AIAA AVIATION Forum
June 25-29, 2018, Atlanta, Georgia
2018 Multidisciplinary Analysis and Optimization Conference

10.2514/6.2018-4176

Gaussian Process Regression for Bayesian Fusion of Multifidelity Information Sources

S.F. Ghoreishi* and D.L. Allaire†

Department of Mechanical Engineering, Texas A&M University, College Station, TX, 77843

Predictions and design decisions for complex systems often can be made or informed by a variety of information sources. These information sources describe the quantity of interest with different levels of fidelity. We propose a Bayesian approach in which prior beliefs about the information sources are represented in terms of Gaussian processes and we utilize these sources to generate a fused model with superior predictive capability than any of the constituent models. For this, we implement a multifidelity co-Kriging model aimed at constructing an accurate estimate of the quantity of interest by exploiting data from all models. The key feature of our proposed approach is the relaxation of the assumption of hierarchical relationships among information sources. Instead, we consider an autoregressive model where each information source is related to the highest fidelity information source. The approach is demonstrated on a one-dimensional example test problem and an aerodynamic design problem.

I. Introduction

In engineering, science, and technology, it is often the case that several different computer models in addition to experimental data are available to support decision-making. These information sources typically encompass different mathematical formulae, different resolutions, different physics, and different modeling assumptions that simplify the problem. This leads to information sources with varying degrees of discrepancy from the true quantity of interest, or fidelity, as well as different querying costs. While some information sources may be considered to be of higher fidelity than others, all information sources potentially contribute some amount of information that should be considered in any decision process. Thus, all available information sources should be considered when making inference about a quantity of interest.

One of the main purposes of employing multifidelity approaches is to replace expensive information source queries with several less expensive queries to lower fidelity information sources. These lower fidelity sources often are in the form of surrogates or hierarchies of surrogates.[1] The task then is to create mathematical approaches for fusing the information from these different sources, which usually entails the exploitation of cross-correlations between the outputs of the sources. In Ref. 2, a multifidelity model is proposed based on a linear regression formulation. This model is then improved in Ref. 3, which describes an approach that combines the information from both approximate and accurate models into a single multiscale emulator for the computer model by using a Bayes linear formulation. These methods can suffer from a lack of accuracy since they are based on a linear regression formulation. Ref. 4 presents a co-Kriging model in a Bayesian setting, which is an extension of Kriging for multiple response models. The proposed co-Kriging model is based on an autoregressive relation between the different information sources. This method has been extensively used for different applications, such as multifidelity optimization[5] and a Bayesian formulation proposed in Ref. 6. While co-Kriging models generally provide good predictions they are often computationally expensive to construct. The expense typically arises when large data sets are considered, which can also lead to numerical issues, such as ill-conditioned covariance matrices. Generally Kriging models are known to suffer from these concerns, however, they are even more severe for co-Kriging models due to the increased size of the data set caused by the inclusion of observations from all available information sources. These complexities are mitigated in the works of Refs. 7–10 by dividing the whole set of simulations into groups of simulations

*Ph.D. Candidate, AIAA Student Member, f.ghoreishi88@tamu.edu
†Assistant Professor, AIAA Member, dallaire@tamu.edu

corresponding to certain levels. This is achieved by using a co-Kriging model with an original recursive formulation. Ref. 11 presents a non-intrusive framework based on treed multi-output Gaussian processes, in which the response statistics are obtained through sampling a properly trained surrogate model of the physical system. In Ref. 12, a multifidelity approach is proposed to minimize the number of high-fidelity model evaluations, where the statistics of the high-fidelity model are computed based on realizations of a corrected low-fidelity surrogate. The correction function can be additive, multiplicative, or a combination of the two and it may be updated by high-fidelity model evaluations. In Ref. 13, a methodology is proposed to construct the response surfaces of complex stochastic dynamical systems by blending multiple information sources via auto-regressive stochastic modeling. Ref. 14 presents a multifidelity Gaussian process regression (GPR) approach for prediction of finite-dimensional random fields based on observations of surrogate models or hierarchies of surrogate models. In Ref. 15, a multimodel fusion-based sequential optimization approach is proposed to allocate samples from nonhierarchical multifidelity models for design optimization. Ref. 16 presents a fusion-based multi-information source optimization approach using knowledge gradient policies.

There are several techniques used in practice for combining information from multiple information sources. Among them are the adjustment factors approach,[17–19] Bayesian model averaging,[20–24] and fusion under known and unknown correlation.[25,26] This paper is concerned with the development of an approach for incorporating different available information sources, with potentially differing levels of fidelity and cost, to enable accurate and efficient inference of a real world quantity of interest. Here, we relate the fidelity to uncertainty due to model inadequacy, which is uncertainty due to the omission of some aspects of reality, improper modeling, or unrealistic assumptions.[27–33] This fidelity is characterized by assigning a probability distribution to the output of each individual model on the basis of the model inadequacy associated with that particular model. Surrogate models are constructed for information sources using Gaussian processes,[34] and a single Gaussian process is constructed for the most accurate estimate of the true quantity of interest by using the information obtained from all available information sources. We use the recursive co-Kriging technique proposed in Refs. 7–9 aimed at constructing an accurate estimate of the quantity of interest by leveraging data from all information sources. One of the distinguishing features of our proposed approach is the relaxation of the assumption of hierarchical relationships among information sources. We instead enforce a relationship between all information sources and the highest fidelity information source via an autoregressive model. Our proposed multifidelity approach is applied to a one-dimensional example test problem and an aerodynamic design example.

The rest of the paper is organized as follows. Section II presents the approach proposed here. In Section III, the approach is applied to a one-dimensional test function and an aerodynamic example problem. Conclusions are drawn in Section IV.

