Ninth International Conference on Computational Fluid Dynamics (ICCFD9), Istanbul, Turkey, July 11-15, 2016

ICCFD9-2016-148

Construction of Multi-Fidelity Surrogate Models for Aerodynamic Databases *

Markus Peer Rumpfkeil[1] and Philip Beran[2]

[1] Department of Mechanical and Aerospace Engineering, University of Dayton, USA
[2] U.S. Air Force Research Laboratory, Wright-Patterson Air Force Base, USA
Corresponding author: Markus.Rumpfkeil@udayton.edu

Abstract: In order to enable the technological advancements required for new aircraft, the conceptual design process must accommodate both low- and high-fidelity multidisciplinary engineering analyses. The results of these analyses can be used to create accurate multi-fidelity surrogate models which can be enhanced with derivative information and augmented with dynamic training point selection and local optimization to reduce overall computational cost. In this paper such accurate multi-fidelity surrogate models are used for aerodynamic database creations capitalizing on the computationally cheap surrogate function evaluations.

Keywords: Surrogate Model, Kriging, Aerodynamic Database, Multi-fidelity, Variable-fidelity

1 Motivation and Background

The Air Force Research Laboratory's Multidisciplinary Science and Technology Center is currently investigating conceptual design processes and computing frameworks that could significantly impact the design of the next-generation Efficient Supersonic Air Vehicle (ESAV) [1]. The ESAV is an aircraft concept designed to meet the ever-growing Air Force requirements for mission capability, combat survivability, and lifetime sustainability of future military aircraft. The supersonic, likely tailless, low-observable, and embedded-engine configuration requires a multidisciplinary design and analysis approach. This approach makes it possible to achieve these requirements while capturing the complex, often coupled, physical phenomena present in the operating environment and flight regime, e.g., nonlinear aeroelastic, aerodynamic, and thermal-structural effects. It also allows researchers to exploit these effects and their interactions to achieve advanced aircraft capabilities and configurations otherwise unattainable. These coupled analyses are computationally very expensive, which poses a huge challenge since a large number of configurations must be analyzed [1]. Thus, for inclusion in the overall design routine, the analysis must balance a trade-off between fidelity of the solution and computational time [2]. In addition, being tailless presents unique challenges in lateral control requiring the use of multiple unconventional control effectors for which no previous examples and no empirical knowledge in which to base these designs exist.

In order 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 which can be probed exhaustively. Especially the polynomial chaos [3, 4, 5] and the kriging model [6, 7, 8, 9, 10] have gained popularity. The kriging surrogate model, originally developed in the field of geological statistics, predicts the function value by using stochastic processes, and has the flexibility to represent multimodal functions. Because an efficient gradient evaluation method based on adjoint formulations is available, the introduction of gradient information within surrogate models as additional training data has also attracted attention. The reason for this is that, for computational high-fidelity applications targeting a single output objective, the effort for computing the full gradient is, thanks to adjoint techniques, comparable to the effort of computing the objective function itself. Therefore, as the number of design inputs, M,

*Cleared for Public Release, Distribution Unlimited: 88ABW-2016-3232

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. Thus, one can reasonably expect to have to compute the output function overall far fewer times to obtain a good surrogate model when using gradient information. While the computational effort for adjoint techniques is proportional to the number of output objectives this should not limit the applicability to the ESAV design since in all studies thus far this number has been very small. Gradient enhanced polynomial chaos [3, 4, 5] as well as kriging [11, 12, 13, 14] models have been developed in the surrogate model community and have shown very beneficial results.

In order to obtain a globally accurate surrogate model, the first author refined the construction of surrogate models 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 (i.e., randomly) [15, 16]. 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. The adopted strategy for the dynamic training point selection is to select a large set of test candidates randomly. One can then construct a local response surface (Multi-variate Interpolation and Regression (MIR) [17] is employed in this work) using available function and gradient information in the neighborhood of a test candidate and compare the global surrogate model function value prediction with the local one. 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) value for these points. This approach has been demonstrated to yield better monotonicity in the convergence (i.e., using more training points leads to a more accurate surrogate model) and more accurate surrogates for the same number of training points [15, 16]. It also provides a convergence criteria as shown in Figure 1. Here, the norm of the differences between the local and global surrogate (denoted as DIFF) is used in lieu of the actual root-mean squared error (RMSE) with hardly any computational cost involved in computing DIFF as opposed to the popular Leave-one-out cross validation [18, 19] which can be quite expensive.

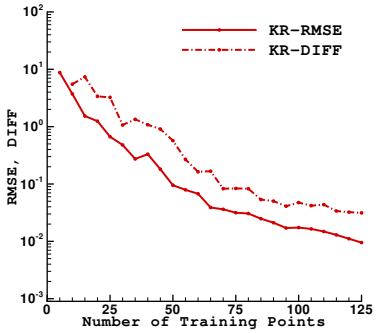

Figure 1: Comparison between actual RMSE and norm of differences between the local and global surrogates.

Kriging also supports the usage of both high- and low-fidelity training points [20, 21, 14, 22, 23]. The general idea is to combine trends from low-fidelity data (e.g., coarser meshes, less sophisticated models) with interpolations of high-fidelity data (e.g., finer meshes, better models, experimental data). An application to an analytic function example where a more accurate surrogate model is constructed by using the trends from low-fidelity functions is shown in Figure 2.

