Construction of Multi-Fidelity Locally Optimized Surrogate Models for Uncertainty Quantification

Markus P. Rumpfkeil *
University of Dayton, Ohio, 45469, USA

Kyohei Hanazaki †
Nagaoka University of Technology, Niigata 940-2188, Japan

Philip Beran ‡
U.S. Air Force Research Laboratory, Wright-Patterson Air Force Base, USA

In order to enable technological advancements in aircraft designs the conceptual design phase should include both low- and high-fidelity engineering analyses to enable reasonable computational cost and accuracy trade-offs for design space exploration and uncertainty quantification. In this paper it will be demonstrated that accurate multi-fidelity locally optimized surrogate (MFLOS) models can be obtained at lower computational cost than global single-fidelity ones. Once an accurate surrogate model is constructed it can be used for evaluating a large number of conceptual designs and/or Monte-Carlo samples for uncertainty quantification since one can capitalize on the computationally cheap surrogate function evaluations. To demonstrate the power of the MFLOS models analytic test functions as well as a transonic NACA 00xx airfoil case will be employed.

I. Motivation and Background

Computational methods have been playing an increasingly important role in science and engineering analysis and design over the last several decades, due to the rapidly advancing capabilities of computer hardware, as well as increasingly sophisticated and capable numerical algorithms. However, in spite of the rapid advances and acceptance of numerical simulations, serious deficiencies remain in terms of accuracy, uncertainty, and validation for many applications. For example, many real-world problems involve input data that is noisy or uncertain, due to measurement or modeling errors, approximate modeling parameters, manufacturing tolerances, in-service wear-and-tear, or simply the unavailability of information. These imprecise or unknown inputs are important in the design process and need to be quantified in some fashion. Thus, uncertainty quantification (UQ) has emerged as an important area in modern computational engineering in which a probability distribution function (pdf) or interval bound of the simulation objectives is calculated as a function of the inherent uncertainties in order to establish confidence levels over a range of performance predictions.

UQ typically consists of three major phases: (i) characterization of the uncertainty in the input parameters from observations and physical evidence; (ii) uncertainty propagation in which the input variabilities are propagated through the mathematical model; and (iii) calculation of the statistical properties and confidence bounds of the output quantities of interest. Arguably, the computationally most expensive part of UQ is the second phase. The simplest approach to obtain the output statistics in response to input distributions is the Monte-Carlo (MC) method, in which a large number of independent calculations are computed; however, in many practical cases the number of realizations required is too large and results in prohibitively high computational cost, especially for complex, high-fidelity, physics-based simulations.

* Associate Professor, Dept. of Mechanical and Aerospace Eng., Markus.Rumpfkeil@udayton.edu, Senior Member AIAA
† Graduate Student, Department of Mechanical Engineering
‡ Principal Research Aerospace Engineer, Associate Fellow AIAA

Cleared for Public Release, Distribution Unlimited: 88ABW-2016-5983
To save computational time, the use of surrogate models is a very attractive option. The idea of a surrogate model is to replace expensive function evaluations with an approximate but inexpensive functional representation. Especially the polynomial chaos\(^6–^8\) and the kriging model\(^9–^{13}\) have gained popularity in the scientific community. Because an efficient gradient evaluation method based on adjoint formulations is available, gradient enhanced polynomial chaos\(^6–^8\) as well as kriging\(^^{14–17}\) models have been developed in the surrogate model community and have shown very beneficial results. The intuitive reason for this is that, for computational high-fidelity applications targeting a single output objective, the effort for computing the full gradient is, owing to adjoint techniques, comparable to the effort of computing the objective function itself. Therefore, as the number of design parameters, \(M\), increases, using the output function and its derivative information is appealing, because it provides \(M + 1\) pieces of information for roughly the cost of two function evaluations. Kriging also supports the usage of both high- and low-fidelity training points\(^{17–21}\) called variable- or multi-fidelity kriging which will be used in this work. The general idea is to combine trends from low-fidelity (LF) data (e.g., coarser meshes, less sophisticated models) with interpolations of high-fidelity (HF) data (e.g., finer meshes, better models, experimental data).