II. Approach

In this section, our proposed approach to build a multifidelity surrogate model of a quantity of interest using the information obtained from the multiple levels of fidelity is introduced. Here, we assume we have available some set of information sources, $\bar{f}_i(\mathbf{x})$, where $i \in \{1, 2, ..., S\}$, that can be used to estimate the quantity of interest, $f(\mathbf{x})$, at design point $\mathbf{x}$. Furthermore, we consider that $\bar{f}_S(\mathbf{x})$ is the highest fidelity and the most expensive information source, and $\bar{f}_i(\mathbf{x})$, $i \in \{1, 2, ..., S-1\}$, are the cheaper and less accurate information sources. In order to predict the output of each information source at locations that data are not available, an intermediate surrogate is constructed for each information source using Gaussian processes.[34] These surrogates are denoted by $f_{GP,i}(\mathbf{x})$.

We consider the prior distributions of the information sources modeled by Gaussian processes as

$$f_{GP,i}(\mathbf{x}) \sim \mathcal{GP}\left(\mathbf{0}, k_i(\mathbf{x}, \mathbf{x})\right), \tag{1}$$

where $k_i(\mathbf{x}, \mathbf{x})$ is a real-valued kernel function over the input space. For the kernel function, we consider the commonly used squared exponential covariance function, which is specified as

$$k(\mathbf{x}, \mathbf{x}') = \sigma_s^2 \exp\left(-\sum_{h=1}^{d} \frac{(x_h - x'_h)^2}{2l_h^2}\right), \tag{2}$$

where d is the dimension of the input space, σ_s^2 is the signal variance, and l_h is the characteristic length-scale that indicates the correlation between the points within dimension h. The parameters σ_s^2 and l_h associated with each information source can be estimated via maximum likelihood.

To construct posterior distributions of $f_{GP,i}(\mathbf{x})$ at any point $\mathbf{x}$ in the input space, we assume we have available N_i evaluations of information source i. We denote this information as $\mathbf{X}_{N_i}, \mathbf{y}_{N_i}$, where $\mathbf{X}_{N_i} = (\mathbf{x}_{1,i}, ..., \mathbf{x}_{N_i,i})$ represents the N_i input samples to information source i and $\mathbf{y}_{N_i}$ represents the corresponding outputs from information source i. Given this information, the posterior distributions are given as

$$f_{GP,i}(\mathbf{x}) \mid \mathbf{X}_{N_i}, \mathbf{y}_{N_i} \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma_i^2(\mathbf{x})\right), \tag{3}$$

where

$$\mu_i(\mathbf{x}) = K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma_{n,i}^2 I]^{-1} \mathbf{y}_{N_i}, \tag{4}$$

$$\sigma_i^2(\mathbf{x}) = k_i(\mathbf{x}, \mathbf{x}) - K_i(\mathbf{X}_{N_i}, \mathbf{x})^T [K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i}) + \sigma_{n,i}^2 I]^{-1} K_i(\mathbf{X}_{N_i}, \mathbf{x}), \tag{5}$$

where $K_i(\mathbf{X}_{N_i}, \mathbf{X}_{N_i})$ is the $N_i \times N_i$ matrix whose mn^{th} entry is $k_i(\mathbf{x}_{m,i}, \mathbf{x}_{n,i})$, and $K_i(\mathbf{X}_{N_i}, \mathbf{x})$ is the $N_i \times 1$ vector whose m^{th} entry is $k_i(\mathbf{x}_{m,i}, \mathbf{x})$ for information source i.

In order to estimate the quantity of interest, $f(\mathbf{x})$, using the knowledge available for the information sources, we employ the co-Kriging method by assuming that $f_S(\mathbf{x})$ is the most accurate model that can represent $f(\mathbf{x})$. Co-Kriging was first proposed by Kennedy and O'Hagan.[4] This approach can be computationally infeasible if many levels of fidelity and/or a large number of data observations are available for information sources. In order to overcome this computational complexity, a recursive scheme is proposed by Refs. 8 and 9, which allows the computation of the posterior distribution by a sequence of distinct inferences of smaller dimensions. In this approach, each model output is represented in terms of a Gaussian random field and then such fields are hypothesized to be related to each other by the autoregressive model. In our approach, in order to relax the assumption of hierarchical relationship among the information sources, we assume that all the information sources are only related to the highest-fidelity information source, $f_S(\mathbf{x})$, not to each other, in an autoregressive co-Kriging scheme as

$$f_S(\mathbf{x}) = \gamma_i(\mathbf{x}) f_i(\mathbf{x}) + \phi_i(\mathbf{x}), \tag{6}$$

where $i \in \{1, 2, ..., S-1\}$, $\gamma_i(\mathbf{x})$ is a regression-like parameter representing a scale factor between $f_S(\mathbf{x})$ and $f_i(\mathbf{x})$, and $\phi_i(\mathbf{x})$ is a Gaussian random field independent of all the information sources. In a Bayesian setting, $\gamma_i(\mathbf{x})$ is treated as a random field with an assigned prior distribution that is later fitted to the data through inference. Here, we construct Gaussian processes to predict $\gamma_i(\mathbf{x})$ and $\phi_i(\mathbf{x})$ at all $\mathbf{x}$ in the design space.