In summary, the general goal of this work is to build highly accurate surrogate models at small overall computational cost assuming that obtaining high-fidelity training point information is the dominant cost factor. The strategy to achieve this is to enhance 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 surrogate models. The last two aspects will be discussed in more detail in the next two Sections, 2 and 3, before the global accuracy is demonstrated for analytic test functions and a transonic aero-database in Section 4. Section 5 concludes this paper.

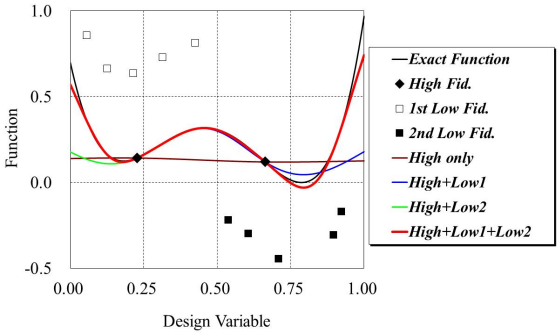

Figure 2: Analytic function example for the use of variable-fidelity kriging.

2 Variable-Fidelity Kriging Surrogate Modeling (VFM)

The key idea of any variable-fidelity surrogate model is to map the trend of the unknown function underlying the intensively sampled low-fidelity (LF) data to the less intensively sampled high-fidelity (HF) data. The most popular method currently used is a correction-based method [23]. The correction is called bridge function, scaling function or calibration. The correction can be multiplicative [24], additive [25, 26] or hybrid multiplicative/additive [27, 28]. 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 variable-fidelity optimization or for data fusion [23]. 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 [23]. However, Gano *et al.* [27] 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 [27, 28] are particularly designed for an optimization context, however, the problem is quite different for the construction of globally accurate surrogates in a possibly relatively large parameter space as needed here.

For the present work a hybrid bridge function approach adopted from Han *et al.* [23] 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:

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

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 variable-fidelity model (VFM) 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 $\gamma_2(x) = y_{LF_2}(x) - \hat{\phi}_2(x)\hat{y}_{LF_1}(x)$ 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 $(.)_{LF_2} = (.)_{HF}$ stop, otherwise repeat steps 2 and 3 until highest fidelity level has been reached

To demonstrate the advantage of a hybrid bridge function compared to an additive one a transonic CFD example is presented next, which involves the steady inviscid flow around a NACA 00xx airfoil. The computational mesh for the employed finite-volume solver SU2 [29, 30] consisted of 10,216 triangular elements. The variations of the lift coefficient with changes in Mach number ($0.8 \leq M_\infty \leq 1.2$), angle of attack ($2° \leq \alpha \leq 8°$) and thickness to chord ratio ($4\% \leq tc \leq 12\%$) are studied. An "exact" database is obtained

from Euler flow solves on a Cartesian mesh of $21 \times 13 \times 9 = 2457$ equispaced nodes and is used for comparisons. The low-fidelity data is calculated using ZEUS [31], which solves the Euler equations with a small disturbance boundary condition and is roughly four times faster than SU2. Figure 3 shows the exact and VFM kriging isosurfaces for three different C_l values. It can be inferred that the VFM is in good agreement with the exact model and is able to capture the transonic behavior very well, demonstrating its ability to model non-smooth functions. One can also observe that the dynamic training point selection clustered high-fidelity points in the more varying transonic region as opposed to the relatively flat supersonic region.

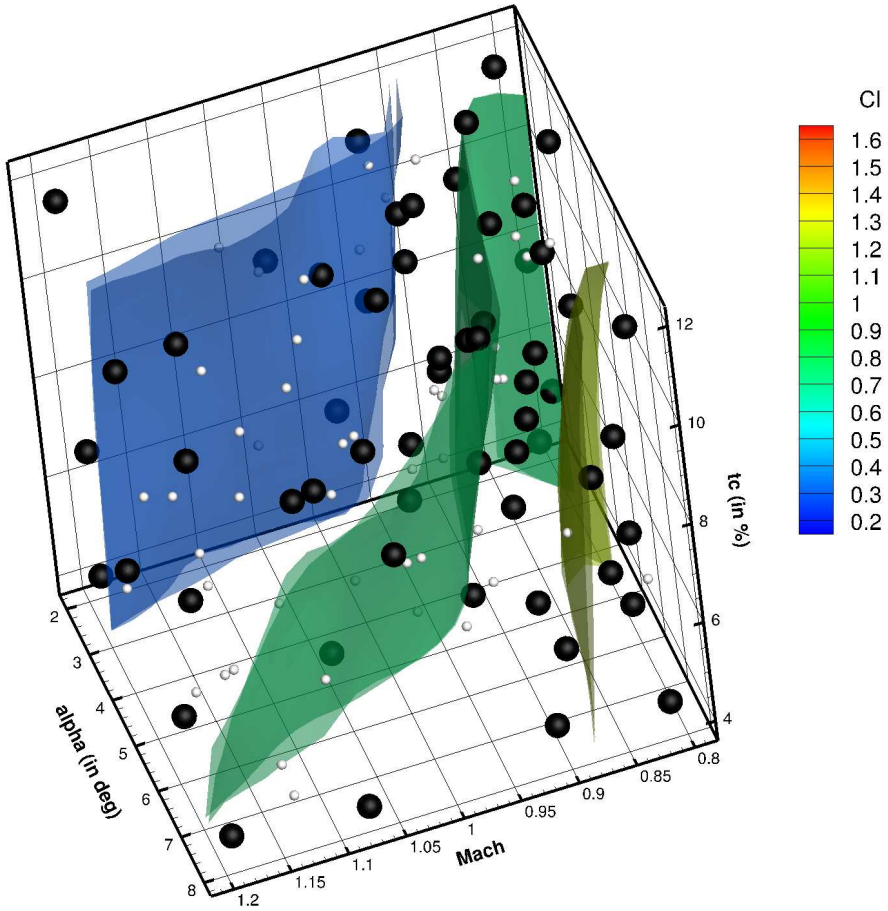

Figure 3: Three isosurfaces of truth model and VFM kriging with 60 high-fidelity and 50 low-fidelity training points using a hybrid bridge function.