In order to obtain globally accurate surrogate models, the construction can be refined by a dynamic training point selection with a stopping criteria rather than only specifying the sample size at the beginning and picking the training points through latin hypercube sampling.\(^{22,23}\) This is similar to the concept of expected improvement (EI) when performing optimizations with a kriging model where a potential for improvement is used which considers both estimated function values and uncertainties in the surrogate model, thereby keeping the balance between global and local search performance. One strategy for the dynamic training point selection is to select a large number of test candidates randomly. One can then construct a local response surface using available function (and gradient) information in the neighborhood and compare the global surrogate model function value predictions for the test candidates with the ones from the local model (multi-variate interpolation and regression (MIR)\(^{24}\) is employed here). A few test candidates with the worst discrepancy between the two predictions can then be added to the set of training points, only then evaluating the real function (and gradient) values for these points. This approach has been demonstrated to yield better monotonicity in the convergence (i.e., employing more training points leads to a more accurate surrogate model) and more accurate surrogates for the same number of training points.\(^{22,23}\)

It also provides a convergence criteria as shown in Figure 1. Here, the mean difference between the local and global surrogate (denoted as KR-DIFF) is used in lieu of the actual root-mean squared error (KR-RMSE) and there is no additional computational cost involved in computing KR-DIFF as opposed to the popular leave-one-out cross validation\(^{25,26}\) which can be quite expensive.

An important aspect in UQ is to differentiate between epistemic and aleatory uncertainty. Epistemic uncertainty (or type B, or reducible uncertainty) represents a lack of knowledge about the appropriate value to use for a quantity.\(^{27}\) In contrast, uncertainty characterized by inherent randomness is called aleatory uncertainty (or type A, or irreducible uncertainty). Epistemic uncertainty may or may not be modeled probabilistically, however, regulatory agencies and design teams are increasingly being asked to specifically characterize and quantify epistemic uncertainty and separate its effect from that of aleatory uncertainty.\(^{28}\) For epistemic uncertainties MC methods may be employed, but the results can only be interpreted with regards to the interval produced on the output functional, with no inferred statistical distribution. Other approaches for propagating epistemic uncertainties, such as Dempster-Shafer evidence theory\(^{29–31}\) also typically require a large number of function evaluations in part because it is a generalization of classical proba-
bility theory which is non-intrusive. Typically, for any method employed for UQ the construction of accurate surrogate models is one of the best options since one can then capitalize on the cheap function evaluations of the surrogate. In the context of MC this approach is typically called inexpensive MC (IMC) method. Thus, the development of efficient surrogate models potentially enhanced with lower fidelity information constitutes an important avenue for reducing the cost of UQ for both aleatory and epistemic uncertainties. To address the “curse of dimensionality”, whereby the cost of quantifying uncertainty increases rapidly with $M$, one can combine two different strategies: firstly, select only the input parameters that are truly relevant to the simulation outcome through a sensitivity analysis and thus reduce the dimension of the problem at the outset; secondly, exploit the information gain at reduced additional cost from derivatives and LF data as described above.

For aleatory uncertainty, if one is only interested in the mean and standard deviation of an objective function, moment methods can be a good choice. Moment methods are based on Taylor series expansions of the original non-linear objective function $J(D)$ about the mean of the parameter distribution, $\mu_D$, given standard deviations, $\sigma_{D_j}$. The resulting mean $\mu_J$ and variance $\text{Var}_J$ of the objective function are given to first order (MM1) by

$$ \mu_J = J(\mu_D) \quad \text{Var}_J = \sum_{j=1}^{M} \left( \frac{dJ}{dD_j} \bigg|_{\mu_D} \sigma_{D_j} \right)^2 $$

(1)

However, the method of moments provides no information on the distribution function of the output and when a complete pdf of the objective function is desired, an IMC represents the most straight-forward approach for propagating uncertainties through the simulation process. One of the most simple and often used IMC models is extrapolation around the function value of the mean of the inputs, $J(\mu_D)$. A linear extrapolation for the function value in sample point, $x$, is given by

$$ J_{\text{lin}}(x) = J(\mu_D) + \sum_{j=1}^{M} \frac{dJ}{dD_j} \bigg|_{\mu_D} (x_j - \mu_D) $$

(2)

In this paper, MM1 and IMC with linear extrapolation, as well as with more sophisticated surrogate models, will be applied to analytic test functions and a transonic NACA 00xx airfoil case. In summary, the general goal of this work is to build highly accurate global surrogate models at small overall computational cost for applications to UQ assuming that obtaining high-fidelity training point information is the dominant computational cost factor. The strategy to achieve this is to enhance classic surrogate modeling techniques by adaptively selecting training points as well as utilizing derivative information and to employ lower fidelity information as well as locally-optimized covariances. The last two aspects will be discussed in more detail in Section II, before applications to UQ are demonstrated for analytic test functions and a transonic aero-database in Sections III and IV, respectively. Section V concludes this paper.