In order to construct a Gaussian process for $\gamma_i(\mathbf{x})$, we use the ratio of the output of the highest fidelity model and the information source i as training data. To do so, we accumulate the input samples to information source i, $\mathbf{X}_{N_i}$, and the input samples to information source S, $\mathbf{X}_{N_S}$, in a set represented by $\mathbf{X}_{i,S}$. According to Equation (3), the values of information source i at a design point $\mathbf{x}$ are distributed normally with mean $\mu_i(\mathbf{x})$ and variance $\sigma_i^2(\mathbf{x})$. Therefore, starting from the information source i, we draw N_q independent samples of output values for information sources i and S at each design point $\mathbf{x} \in \mathbf{X}_{i,S}$, as

$$f_i^q(\mathbf{x}) \sim \mathcal{N}\left(\mu_i(\mathbf{x}), \sigma_i^2(\mathbf{x})\right), \quad \text{for } q = 1, \ldots, N_q \text{ and } \mathbf{x} \in \mathbf{X}_{i,S}. \tag{7}$$

Then, at each sample $\mathbf{x} \in \mathbf{X}_{i,S}$, we compute the ratio of the generated output samples of the highest fidelity model, $f_S^q(\mathbf{x})$, and those of the information source i, $f_i^q(\mathbf{x})$, as

$$\gamma_i^q(\mathbf{x}) = \frac{f_S^q(\mathbf{x})}{f_i^q(\mathbf{x})}, \tag{8}$$

and compute the mean and variance of these ratios as

$$\bar{\gamma}_i(\mathbf{x}) = \frac{1}{N_q} \sum_{q=1}^{N_q} \gamma_i^q(\mathbf{x}), \tag{9}$$

$$\sigma_{\gamma_i}^2(\mathbf{x}) = \frac{1}{N_q} \sum_{q=1}^{N_q} \left(\gamma_i^q(\mathbf{x}) - \bar{\gamma}_i(\mathbf{x})\right)^2. \tag{10}$$

These values are then evaluated for all $\mathbf{x} \in \mathbf{X}_{i,S}$ and denoted as vectors of $\bar{\boldsymbol{\gamma}}_i$ and $\boldsymbol{\sigma}^2_{\gamma_i}$. Assuming the squared exponential covariance function and obtaining the parameters by performing the maximum likelihood method, we construct a Gaussian process for γ_i with mean function μ_{γ_i} and covariance matrix Σ_{γ_i} as

$$\gamma_i \sim \mathcal{N}(\mu_{\gamma_i}, \Sigma_{\gamma_i}), \tag{11}$$

where

$$\mu_{\gamma_i} = K_{\gamma_i}(\mathbf{X}_{i,S}, \mathbf{X})^T [K_{\gamma_i}(\mathbf{X}_{i,S}, \mathbf{X}_{i,S}) + \Sigma^{\gamma}_{i,S}]^{-1} \bar{\boldsymbol{\gamma}}_i, \tag{12}$$

$$\Sigma_{\gamma_i} = K_{\gamma_i}(\mathbf{X}, \mathbf{X}) - K_{\gamma_i}(\mathbf{X}_{i,S}, \mathbf{X})^T [K_{\gamma_i}(\mathbf{X}_{i,S}, \mathbf{X}_{i,S}) + \Sigma^{\gamma}_{i,S}]^{-1} K_{\gamma_i}(\mathbf{X}_{i,S}, \mathbf{X}), \tag{13}$$

where $\mathbf{X}$ is any set of samples in the design space, and $\Sigma^{\gamma}_{i,S}$ is a diagonal matrix with diagonal elements of $\boldsymbol{\sigma}^2_{\gamma_i}$.

After constructing Gaussian process for $\gamma_i(\mathbf{x})$, we draw N_γ realizations from $\gamma_i(\mathbf{x})$, and for each realization $\gamma_i^j(\mathbf{x})$, $j \in \{1, 2, ..., N_\gamma\}$, at each $\mathbf{x} \in \mathbf{X}_{i,S}$, the mean and variance of $\phi_i^j(\mathbf{x})$ are computed as

$$\bar{\phi}_i^j(\mathbf{x}) = \mu_{S,i-1}(\mathbf{x}) - \gamma_i^j(\mathbf{x}) \mu_i(\mathbf{x}), \tag{14}$$

$$\sigma^2_{\phi_i^j}(\mathbf{x}) = \sigma^2_{S,i-1}(\mathbf{x}) + {\gamma_i^j}^2(\mathbf{x}) \sigma_i^2(\mathbf{x}), \tag{15}$$

where $\mu_{S,0}(\mathbf{x}) = \mu_S(\mathbf{x})$ and $\sigma^2_{S,0}(\mathbf{x}) = \sigma^2_S(\mathbf{x})$ for $i = 1$. These values are evaluated for all $\mathbf{x} \in \mathbf{X}_{i,S}$ and denoted as vectors of $\bar{\boldsymbol{\phi}}_i^j$ and $\boldsymbol{\sigma}^2_{\phi_i^j}$. Assuming again the squared exponential covariance function and obtaining the parameters by performing the maximum likelihood method, we construct a Gaussian process for ϕ_i^j with mean function $\mu_{\phi_i^j}$ and covariance matrix $\Sigma_{\phi_i^j}$ as

$$\phi_i^j \sim \mathcal{N}(\mu_{\phi_i^j}, \Sigma_{\phi_i^j}), \tag{16}$$

where

$$\mu_{\phi_i^j} = K_{\phi_i^j}(\mathbf{X}_{i,S}, \mathbf{X})^T [K_{\phi_i^j}(\mathbf{X}_{i,S}, \mathbf{X}_{i,S}) + \Sigma^{\phi^j}_{i,S}]^{-1} \bar{\boldsymbol{\phi}}_i^j, \tag{17}$$

$$\Sigma_{\phi_i^j} = K_{\phi_i^j}(\mathbf{X}, \mathbf{X}) - K_{\phi_i^j}(\mathbf{X}_{i,S}, \mathbf{X})^T [K_{\phi_i^j}(\mathbf{X}_{i,S}, \mathbf{X}_{i,S}) + \Sigma^{\phi^j}_{i,S}]^{-1} K_{\phi_i^j}(\mathbf{X}_{i,S}, \mathbf{X}), \tag{18}$$

where $\mathbf{X}$ is any set of samples in the design space, and $\Sigma^{\phi^j}_{i,S}$ is a diagonal matrix with diagonal elements of $\boldsymbol{\sigma}^2_{\phi_i^j}$.