In Figure 4 the performances of just an additive and the hybrid bridge function (using a linear multiplier) are compared in more detail. As a benchmark the kriging model constructed using HF training points only is also shown (solid red line). All models started with the same nine HF training points (the corners and center of the domain) together with the total number of LF training points (0-300) which are selected via latin hypercube sampling. Then the adaptive training point framework added three HF training points per iteration until a maximum amount was reached. One can observe that the hybrid bridge functions yields much more accurate results for the same computational cost compared to the additive bridge function. The VFM outperforms the single-fidelity one even when the cost for obtaining the LF samples is taken into account.

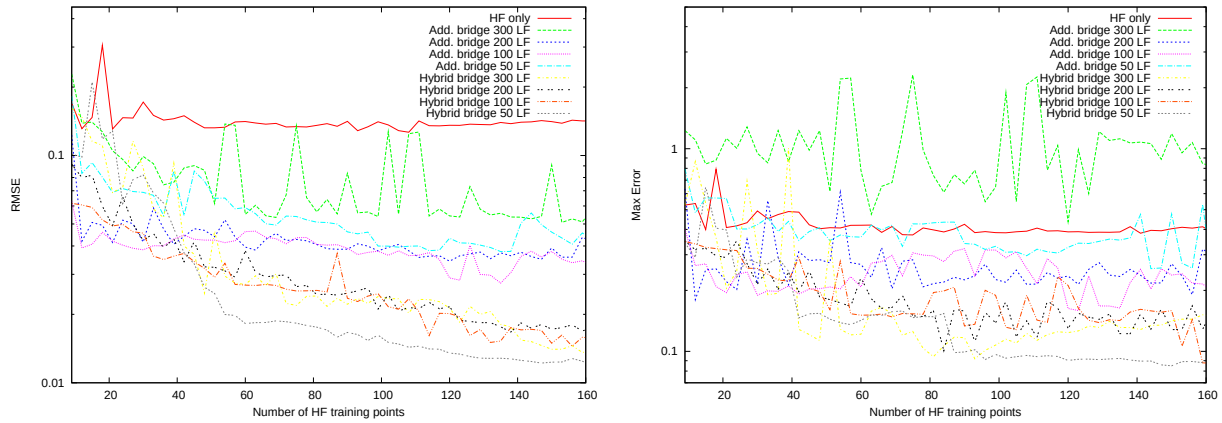

Figure 4: Root-mean squared error (RMSE) (left) and maximum error (right) between the truth model and VFMs as a function of number of high-fidelity training points.

3 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 [32, 33] and Liem *et al.* [34].

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.* [35] 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 (GMM) [36] 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 (MELE), whereas the classical ME model is categorized as the mixture of implicitly localized experts (MILE). Masoudnia and Ebrahimpour [37] 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 [38].

Following Liem *et al.* [34] an explicit mixture-of-experts approach is used here. In their approach, 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:

$$\hat{y}(x) = \sum_{k=1}^{K} \Pi_k(x) \hat{y}_k(x) \tag{2}$$

where $\hat{y}_k(x)$, $k = 1, \ldots, K$ are the local surrogate model predictions (ie. the local experts) where 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

$$\Pi_j(x) = \frac{\exp[\omega a_j(x)]}{\sum_{k=1}^{K} \exp[\omega a_k(x)]} \tag{3}$$

with ω 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) [39] 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, $\hat{y}_k(x) \quad k = 1, \ldots, K$, and compute the corresponding mixing proportion, $\Pi_k(x)$, using equation (3). 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 (2).

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 and return to step 2 until a computational budget is exhausted or convergence.

4 Global Accuracy Results

4.1 Analytic Test Functions

In this subsection the root-mean squared error (RMSE) and maximum error of the various surrogate modeling strategies are compared for three analytic test functions defined on unit cubes of increasingly higher dimensionality.

One-dimensional Analytic Test Function

In Figure 5 the performances of agglomerated locally optimized surrogates (ALOS) and a global kriging surrogate for a one-dimensional analytic test function are compared. All models started with the same seventeen training points spread non-uniformly over the domain. Then the adaptive training point framework added one training point per iteration until a maximum amount of forty was reached. The initial training point distribution and function is taken from Clark and Bae [33] and is given by

$$f(x) = \sin[30(x - 0.9)^4] \cos[2(x - 0.9)] + (x - 0.9)/2$$

One can observe 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, θ, 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. Moreover, each local expert is free to select the best model parameters to better reflect the characteristics of the underlying function in its subregion (e.g., different distance weights or covariance functions for each local kriging model).