II. Enhancements for Classic Surrogate Models

In this section it will be explained how lower fidelity information as well as locally-optimized covariances are employed in this work. These ideas were already presented and used to demonstrate the global accuracy of MFLOS in Rumpfkeil and Beran, but are repeated here for the convenience of the reader.

II.A. Multi-Fidelity Surrogate (MFS) Modeling

The key idea of any multi-fidelity surrogate model is to map the trend of the unknown function underlying the intensively sampled LF data to the less intensively sampled HF data. The most popular method currently used is a correction-based method. The correction is called a bridge function, scaling function or calibration. The correction can be multiplicative, additive or hybrid multiplicative/additive. A multiplicative bridge function is used to locally scale the LF function to approximate the HF function and is typically a low-order polynomial (of constant, linear or second order). An additive bridge function was developed as a global correction and has become the most popular method for multi-fidelity optimization or for data fusion. The additive bridge function should also be of low order but of higher order than the multiplicative one. In general, additive bridge functions are more accurate and robust than multiplicative bridge functions. However,
Gano et al. showed that additive bridge functions are not always better than multiplicative ones. Hence, Gano et al. developed an adaptive hybrid method that combines the multiplicative and additive methods. Most hybrid bridge functions are particularly designed for an optimization context, however, the problem is quite different for the construction of globally accurate surrogates in a relatively large parameter space as needed here.

For the present work a hybrid bridge function approach using kriging adopted from Han et al. was implemented. The relationship between the high- and low-fidelity surrogate model (surrogate values indicated by a hat) in any location $x$ is expressed as:

$$\hat{y}_{HF}(x) = \hat{\phi}(x)\hat{y}_{LF}(x) + \hat{\gamma}(x)$$  \hspace{1cm} (3)

where $\hat{\gamma}(x)$ is an additive bridge function, $\hat{\phi}(x) = f^T(x)\hat{\rho}$ is a low-order polynomial with $q+1$ basis functions $f^T(x) = [1, f_1(x), \ldots, f_q(x)]$ and corresponding coefficients $\hat{\rho} = [\hat{\rho}_0, \hat{\rho}_1, \ldots, \hat{\rho}_q]^T$. The implemented framework supports an arbitrary number of fidelity levels via an approach similar to a multi-grid strategy. The construction of a multi-fidelity surrogate (MFS) model is accomplished via the following four steps:

1. Build kriging model for lowest fidelity data, $\hat{y}_{LF_1}$, using $N_{LF_1}$ lowest fidelity training points
2. Build another kriging model for additive bridge function, $\hat{\gamma}_2$, to connect to next fidelity level, where $\hat{\gamma}_2(x) = \hat{y}_{LF_2}(x) - \hat{\phi}_2(x)\hat{y}_{LF_1}(x)$ in $N_{LF_2}$ next fidelity level training points
3. Compute optimal $\hat{\rho}_2$ during the maximum likelihood estimation updates for $\hat{\gamma}_2$ which yields $\hat{y}_{LF_2}(x) = \hat{\phi}_2(x)\hat{y}_{LF_1}(x) + \hat{\gamma}_2(x)$
4. If $(\cdot)_{LF_2} = (\cdot)_{HF}$ stop, otherwise repeat steps 2 and 3 until highest fidelity level has been reached

II.B. Agglomeration of Locally Optimized Surrogates (ALOS)

The idea is to use multiple local models to overcome the limited modeling flexibility of a single global model when there is heterogeneity in the governing function. In a kriging context if an underlying function shows varying response behavior a stationary covariance structure will result in low quality prediction and overly conservative expected mean squared errors. This effect can be amplified by data collected adaptively and unevenly as done in this work. Thus, utilizing a non-stationary kriging methodology via locally-optimized covariance holds promise as demonstrated by Clark and Bae and Liem et al.