The mean and variance of the fused multifidelity model resulting from the j^{th} realization of $\gamma_i^j(\mathbf{x})$ after fusion of the first i information sources, represented by $\mu^j_{S,i}$ and ${\sigma^2}^j_{S,i}$ are given as[8,9]

$$\mu^j_{S,i}(\mathbf{x}) = \gamma_i^j(\mathbf{x})\mu_i(\mathbf{x}) + \mu_{\phi_i^j}(\mathbf{x}) + K_{\phi_i^j}(\mathbf{x}, \mathbf{X}_{N_S}) K^{-1}_{\phi_i^j}(\mathbf{X}_{N_S}, \mathbf{X}_{N_S}) \left(\mathbf{y}_{N_S} - \gamma_i^j(\mathbf{X}_{N_S}) \circ \mu_i(\mathbf{X}_{N_S}) - \mu_{\phi_i^j}(\mathbf{X}_{N_S}) \right), \tag{19}$$

$${\sigma^2}^j_{S,i}(\mathbf{x}) = (\gamma_i^j(\mathbf{x}))^2 \sigma_i^2(\mathbf{x}) + K_{\phi_i^j}(\mathbf{x}, \mathbf{x}) - K_{\phi_i^j}(\mathbf{x}, \mathbf{X}_{N_S}) K^{-1}_{\phi_i^j}(\mathbf{X}_{N_S}, \mathbf{X}_{N_S}) K^T_{\phi_i^j}(\mathbf{x}, \mathbf{X}_{N_S}), \tag{20}$$

where $\circ$ denotes the "Hadamard" product, which is the element by element matrix product. We denote the vector containing the values of $\gamma_i^j(\mathbf{x})$ for $\mathbf{x} \in \mathbf{X}_{N_S}$ by $\gamma_i^j(\mathbf{X}_{N_S})$. These posterior values are evaluated for all realizations of $\gamma_i^j(\mathbf{x})$, $j \in \{1, 2, ..., N_\gamma\}$, for the i^{th} information source, and the resulting posterior mean and variance of the multifidelity model after fusion of the first i information sources at design point $\mathbf{x}$ are

$$\mu_{S,i}(\mathbf{x}) = \frac{1}{N_\gamma} \sum_{j=1}^{N_\gamma} \mu^j_{S,i}(\mathbf{x}), \tag{21}$$

$$\sigma^2_{S,i}(\mathbf{x}) = \frac{1}{N_\gamma} \sum_{j=1}^{N_\gamma} {\sigma^2}^j_{S,i}(\mathbf{x}). \tag{22}$$

These steps are performed for the other information sources, $i \in \{1, 2, ..., S-1\}$. The values for $f_S^q(\mathbf{x})$ in Equation (7) are now generated based on Equations (21) and (22) and also these equations are used to compute the first terms in Equations (14) and (15) for the next information source. After performing these steps for all the information sources, the resulting $\mu_{S,S-1}$ and $\sigma^2_{S,S-1}$ are the most accurate knowledge that we can gain from the available information sources to estimate the quantity of interest. These values can be used to construct a Gaussian process for our fused multifidelity model as discussed for γ and ϕ. Our proposed approach for Bayesian fusion of multifidelity information sources is presented in Algorithm 1.

Algorithm 1: Bayesian Fusion of Multifidelity Information Sources

---

1: Construct Gaussian processes for all the information sources, $i \in \{1, 2, ..., S\}$, given their available data.
for $i \in \{1, 2, ..., S-1\}$
2: Draw N_q independent samples of output values for information sources i and S at each design point $\mathbf{x} \in \mathbf{X}_{i,S}$ as in Equation (7).
3: Compute the mean and variance of the ratio of the generated output samples according to Equations (8-10).
4: Construct a Gaussian process for γ_i according to Equations (11-13).
5: Draw N_γ realizations from the Gaussian process of γ_i.
for $j \in \{1, 2, ..., N_\gamma\}$
6: Compute the mean and variance of $\phi_i^j(\mathbf{x})$ according to Equations (14-15).
7: Construct a Gaussian process for ϕ_i^j according to Equations (16-18).
8: Compute the mean and variance of the fused multifidelity model according to Equations (19-20).
end for
9: Compute the posterior mean and variance of the multifidelity model after fusion of i information sources according to Equations (21-22).
end for

III. Application and Results

In this section, we present the results of two demonstrations of our methodology. The first is an analytical example problem with one-dimensional input and output. The second demonstration has a two-dimensional input and uses data from two computational fluid dynamics simulators, XFOIL[35] and Stanford University Unstructured (SU2).[36] Details regarding these simulators and their implementation are discussed in Section B.

A. One-Dimensional Example

We consider three information sources with a one-dimensional input. In this case, models 1 and 2 are considered to be the low-fidelity and the high-fidelity models respectively, and model 3 is the highest fidelity model that here, represents the true quantity of interest. These information sources are defined as

$$\begin{aligned} f_1(\mathbf{x}) &= 2 - (1.8 - 3x)\sin(18x + 0.1), \\ f_2(\mathbf{x}) &= 2 - (1.6 - 3x)\sin(18x), \\ f_3(\mathbf{x}) &= 2 - (1.4 - 3x)\sin(18x), \end{aligned} \tag{23}$$

where the domain of $\mathbf{x}$ is limited to $0 \leq \mathbf{x} \leq 1.2$. An illustration of the information sources is shown in Figure 1. We assume that 15 samples are available for models 1 and 2 and 10 samples for the highest fidelity model.

Figure 2 shows the 95% confidence interval and mean of Gaussian processes of the three information sources given their available samples as well as the 95% confidence interval and mean of the fused multifidelity model obtained by our approach.