Plots of the initial models are shown in Figure 6 and one can observe the larger fluctuation of the global model in the right half of the domain which is due to the required compromise in θ. The optimal distance weight for the first local model (left half) is $\theta = 7.43$ whereas for the second (right half) it is $\theta = 2.41$ where

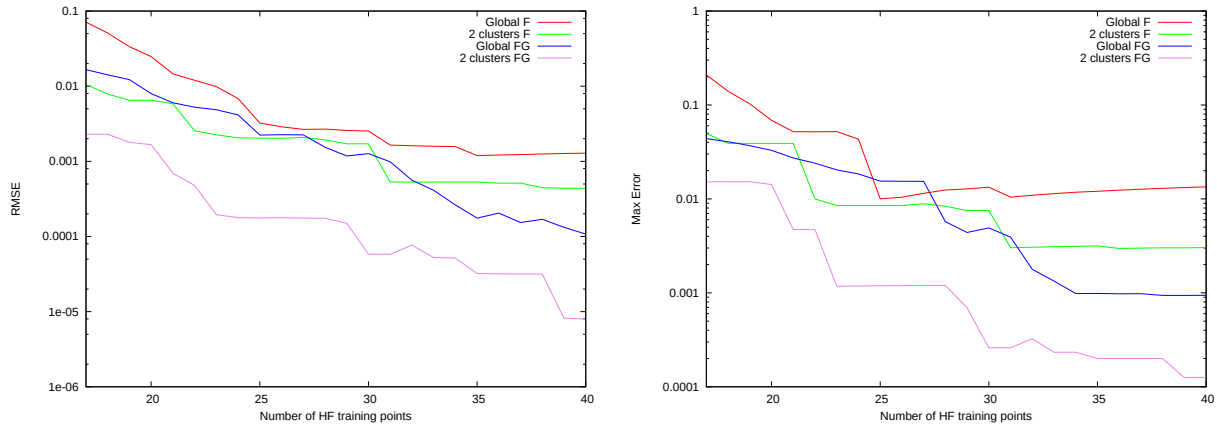

Figure 5: RMSE (left) and maximum error (right) between the one-dimensional truth model and ALOS with 2 clusters as well as a global surrogate as a function of number of high-fidelity training points.

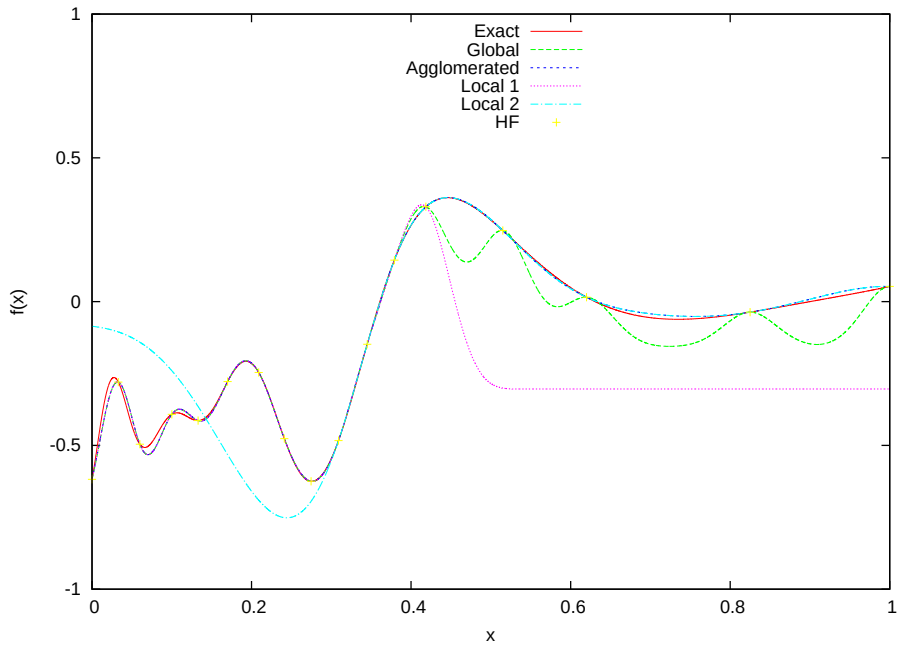

Figure 6: Plot of one-dimensional truth model (red), global kriging (green), two local krigings (pink and light blue) as well as the agglomerated model (dark blue) using the initial seventeen points.

a lower θ implies a stronger correlation between the training points. In contrast, for the global kriging the optimal θ is 7.20 which shows the required compromise and yields a seven times larger RMSE as can be inferred from Figure 5 and visually seen in Figure 6.

Figure 7 shows the performance when the global and ALOS models are enhanced with low-fidelity data. 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 LF data is governed by