For a more local approach the problem domain must be partitioned into several subregions. One promising way to achieve this is a mixture-of-experts (ME) approach. In the classical ME approach, the partitioning and learning of the problem domain are based on the same algorithm. Tang et al. proposed another approach, which relies on a cluster-based preprocessing step, thus separating the partitioning and learning processes. In the partitioning process similar data is clustered together based on one attribute that reflects the function profile to be modeled (e.g., function value or derivative information). Thus, the clustering or unsupervised learning algorithm is applied to a much simpler one-dimensional problem and several algorithms such as self-organizing feature maps, K-means, distance-based measures, or Gaussian mixture models can be employed. Each local expert is then trained within its own smaller problem domain, which is likely more homogeneous than the entire domain. This approach is typically referred to as the mixture of explicitly localized experts (MILE), whereas the classical ME model is categorized as the mixture of implicitly localized experts (MILE). Masoudnia and Ebrahimpour present a thorough survey of the different ME methods and discuss the advantages and disadvantages of MILE and MELE. One of the main challenges in ME modeling is the automatic determination of the number of experts a priori, which has been identified as a difficult problem in data clustering in general.

Following Liem et al. an explicit mixture-of-experts approach is employed here. The problem domain is first partitioned into several subregions via a GMM unsupervised learning algorithm, which is followed by local expert (surrogate model) training in each subregion. The local predictions can then be combined probabilistically to yield the agglomerated final prediction, $\hat{y}(x)$, in any location $x$:

$$\hat{y}(x) = \sum_{k=1}^{K} \Pi_k(x)\hat{y}_k(x)$$  \hspace{1cm} (4)

American Institute of Aeronautics and Astronautics
where \( \hat{y}_k(x), k = 1, \ldots, K \) are the local surrogate model predictions (ie., the local experts) and \( K \) is the total number of experts. \( \Pi_k(x) \) is the mixing proportion with \( 0 \leq \Pi_k(x) \leq 1 \) and \( \sum_{k=1}^{K} \Pi_k(x) = 1 \). Here, this mixing proportion is given by a modified softmax function

\[
\Pi_j(x) = \frac{\exp[\omega_j(x)]}{\sum_{k=1}^{K} \exp[\omega_k(x)]}
\]

(5)

with \( \omega \) set to 3.0 and \( a_j(x) = \ln[p(x|z_j = 1)p(z_j = 1)] \) where \( p() \) is a probability determined by a Gaussian Naive Bayes supervised learning algorithm and \( z_j \) is a \( K \)-dimensional binary random variable with \( \sum_{k=1}^{K} z_k = 1 \), ie., the \( j \)-th local model is active when \( z_j = 1 \). The agglomeration of locally optimized surrogates (ALOS) framework can be summarized as follows:

1. Perform unsupervised learning using a GMM to initially cluster the available training data. The program EMMIX (Expectation-Maximization-based MIXture analysis)\(^{49}\) is employed here. The user must decide on the clustering criterion (default is function values). The number of clusters is maximized by the algorithm such that at least a user defined minimum number of training points is in each cluster. The algorithm can check for “islands” which may have been created and remove them automatically by reassigning all training points in the island to the largest touching cluster.

2. Build a separate local surrogate model within each cluster, \( \hat{y}_k(x), k = 1, \ldots, K \), and compute the corresponding mixing proportion, \( \Pi_k(x) \), using equation (5). The clusters can either have no overlap at all (ie., they are a partition of unity) or a user defined number of training points from each neighboring cluster can be added to the local training set to promote continuity between clusters.

3. Compute the agglomerated estimation, \( \hat{y}(x) \), using equation (4).

4. Add additional training points to the cluster with the worst discrepancy between the global ALOS and local MIR models subject to a load balancing constraint (not all additional points should always be added to the same cluster) and return to step 2 until a computational budget is exhausted or convergence.

### III. Uncertainty Quantification Applied to Analytic Test Functions

In this section three analytic functions defined on unit cubes of increasingly higher dimensionality are considered as test applications. These were already used to demonstrate the global accuracy of the developed multi-fidelity locally optimized surrogate (MFLOS) models in previous work\(^{36}\) and are again presented in Table 1 along with the employed LF models.

