Bruno Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of computational Physics*, 230(6):2345–2367, 2011.
- [2] Thierry Crestaux, Olivier Le Maître, and Jean-Marc Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172, 2009.
- [3] Mark Drela and Harold Youngren. Athena vortex lattice (AVL). *Computer software. AVL*, 4, 2008.
- [4] Loic Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, Université Paris-Diderot, 2013.
- [5] Loic Le Gratiet and Josselin 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.
- [6] M. Kennedy and A. O’Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, March 2000.
- [7] Daniel G Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139, 1951.
- [8] GJA Loeven, JAS Witteveen, and H Bijl. Probabilistic collocation: an efficient non-intrusive approach for arbitrarily distributed parametric uncertainties. In *45th AIAA Aerospace Sciences Meeting and Exhibit*, page 317, 2007.
- [9] Georges Matheron. Principles of geostatistics. *Economic geology*, 58(8):1246–1266, 1963.
- [10] Leo Wai-Tsun Ng and Michael Eldred. Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, page 1852, 2012.
- [11] Sergey Oladyshkin and Wolfgang Nowak. Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion. *Reliability Engineering & System Safety*, 106:179–190, 2012.

- [12] Pramudita Satria Palar, Lavi Rizki Zuhail, Koji Shimoyama, and Takeshi Tsuchiya. Global sensitivity analysis via multi-fidelity polynomial chaos expansion. *Reliability Engineering & System Safety*, 170:175–190, 2018.
- [13] Jooyoung Park and Irwin W Sandberg. Universal approximation using radial-basis-function networks. *Neural computation*, 3(2):246–257, 1991.
- [14] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006. OCLC: ocm61285753.
- [15] Melissa Rivers and Ashley Dittberner. Experimental Investigations of the NASA Common Research Model (Invited). In *28th AIAA Applied Aerodynamics Conference*, Chicago, Illinois, June 2010. American Institute of Aeronautics and Astronautics.
- [16] Melissa Rivers, Craig Hunter, and Richard Campbell. Further Investigation of the Support System Effects and Wing Twist on the NASA Common Research Model. In *30th AIAA Applied Aerodynamics Conference*, New Orleans, Louisiana, June 2012. American Institute of Aeronautics and Astronautics.
- [17] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability engineering & system safety*, 93(7):964–979, 2008.