$$f_l(x) = (f - 1.0 + x)/(1.0 + 0.25x)$$

which 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

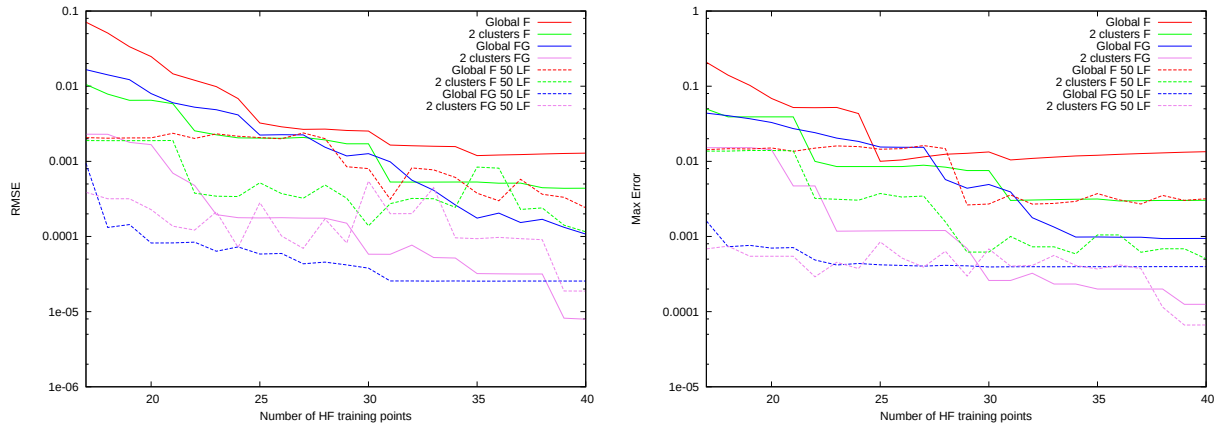

Figure 7: RMSE (left) and maximum error (right) for ALOS with 2 clusters and a global surrogate both enhanced with low-fidelity data as a function of number of high-fidelity training points in one dimension.

computing $\gamma(x) = y_{HF}(x) - \hat{\phi}(x)\hat{y}_{LF}(x)$ in all N_{HF} high-fidelity training point locations. The low- and high-fidelity data use the same information, ie. they are both function value only (F) or both function and gradient values (FG). Initially 50 LF training points are used where 17 are dictated by the HF training locations and the remaining 33 points are picked via latin hypercube sampling subject to a distance constraint to the 17 existing points. Whenever a HF point is added via the dynamic training point algorithm the corresponding LF point is added to the set as well.

Two-dimensional Analytic Test Function

In Figure 8 the performances of ALOS with two clusters and a global kriging surrogate for a two-dimensional analytic test function are compared. All models started with the same twenty-five training points. Then the adaptive training point framework added two training points per iteration until a maximum amount (63) was reached. The function and initial training point distribution is again taken from Clark and Bae [33] and is given by

$$g(x, y) = \sin[21(x - 0.9)^4] \cos[2(x - 0.9)] + (x - 0.7)/2 + 2y^2 \sin[xy]$$

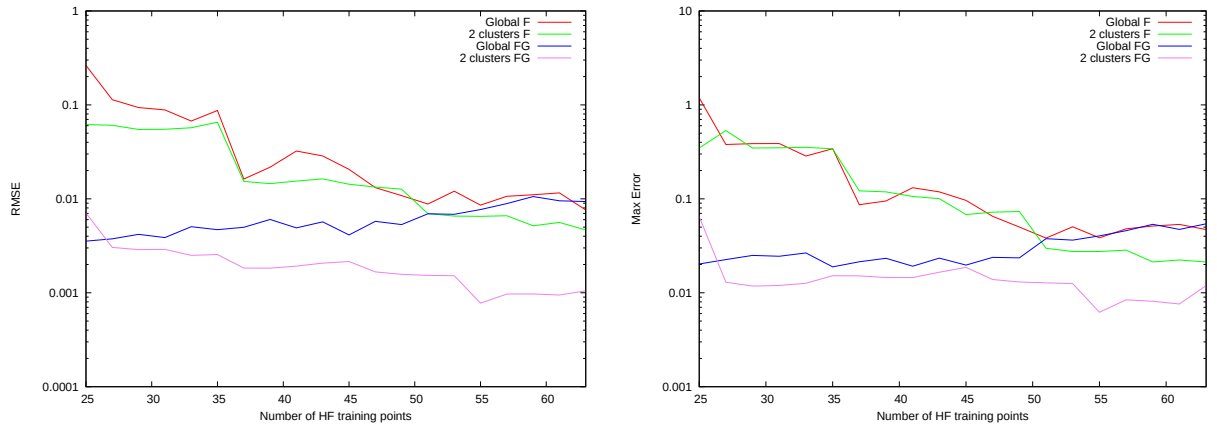

Figure 8: RMSE (left) and maximum error (right) between the two-dimensional truth model and ALOS with 2 clusters as well as a global surrogate as a function of number of high-fidelity training points.

One can infer that ALOS outperforms the global approach in the beginning and end for function value only (F) whereas it is worse than the global approach for function and gradient values (FG) in the beginning

but then does much better than the global approach. Figure 9 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 put the points especially in the second (blue) cluster (see also Figure 11) which also features the largest θ values as shown in Table 1. One can also observe a smoother transition and thus less agglomerated error between the two clusters compared to the beginning.