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Functional expression</th>
<th># of initial and final training points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( f(x) = \sin[30(x - 0.9)^4]\cos[2(x - 0.9)] + (x - 0.9)/2 )</td>
<td>17 40</td>
</tr>
<tr>
<td></td>
<td>( f_{LF}(x) = (f - 1.0 + x)/(1.0 + 0.25x) )</td>
<td>50 73</td>
</tr>
<tr>
<td>2</td>
<td>( g(x, y) = \sin[21(x - 0.9)^4]\cos[2(x - 0.9)] + (x - 0.7)/2 + 2y^2\sin(xy) )</td>
<td>25 63</td>
</tr>
<tr>
<td></td>
<td>( g_{LF}(x, y) = (g - 2.0 + x + y)/(5.0 + 0.25x + 0.5y) )</td>
<td>50 88</td>
</tr>
<tr>
<td>3</td>
<td>( h(x, y, z) = \sin[21(x - 0.9)^4]\cos[2(x - 0.9)] + (x - 0.7)/2 + 2y^2\sin(xy) + 3z^3\sin[xyz] )</td>
<td>64 121</td>
</tr>
<tr>
<td></td>
<td>( h_{LF}(x, y, z) = (h - 2.0 + x + y + z)/(5.0 + 0.25x + 0.5y - 0.75z) )</td>
<td>150 207</td>
</tr>
</tbody>
</table>

Table 1. Pertinent information about employed analytic test functions.

The choice of the LF functions implies that the employed linear multiplicative bridge function can recover the HF data exactly. Intuition suggests and numerical experimentation confirmed that LF training point locations should contain the HF locations as a true subset to ensure that no additional error through \( \hat{y}_{LF}(x) \) is introduced when computing \( \gamma(x) = y_{HF}(x) - \hat{y}(x) \) in all \( N_{HF} \) high-fidelity training point locations. The remaining \( (N_{LF} - N_{HF}) \) LF points are picked via latin hypercube sampling (LHS) subject to a distance constraint to the existing HF points. Also, whenever a HF point is added via the dynamic training point algorithm the corresponding LF point is added to the set as well.
The one- and two-dimensional high-fidelity functions are taken from Clark and Bae\textsuperscript{43} and are displayed together with their low-fidelity approximations in Figure 2.

For the uncertainty quantification in this section, 50,000 normally distributed MC sample points are created through LHS with a mean of $\mu_{D_i} = 0.5$ and a standard deviation of $\sigma_{D_i} = 0.15$ for all three analytic test functions presented above. Therefore, the surrogate model domain encompasses about three standard deviations in all coordinate directions. This implies that for the normally distributed input variables more than 99 percent of all samples fall within the surrogate domain and the much less accurate extrapolation capabilities of a surrogate model have only to be used for a small fraction of the samples.

Results for the mean and variance predictions using a full MC, an inexpensive MC using linear extrapolation around the mean value (Lin) as well as MM1 are shown in Table III. As can be seen, MM1 and Lin yield very similar results, as expected from their leading error and they are both far off from the actual mean and variance values as predicted by MC.

<table>
<thead>
<tr>
<th></th>
<th>f (1d) Mean</th>
<th>f (1d) Variance</th>
<th>g (2d) Mean</th>
<th>g (2d) Variance</th>
<th>h (3d) Mean</th>
<th>h (3d) Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>0.0632</td>
<td>0.0643</td>
<td>0.3052</td>
<td>0.0561</td>
<td>0.3780</td>
<td>0.0729</td>
</tr>
<tr>
<td>Lin</td>
<td>0.2847</td>
<td>0.1241</td>
<td>0.3802</td>
<td>0.0809</td>
<td>0.4275</td>
<td>0.1033</td>
</tr>
<tr>
<td>MM1</td>
<td>0.2840</td>
<td>0.1245</td>
<td>0.3805</td>
<td>0.0804</td>
<td>0.4272</td>
<td>0.1031</td>
</tr>
<tr>
<td>Rel. error</td>
<td>349%</td>
<td>94%</td>
<td>25%</td>
<td>43%</td>
<td>13%</td>
<td>41%</td>
</tr>
</tbody>
</table>

Table 2. Comparison of Mean and Variance predictions for the three analytic test functions.

The results of applying an IMC using MFLOS for the uncertainty quantification of the individual test functions will be presented in the next three subsections.

III.A. One-dimensional Analytic Test Function

In Figure 3, the relative errors (in percent) for the mean and variance predictions between MC and an inexpensive MC using agglomerated locally optimized surrogates (ALOS) with two clusters and a global kriging surrogate are shown versus the number of HF training points used to construct the surrogate models. All models started with the same seventeen training points spread non-uniformly over the domain as shown in the left of Figure 2. Then the adaptive training point framework added one training point per iteration until a maximum number of 40 was reached.