As it can be seen, the information sources cannot represent the true model accurately, while after fusion of their knowledge and the construction of the multi-fidelity model by our approach, the truth is estimated well.

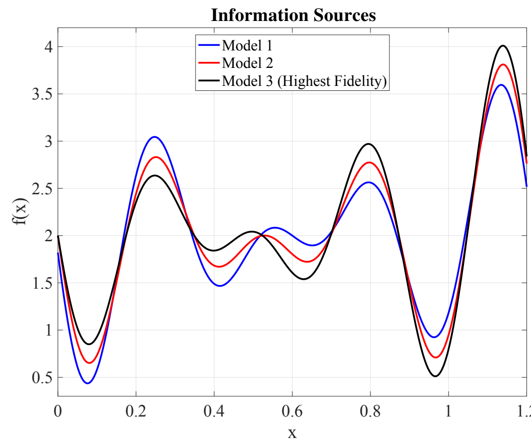


Figure 1: Information sources of problem (23).

B. 2D CFD Demonstration

This demonstration uses the computational fluid dynamics programs XFOIL[35] and SU2[36] as the two simulators. The airfoil of interest is the NACA 0012, a common validation airfoil. The "truth" model used to validate the method is real-world wind tunnel data of the NACA 0012 airfoil,[37,38] which includes 68 data points throughout the design space. The highest fidelity model is built from some set of these real-world wind tunnel data. For this case, the Mach number, M, and the angle of attack, α, are the inputs for the analysis. The quantity of interest is the coefficient of lift, C_L. The design space is $\chi = I_M \times I_\alpha$ with $I_M = [0.15 \quad 0.75]$ and $I_\alpha = [-2.2 \quad 13.3]$.

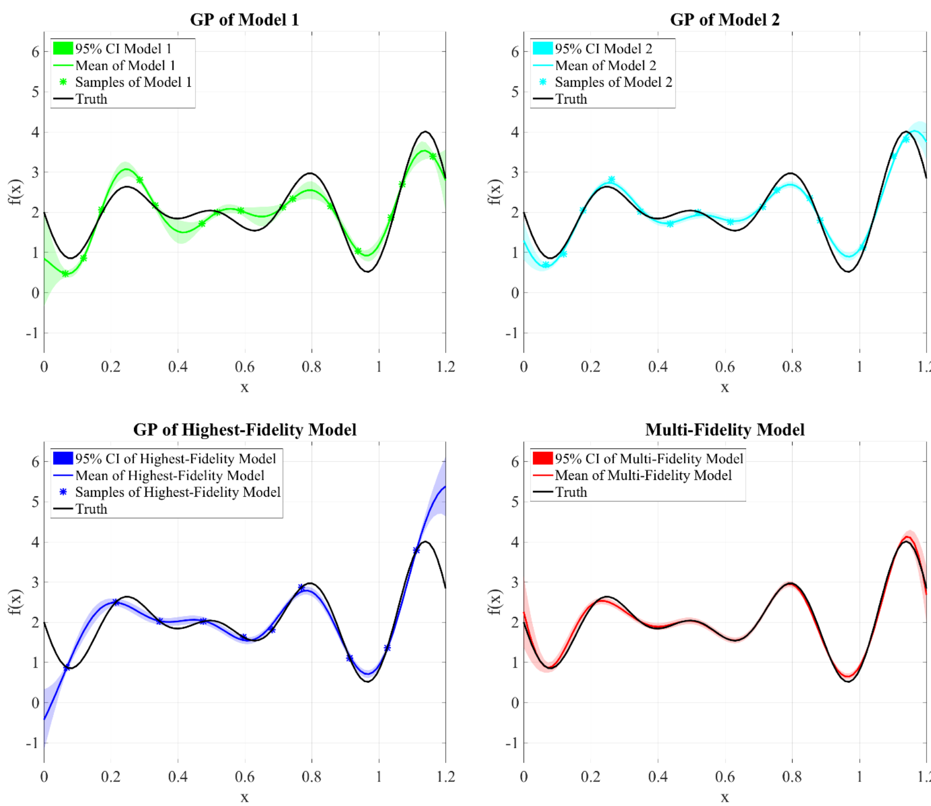
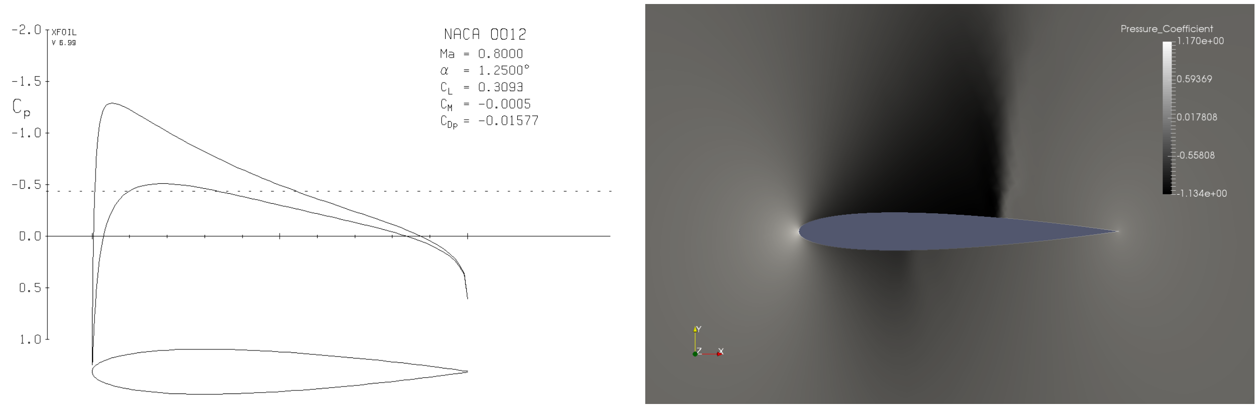


Figure 2: Gaussian processes of the information sources and the multifidelity model obtained by our approach, as well as the true model of problem (23).