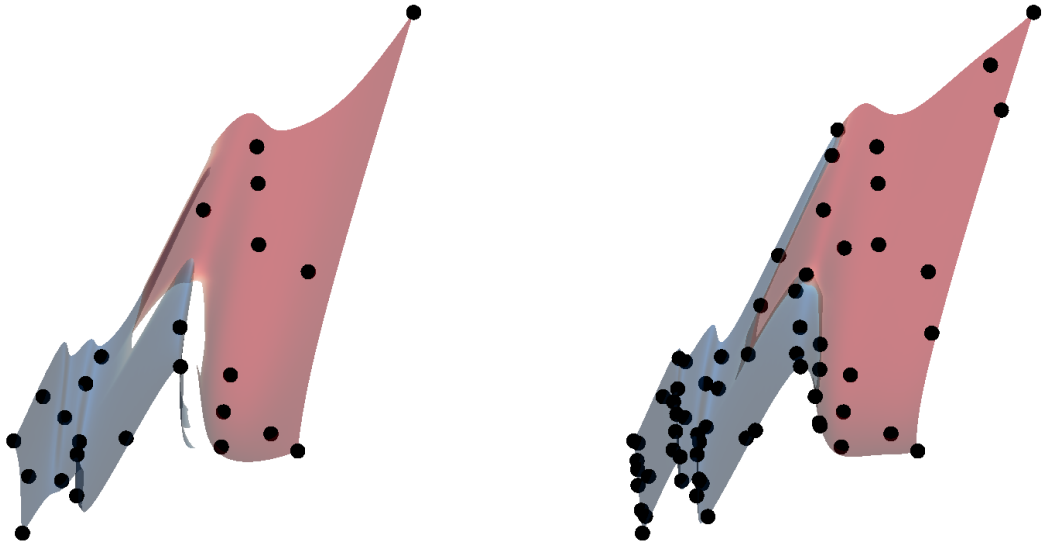

Figure 9: Plot of $\Pi_k(x)$ using the 25 initial points (left) and final 63 points (right). Different colors correspond to different clusters. The highest color intensity within each cluster corresponds to $\Pi_k(x) = 1$.

The local kriging and gradient-enhanced kriging (GEK) models have different optimum length scales $\theta = [\theta_x, \theta_y]$ in the partitioned input space as the global kriging and GEK as shown in Table 1, 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 θ-value of 2.55 is a compromise between the blue cluster's 4.20 and red cluster's 0.59 where a lower θ implies a stronger correlation between the training points. A consequence of this is shown in Figure 10

Number of clusters	Length scales, θ
kriging	
1	[2.55, 0.27]
2	[0.59, 0.21], [4.20, 0.34]
GEK	
1	[0.55, 0.04]
2	[0.32, 0.06], [0.59, 0.04]]

Table 1: Optimum length scales for kriging and GEK using 1 and 2 clusters in two dimensions using 25 training points.

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 9.

A plot of the final models after 38 points have been dynamically added (for a total of 63) is shown in Figure 11 where an excellent agreement between ALOS and the truth model can be visually inferred.

Figure 12 shows the performance with enhancement through low-fidelity data. The LF data tends to reduce the errors for all numbers of high-fidelity training points considered (compare dashed and solid lines

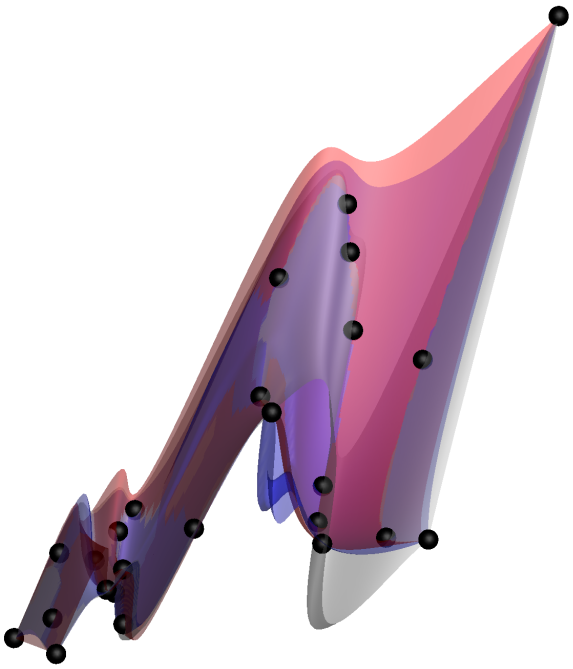

Figure 10: Plot of truth model (red) as well as the global kriging (white) and ALOS with 2 clusters (blue) using the 25 initial points.

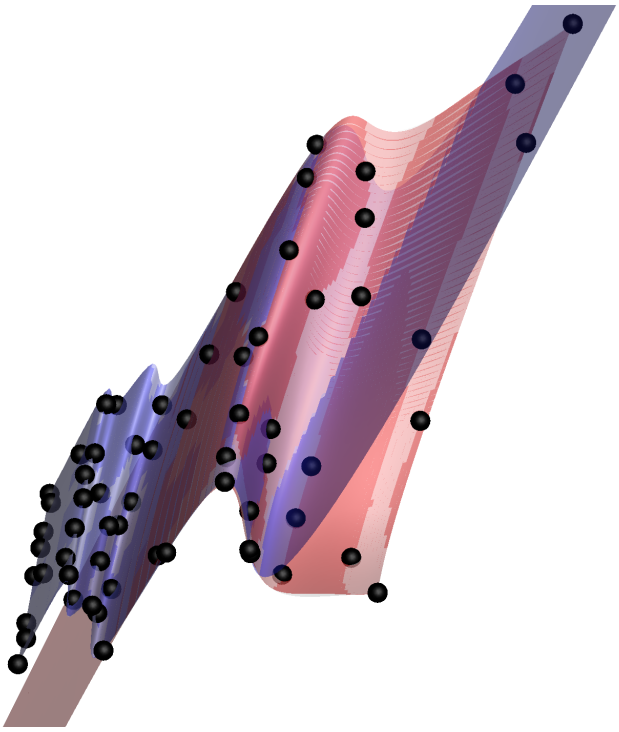

Figure 11: Plot of truth model (white) as well as the two local surrogate models (same color code as in Figure 9) using 63 points.