The most important observation is that for all surrogate models the mean and variance predictions are quite good (within 1%) using only a fraction of the training points required for a full MC (dozens compared to tens of thousands of function evaluations). One can also see that the ALOS approach outperforms the global approach for both function value only (F) and function and gradient values (FG). The global kriging model errors tend to be larger due to the required compromise in the distance weight or length scale, $\theta$,
in the kriging model which has to be a single value in each dimension. When the modeling is distributed to local experts one can disregard the correlation between samples from different subregions and each local expert is free to select the best model parameters (for example, different distance weights or covariance functions) to better reflect the characteristics of the underlying function in its subregion. This is shown in Table 3 where one can see that the local kriging and gradient-enhanced kriging (GEK) models have different optimum length scales $\theta$ in the partitioned input space as the global kriging and GEK models.

<table>
<thead>
<tr>
<th>Number of clusters</th>
<th>Kriging length scales, $\theta$</th>
<th>GEK length scales, $\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.20</td>
<td>5.47</td>
</tr>
<tr>
<td>2</td>
<td>6.82, 2.92</td>
<td>5.13, 2.61</td>
</tr>
</tbody>
</table>

Table 3. Optimum length scales for kriging and gradient-enhanced kriging (GEK) using 1 and 2 clusters in one dimension with 17 training points.

Figure 4 shows the performance when both the global and ALOS models are enhanced with low-fidelity data (starting with fifty points). In this paper, enhancing the global model with low-fidelity data is coined multi-fidelity surrogate (MFS) modeling whereas enhancing the ALOS model with low-fidelity data is called multi-fidelity locally optimized surrogate (MFLOS) modeling. One can infer that the LF data tends to reduce the errors especially in the beginning (compare dashed and solid lines in same color). The poorer
performance at the end is likely due to over-fitting. The low- and high-fidelity data use the same information, i.e., they are both function value only (F) or both function and gradient values (FG).

### III.B. Two-dimensional Analytic Test Function

In Figure 5 the performances of ALOS with two clusters and a global kriging surrogate for the two-dimensional analytic test function are compared. All models started with the same twenty-five training points shown in the right of Figure 2. Then the dynamic training point framework added two training points per iteration until a maximum amount (63) was reached.

![Figure 5](image)

Figure 5. Plot of error in mean (left) and variance (right) for ALOS with 2 clusters and a global surrogate as a function of number of HF training points in two dimensions.

Again, the overall performance of all surrogate models is quite good when using a few dozen HF training points and a much cheaper alternative to full MC. One can infer that ALOS outperforms the global approach in the beginning and end for function value only (F) whereas it is on par with the global approach for function and gradient values (FG). Figure 6 shows the domains of the two clusters and the training point distribution at the beginning and the end of the simulation. At the end one can observe the dynamic training point algorithm placed new points especially in the first (blue) cluster. One can also observe a smoother transition and thus less agglomerated error between the two clusters compared to the beginning.

![Figure 6](image)

Figure 6. Plot of $\Pi_k(x)$ using the 25 initial points (left) and final 63 points (right). Different colors correspond to different clusters. The highest color intensity within each cluster corresponds to $\Pi_k(x) = 1$.

The local kriging and gradient-enhanced kriging (GEK) models have again different optimum length scales $\theta$ in the partitioned input space as the global kriging and GEK models as shown in Table 4, suggesting
that the ALOS approach has the potential to be better at modeling the different characteristics in the function profile. The difference can be especially seen in the more non-stationary $x$-direction where the global kriging’s $\theta_x$-value of 2.55 is a compromise between the blue cluster’s 3.86 and red cluster’s 0.99 where a lower $\theta$ implies a stronger correlation between the training points.

<table>
<thead>
<tr>
<th>Number of clusters</th>
<th>Kriging length scales, $\theta = [\theta_x, \theta_y]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$[2.55, 0.27]$</td>
</tr>
<tr>
<td>2</td>
<td>$[3.86, 0.30], [0.99, 0.31]$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of clusters</th>
<th>GEK length scales, $\theta = [\theta_x, \theta_y]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$[0.54, 0.04]$</td>
</tr>
<tr>
<td>2</td>
<td>$[0.64, 0.04], [0.30, 0.04]$</td>
</tr>
</tbody>
</table>

Table 4. Optimum length scales for kriging and gradient-enhanced kriging (GEK) using 1 and 2 clusters in two dimensions with 25 training points.

A consequence of this is shown in Figure 7 where plots of the initial models are displayed. The global kriging model exhibits additional fluctuation in the relatively smooth right half of the figure whereas ALOS is much more accurate there. However, ALOS exhibits additional fluctuations in the overlap region due to the mismatch of the two local models as shown in the left of Figure 6.