XFOIL and SU2 are both very powerful CFD simulators, but have different performance capabilities in various flow regimes. XFOIL is an airfoil solver for the subsonic regime that combines a panel method with the Karman-Tsien compressibility correction for the potential flow with a two-equation boundary layer model. This causes XFOIL to overestimate lift and underestimate drag.[39] SU2, for the case of airfoil analysis, uses a finite volume scheme, the details of which may be found in Ref. 36. SU2 was set to use the Reynoldsaveraged Navier-Stokes (RANS) method with the Spalart-Allmaras turbulence model. This allowed SU2 to be significantly more accurate than XFOIL in the more turbulent flow regimes at higher values of Mach number and angle of attack. This accuracy comes with orders of magnitude increase in computational expense. Figure 3 shows an example output of the two simulators that illustrates the difference in fidelity levels. Some set of wind tunnel data from NASA and AGARD was used to construct the highest fidelity



(a) XFOIL (b) SU2

Figure 3: Example outputs of NACA 0012 airfoil from XFOIL and SU2.

model by interpolating values between the given data points. A comparison between SU2, XFOIL, and the highest fidelity model when the Mach number is fixed at 0.30 is shown in Figure 4. As expected, SU2 performs better than XFOIL at higher angle of attack.

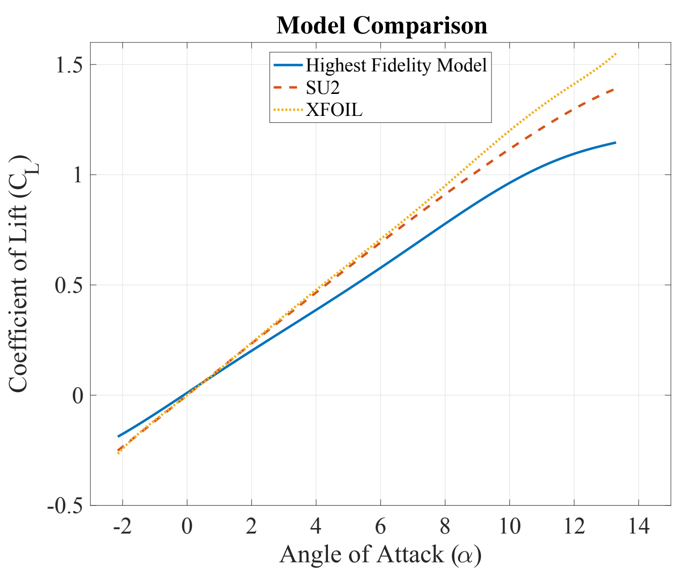


Figure 4: Coefficient of lift estimates from SU2, XFOIL, and the highest fidelity model constructed from wind tunnel data for Mach number fixed at 0.30.

Figure 5 shows the truth model and the fused multifidelity model obtained by our approach for different numbers of available samples for the highest fidelity model (N_S). As it can be seen, as the number of samples available for the highest fidelity model (N_S) increases, our fused multifidelity model gets closer to the truth model.

Figure 6 presents the mean squared errors (MSE) between the fused multifidelity model obtained by our approach and the truth model for different numbers of available samples for the highest fidelity model (N_S). Here, MSE is calculated by sampling a large number of uniformly spaced points in the input space. As it can