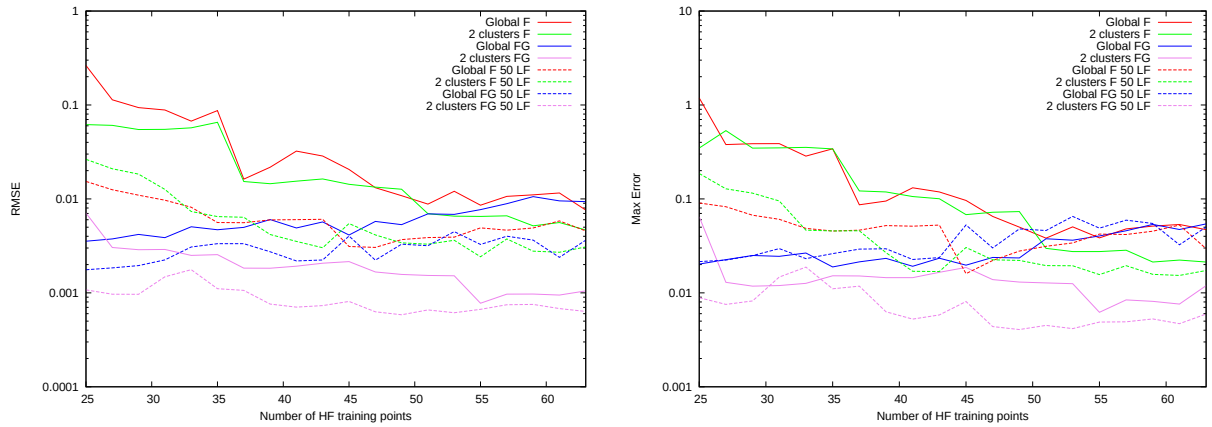

Figure 12: RMSE (left) and maximum error (right) for ALOS with 2 clusters and a global surrogate both enhanced with low-fidelity data as a function of number of high-fidelity training points in two dimensions.

in same color). The LF data is governed by

$$g_l(x, y) = (g - 2.0 + x + y)/(5.0 + 0.25x + 0.5y)$$

which implies again that the employed linear multiplicative bridge function can recover the HF data exactly. Again, 50 LF training points are used initially where this time 25 are dictated by the HF training locations and the remaining 25 points are picked via latin hypercube sampling subject to a distance constraint. Again, whenever a HF point is added via the dynamic training point algorithm the corresponding LF point is added to the set as well.

Three-dimensional Analytic Test Function

In Figure 13 the performances of ALOS with two clusters and a global kriging surrogate for a 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. The function is given by

$$h(x, y, z) = \sin[21(x - 0.9)^4] \cos[2(x - 0.9)] + (x - 0.7)/2 + 2y^2 \sin[xy] + 3z^3 \sin[xyz]$$

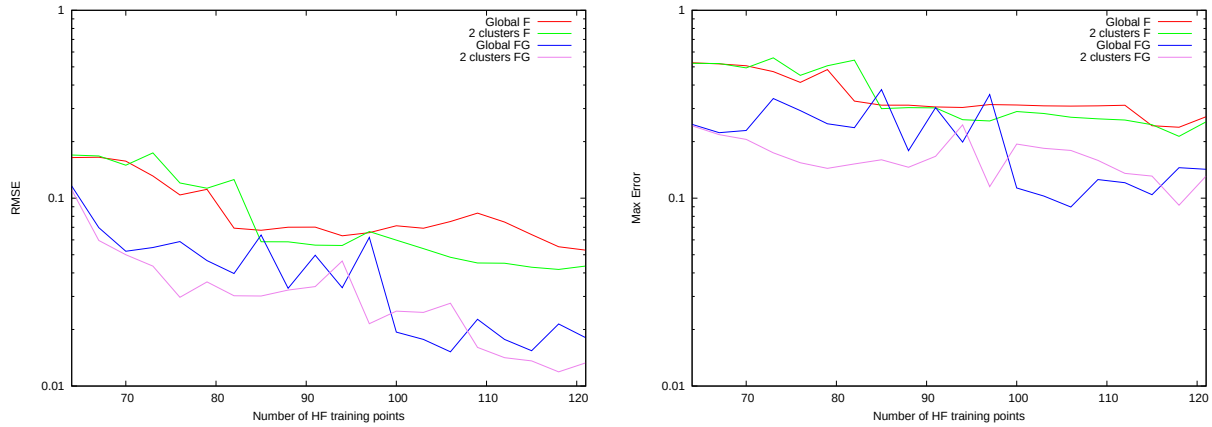

Figure 13: RMSE (left) and maximum error (right) between the three-dimensional truth model and ALOS with 2 clusters as well as a global surrogate as a function of number of high-fidelity training points.

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 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 second (red) cluster (which also features the largest θ values as shown in Table 2).