![Figure 7](image_url)

Figure 7. Plot of truth model (red) as well as the global kriging (white) and ALOS with 2 clusters (blue) using the 25 initial points.

![Figure 8](image_url)

Figure 8. Plot of error in mean (left) and variance (right) for ALOS with 2 clusters and a global surrogate both enhanced with low-fidelity data as a function of number of high-fidelity training points in two dimensions.
Figure 8 shows the performance with enhancement from low-fidelity data (starting with fifty points). Once again the LF data tends to reduce the errors especially in the beginning (compare dashed and solid lines in same color) and suffers from over-fitting at the end.

**III.C. Three-dimensional Analytic Test Function**

In Figure 9 the performances of ALOS with three clusters and a global kriging surrogate for the three-dimensional analytic test function are compared. All models started with the same $4^3 = 64$ equidistantly distributed training points. Then the adaptive training point framework added three training points per iteration until a maximum amount (121) was reached.

![Figure 9. Plot of error in mean (left) and variance (right) for ALOS with 3 clusters and a global surrogate as a function of number of HF training points in three dimensions.](image)

One can infer that ALOS outperforms the global approach in the end for both function value only (F) as well as function and gradient values (FG). Figure 10 shows the domains of the three clusters and the training point distribution at the end of the simulation. One can observe that the dynamic training point algorithm placed new training points especially in the first (red) cluster.

![Figure 10. Plot of $\Pi_k(x)$ using the final 121 points. Different colors correspond to different clusters. The highest color intensity within each cluster corresponds to $\Pi_k(x) = 1$.](image)

The local kriging and GEK models have again different optimum length scales $\theta$ in the partitioned input space as the global kriging and GEK as displayed in Table 5, again suggesting that the ALOS approach has the potential to be better at modeling the different characteristics in the function profile.
Table 5. Optimum length scales for kriging and GEK using 1 and 3 clusters in three dimensions using 64 training points.

<table>
<thead>
<tr>
<th>Number of Clusters</th>
<th>Kriging Length Scales, $\theta = [\theta_x, \theta_y, \theta_z]$</th>
<th>GEK Length Scales, $\theta = [\theta_x, \theta_y, \theta_z]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$[0.27, 0.18, 0.28]$</td>
<td>$[0.84, 0.51, 0.33]$</td>
</tr>
<tr>
<td>3</td>
<td>$[0.29, 0.18, 0.27]$, $[0.33, 0.13, 0.22]$, $[0.33, 0.13, 0.22]$</td>
<td>$[0.63, 0.14, 0.37]$, $[0.63, 0.14, 0.37]$, $[0.63, 0.14, 0.37]$</td>
</tr>
</tbody>
</table>

Figure 11. Plot of error in mean (left) and variance (right) for ALOS with 3 clusters and a global surrogate both enhanced with low-fidelity data as a function of number of HF training points in three dimensions.

Figure 11 shows the performance with enhancement through low-fidelity data (starting with 150 points). The LF data tends to reduce the errors for up to 100 high-fidelity training points (compare dashed and solid lines in same color).

IV. Transonic CFD Application

To demonstrate the true potential of the developed dynamic multi-fidelity locally optimized surrogate (MFLOS) model a transonic CFD example is considered, which involves the steady turbulent flow around a NACA 00xx airfoil. However, a Reynolds-averaged Navier-Stokes (RANS) simulation correctly predicts shock-induced boundary layer separation for larger Mach numbers which is entirely missed by lower fidelity Euler simulations. This is especially prevalent for the computation of the lift coefficient where Euler does not exhibit the same trends as RANS and is thus unsuitable as low-fidelity trend model. Thus, the variations of the drag coefficient, $C_d$, with changes in Mach number ($0.82 \leq M \leq 0.94$), angle of attack ($1^{\circ} \leq \alpha \leq 2^{\circ}$) and thickness to chord ratio ($7\% \leq tc \leq 9\%$) are studied here.

For the uncertainty quantification in this section 3,000 normally distributed sample points are created through LHS with the following means

$$\mu_M = 0.88 \quad \mu_\alpha = 1.5 \quad \mu_{tc} = 8.0$$

and standard deviations

$$\sigma_M = 0.02 \quad \sigma_\alpha = 0.16 \quad \sigma_{tc} = 0.33$$

This means that the surrogate model domain encompasses about three standard deviations in all coordinate directions which covers more than 99% of all MC samples and a much less accurate extrapolation has to be employed for only a very small fraction of the samples.