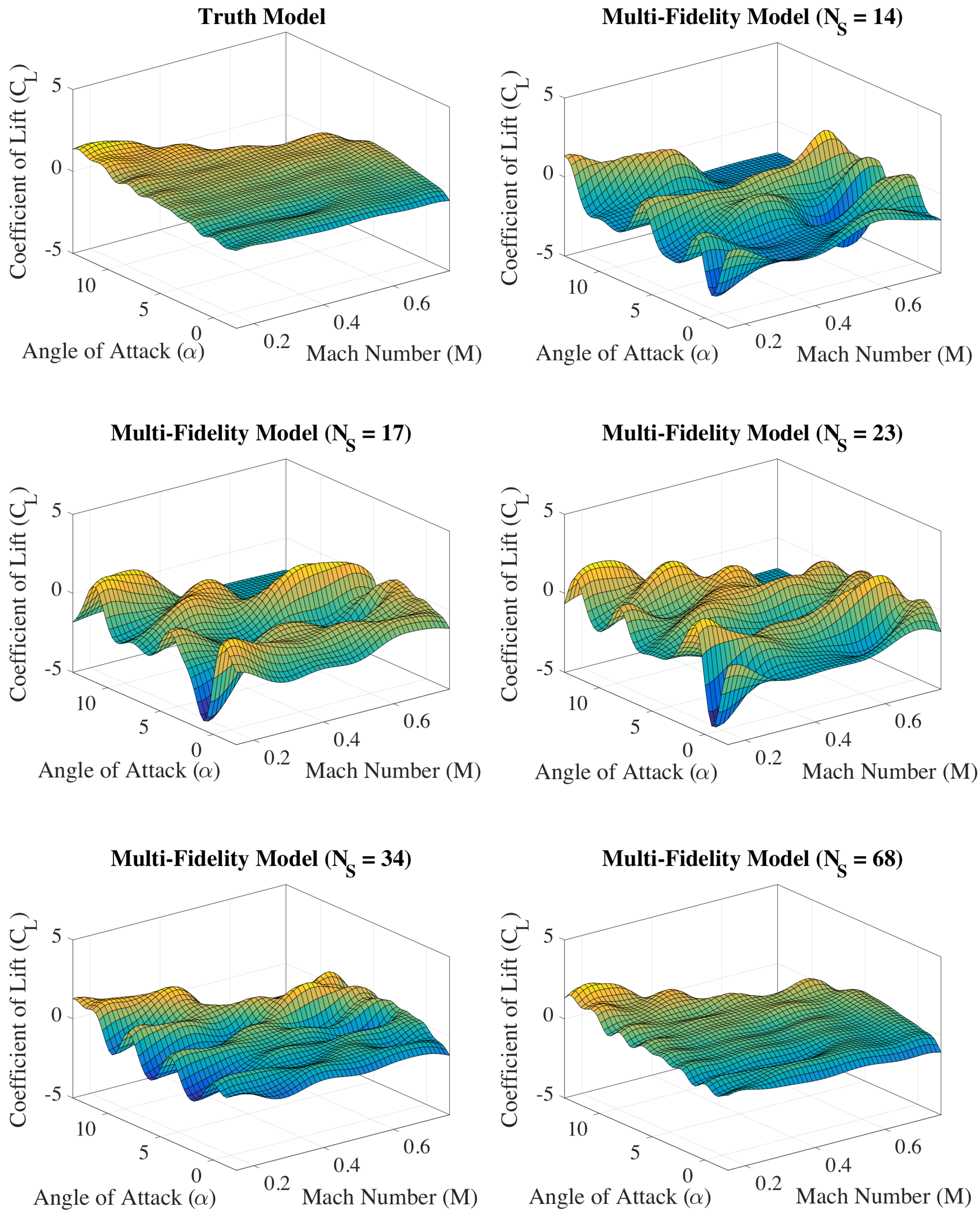


Figure 5: Truth model as well as fused multifidelity models obtained by our approach for different number of available samples for the highest fidelity model (N_S).

be seen, as the number of samples available for the highest fidelity model (N_S) increases, the MSE between our fused multifidelity model and the truth decreases.

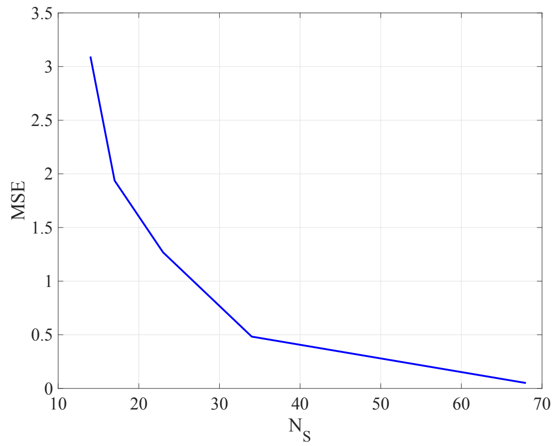


Figure 6: The mean squared error (MSE) between the fused multifidelity model and the truth model for different number of available samples for the highest fidelity model (N_S).

IV. Conclusion

This paper has presented a Bayesian approach to estimate an expensive quantity of interest when different information sources with varying fidelities are available. The approach uses the prior beliefs about the information sources represented in terms of Gaussian processes and utilizes these sources to generate a fused model with superior predictive capability than any of its constituent models. This is achieved by creating a multifidelity co-Kriging model aimed at constructing an accurate estimate of the quantity of interest by leveraging data from all available information sources. The key feature of the proposed approach is the relaxation of the assumption of hierarchical relationships among information sources by making the hypothesis that all the information sources are related to the highest fidelity information source via an autoregressive model. The approach is demonstrated on a one-dimensional example test problem and an aerodynamic design problem. It has been shown that the proposed approach performs well in estimating the quantity of interest.

Acknowledgments

This work was supported partially by the AFOSR MURI on multi-information sources of multi-physics systems under Award Number FA9550-15-1-0038, program manager, Dr. Jean-Luc Cambier and by the National Science Foundation under grant no. CMMI-1663130. Opinions expressed in this paper are of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Allaire, D. and Willcox, K., "Surrogate modeling for uncertainty assessment with application to aviation environmental system models," *AIAA journal*, Vol. 48, No. 8, 2010, pp. 1791–1803.

[2] Craig, P. S., Goldstein, M., Seheult, A., and Smith, J., "Constructing partial prior specifications for models of complex physical systems," *Journal of the Royal Statistical Society: Series D (The Statistician)*, Vol. 47, No. 1, 1998, pp. 37–53.

[3] Cumming, J. A. and Goldstein, M., "Small sample bayesian designs for complex high-dimensional models based on information gained using fast approximations," *Technometrics*, Vol. 51, No. 4, 2009, pp. 377–388.

[4] Kennedy, M. C. and O'Hagan, A., "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, Vol. 87, No. 1, 2000, pp. 1–13.

[5] Forrester, A. I., Sóbester, A., and Keane, A. J., "Multi-fidelity optimization via surrogate modelling," *Proceedings of the royal society of london a: mathematical, physical and engineering sciences*, Vol. 463, The Royal Society, 2007, pp. 3251–3269.

[6] Qian, P. Z. and Wu, C. J., "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments," *Technometrics*, Vol. 50, No. 2, 2008, pp. 192–204.

[7]Le Gratiet, L., *Multi-fidelity Gaussian process regression for computer experiments*, Ph.D. thesis, Université Paris-Diderot-Paris VII, 2013.

[8]Le Gratiet, L., "Bayesian analysis of hierarchical multifidelity codes," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 1, No. 1, 2013, pp. 244–269.

[9]Le Gratiet, L. and Garnier, J., "Recursive co-kriging model for design of computer experiments with multiple levels of fidelity," *International Journal for Uncertainty Quantification*, Vol. 4, No. 5, 2014.

[10]Friedman, S., Isaac, B., Ghoreishi, S. F., and Allaire, D. L., "Efficient Decoupling of Multiphysics Systems for Uncertainty Propagation," *2018 AIAA Non-Deterministic Approaches Conference*, 2018, p. 1661.