The high-fidelity level is given by RANS simulations using Fun3d with the Spalart-Allmaras turbulence model. The computational mesh consists of 37,014 hexahedral elements. The low-fidelity data is obtained from the finite-volume solver SU2 in Euler mode employing a mesh which consists of 10,216 triangular elements. One low-fidelity simulation runs about five to ten times faster than the corresponding high-fidelity
simulation. Figure 12 compares the two fidelity levels in the domain of interest. One can see that the low-fidelity trends match the high-fidelity ones which is encouraging for the use of a multi-fidelity approach.

Figure 12. Left: Three isosurfaces of the two fidelity levels. Right: Contours of $C_d$ for the lowest and highest thickness to chord ratio.

Results for the mean and variance predictions using a full MC, an IMC using linear extrapolation around the mean value (Lin) as well as MM1 are shown in Table IV. Again MM1 and Lin yield very similar results and they are both far off from the actual mean and variance values as predicted by MC.

<table>
<thead>
<tr>
<th></th>
<th>Mean of drag</th>
<th>Variance of drag</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC</td>
<td>0.04267</td>
<td>$9.24 \times 10^{-5}$</td>
</tr>
<tr>
<td>Lin</td>
<td>0.03985</td>
<td>$4.15 \times 10^{-5}$</td>
</tr>
<tr>
<td>MM1</td>
<td>0.03985</td>
<td>$4.06 \times 10^{-5}$</td>
</tr>
<tr>
<td>Rel. error</td>
<td>6.6%</td>
<td>55%</td>
</tr>
</tbody>
</table>

Table 6. Comparison of Mean and Variance predictions for transonic airfoil problem.

In Figure 13 the performances of ALOS with two clusters and a global model with and without the enhancement of lower fidelity data is shown. All models started with the same $3^3 = 27$ equally distanced HF training points. Then the adaptive training point framework added three HF training points per iteration until a maximum amount (84) was reached. When lower fidelity data was used the initial locations again
coincided with the HF training points and the remaining points were picked via LHS subject to a distance constraint. Here, 50 low-fidelity points were employed. Also, similar to the analytical function examples whenever a HF point is added via the dynamic training point algorithm the corresponding lower fidelity point is added to the set as well. One can observe that the MFS model (Global F 50 LF) yields more accurate results compared to using the high-fidelity data alone (Global F) even when the cost for obtaining the lower-fidelity samples is taken into account. This is especially true at the beginning of the simulation. ALOS (2 clusters F) on the other hand does not perform well at the beginning as the individual clusters have too few training points. However, after 39 HF training points are reached ALOS outperforms the global kriging approach. MFLOS (2 clusters F 50 LF) outperforms both ALOS and the single-fidelity global kriging, however, it is only roughly on par with the MFS model.

Figure 14 shows the domains of the two clusters and the training point distribution at the end of the simulation. One can observe that the dynamic training point algorithm placed new training points especially in the red cluster which covers a greater area.

![Figure 14](image)

**Figure 14.** Plot of $\Pi_k(x)$ for the transonic airfoil problem using the final 84 points. Different colors correspond to different clusters. The highest color intensity within each cluster corresponds to $\Pi_k(x) = 1$.

V. Conclusion

The goal of this work is to demonstrate highly accurate surrogate models for applications in uncertainty quantification at small overall computational cost assuming that obtaining high-fidelity training point information is the dominant cost factor. The four-fold strategy to achieve this is to enhance surrogate modeling techniques by adaptively selecting training points, utilize derivative information, employ lower fidelity information, and construct locally optimized surrogate models. Especially the latter two strategies are explained in detail in this paper and all four strategies are applied to inexpensive Monte-Carlo simulations for uncertainty quantification of one-, two- and three-dimensional analytic test functions. Lastly, the developed dynamic multi-fidelity locally optimized surrogate (MFLOS) models are also used for uncertainty quantification applied to the steady turbulent flow around a NACA 00xx airfoil where the influence of Mach number, angle of attack and thickness to chord ratio variability on statistics of the drag coefficient are studied. Fairly accurate surrogate models of the database can be constructed using only 60 or so high-fidelity RANS training points.

**Acknowledgments**

This work was supported by the US Air Force Summer Faculty Fellowship Program and the US Air Force Research Laboratory. The third author acknowledges support of the US Air Force Office of Scientific Research (grant 12RB06COR, Dr. Jean-Luc Cambier, Program Officer).
References