[11]Bilionis, I. and Zabaras, N., "Multi-output local Gaussian process regression: Applications to uncertainty quantification," *Journal of Computational Physics*, Vol. 231, No. 17, 2012, pp. 5718–5746.

[12]Ng, L. W.-T. and Eldred, M., "Multifidelity uncertainty quantification using nonintrusive polynomial chaos and stochastic collocation," *Proceedings of the 14th AIAA Non-Deterministic Approaches Conference, number AIAA-2012-1852, Honolulu, HI*, Vol. 43, 2012.

[13]Perdikaris, P., Venturi, D., Royset, J., and Karniadakis, G., "Multi-fidelity modelling via recursive co-kriging and Gaussian–Markov random fields," *Proc. R. Soc. A*, Vol. 471, The Royal Society, 2015, p. 20150018.

[14]Parussini, L., Venturi, D., Perdikaris, P., and Karniadakis, G. E., "Multi-fidelity Gaussian process regression for prediction of random fields," *Journal of Computational Physics*, Vol. 336, 2017, pp. 36–50.

[15]Chen, S., Jiang, Z., Yang, S., and Chen, W., "Multimodel fusion based sequential optimization," *AIAA Journal*, Vol. 55, No. 1, 2016, pp. 241–254.

[16]Ghoreishi, S. F. and Allaire, D. L., "A Fusion-Based Multi-Information Source Optimization Approach using Knowledge Gradient Policies," *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2018, p. 1159.

[17]Mosleh, A. and Apostolakis, G., "The assessment of probability distributions from expert opinions with an application to seismic fragility curves," *Risk Analysis*, Vol. 6, No. 4, 1986, pp. 447–461.

[18]Reinert, J. M. and Apostolakis, G. E., "Including model uncertainty in risk-informed decision making," *Annals of nuclear energy*, Vol. 33, No. 4, 2006, pp. 354–369.

[19]Riley, M. E., Grandhi, R. V., and Kolonay, R., "Quantification of modeling uncertainty in aeroelastic analyses," *Journal of Aircraft*, Vol. 48, No. 3, 2011, pp. 866–873.

[20]Leamer, E. E., *Specification searches: Ad hoc inference with nonexperimental data*, Vol. 53, John Wiley & Sons Incorporated, 1978.

[21]Madigan, D. and Raftery, A. E., "Model selection and accounting for model uncertainty in graphical models using Occam's window," *Journal of the American Statistical Association*, Vol. 89, No. 428, 1994, pp. 1535–1546.

[22]Draper, D., "Assessment and propagation of model uncertainty," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1995, pp. 45–97.

[23]Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T., "Bayesian model averaging: a tutorial," *Statistical science*, 1999, pp. 382–401.

[24]Imani, M. and Braga-Neto, U., "Multiple model adaptive controller for partially-observed boolean dynamical systems," *Proceedings of the 2017 American Control Conference (ACC2017), Seattle, WA*, 2017.

[25]Allaire, D. and Willcox, K., "Fusing information from multifidelity computer models of physical systems," *Information Fusion (FUSION), 2012 15th International Conference on*, IEEE, 2012, pp. 2458–2465.

[26]Thomison, W. D. and Allaire, D. L., "A Model Reification Approach to Fusing Information from Multifidelity Information Sources," *19th AIAA Non-Deterministic Approaches Conference*, 2017, p. 1949.

[27]Kennedy, M. C. and 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.

[28]Ghoreishi, S. and Allaire, D., "Adaptive uncertainty propagation for coupled multidisciplinary systems," *AIAA Journal*, 2017, pp. 1–11.

[29]Friedman, S., Ghoreishi, S. F., and Allaire, D. L., "Quantifying the Impact of Different Model Discrepancy Formulations in Coupled Multidisciplinary Systems," *19th AIAA Non-Deterministic Approaches Conference*, 2017, p. 1950.

[30]Imani, M. and Braga-Neto, U. M., "Maximum-Likelihood Adaptive Filter for Partially Observed Boolean Dynamical Systems," *IEEE Transactions on Signal Processing*, Vol. 65, No. 2, pp. 359–371.

[31]Ghoreishi, S. F., *Uncertainty analysis for coupled multidisciplinary systems using sequential importance resampling*, Master's thesis, 2016.

[32]Imani, M. and Braga-Neto, U. M., "Control of gene regulatory networks with noisy measurements and uncertain inputs," *IEEE Transactions on Control of Network Systems*, 2017.

[33]Ghoreishi, S. F. and Allaire, D. L., "Compositional uncertainty analysis via importance weighted gibbs sampling for coupled multidisciplinary systems," *18th AIAA Non-Deterministic Approaches Conference*, 2016, p. 1443.

[34]Rasmussen, C. E., "Gaussian processes for machine learning," 2006.

[35]Drela, M., "XFOIL: An analysis and design system for low Reynolds number airfoils," *Low Reynolds number aerodynamics*, Springer, 1989, pp. 1–12.

[36]Palacios, F., Alonso, J., Duraisamy, K., Colonno, M., Hicken, J., Aranake, A., Campos, A., Copeland, S., Economon, T., Lonkar, A., et al., "Stanford University Unstructured (SU 2): an open-source integrated computational environment for multi-physics simulation and design," *51st AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, 2013, p. 287.

[37]Thibert, J., Grandjacques, M., and Ohman, L., "NACA 0012 Airfoil," .

[38]Rumsey, C., "2D NACA 0012 Airfoil Validation Case," *NASA Turbulence Modeling Resource*, 2013.

[39]Barrett, R. and Ning, A., "Comparison of Airfoil Precomputational Analysis Methods for Optimization of Wind Turbine Blades," .