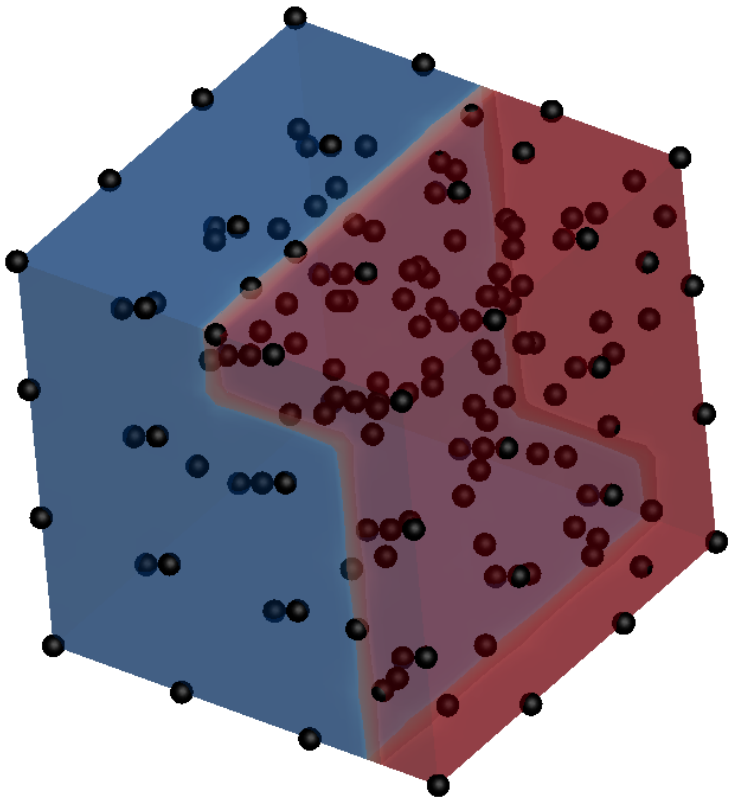

Figure 14: Plot of $\Pi_k(x)$ using the final 121 points. Different colors correspond to different clusters. The highest color intensity within each cluster corresponds to $\Pi_k(x) = 1$.

The local kriging and gradient-enhanced kriging (GEK) models have again different optimum length scales $\theta = [\theta_x, \theta_y, \theta_z]$ in the partitioned input space as the global kriging and GEK as shown in Table 2, again suggesting that the ALOS approach has the potential to be better at modeling the different characteristics in the function profile.

Number of clusters	Length scales, θ
kriging	
1	[0.27, 0.18, 0.28]
2	[0.09, 0.10, 0.19], [0.24, 0.20, 0.24]
GEK	
1	[0.84, 0.51, 0.33]
2	[0.81, 0.44, 0.26], [0.51, 0.29, 0.18]

Table 2: Optimum length scales for kriging and GEK using 1 and 2 clusters in three dimensions using 64 training points.

Figure 15 shows the performance with enhancement through low-fidelity data. The LF data tends to reduce the errors for up to 100 high-fidelity training points (compare dashed and solid lines in same color).

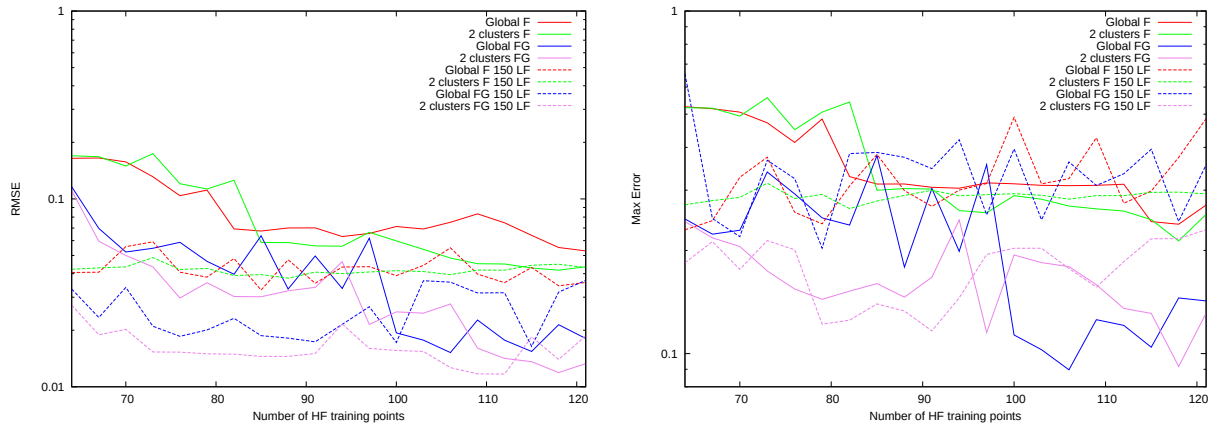

Figure 15: RMSE (left) and maximum error (right) for ALOS with 2 clusters and a global surrogate both enhanced with low-fidelity data as a function of number of high-fidelity training points in three dimensions.

The LF data is governed by

$$h_l(x, y, z) = (g - 2.0 + x + y + z)/(5.0 + 0.25x + 0.5y - 0.75z)$$

which implies again that the employed linear multiplicative bridge function can recover the HF data exactly. This time 150 LF training points are used initially where 64 are dictated by the HF training locations and the remaining 86 points are picked via latin hypercube sampling subject to a distance constraint. Again, whenever a HF point is added via the dynamic training point algorithm the corresponding LF point is added to the set as well.

4.2 Transonic CFD Example

To demonstrate the true potential of the developed dynamic variable-fidelity locally optimized surrogate model another transonic CFD example is considered, which this time 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 ($0.6 \leq M_\infty \leq 1.2$), angle of attack ($0° \leq \alpha \leq 4°$) and thickness to chord ratio ($4\% \leq tc \leq 12\%$) are studied here. The high-fidelity level is given by RANS simulations using Fun3d [40] with the Spalart-Allmaras turbulence model [41]. The computational mesh consists of 37,014 hexahedral elements. An "exact" database is obtained from RANS flow solves on a Cartesian mesh of $31 \times 9 \times 9 = 2511$ equispaced nodes and is used for comparisons. The low-fidelity level is given by the finite-volume solver SU2 [29, 30] 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 16 compares the two fidelity levels in the domain of interest. One can see that the low-fidelity trends match the high-fidelity one which is encouraging for the use of a variable-fidelity approach.

In Figure 17 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 (63) 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 latin hypercube sampling 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 VFM yields more accurate results

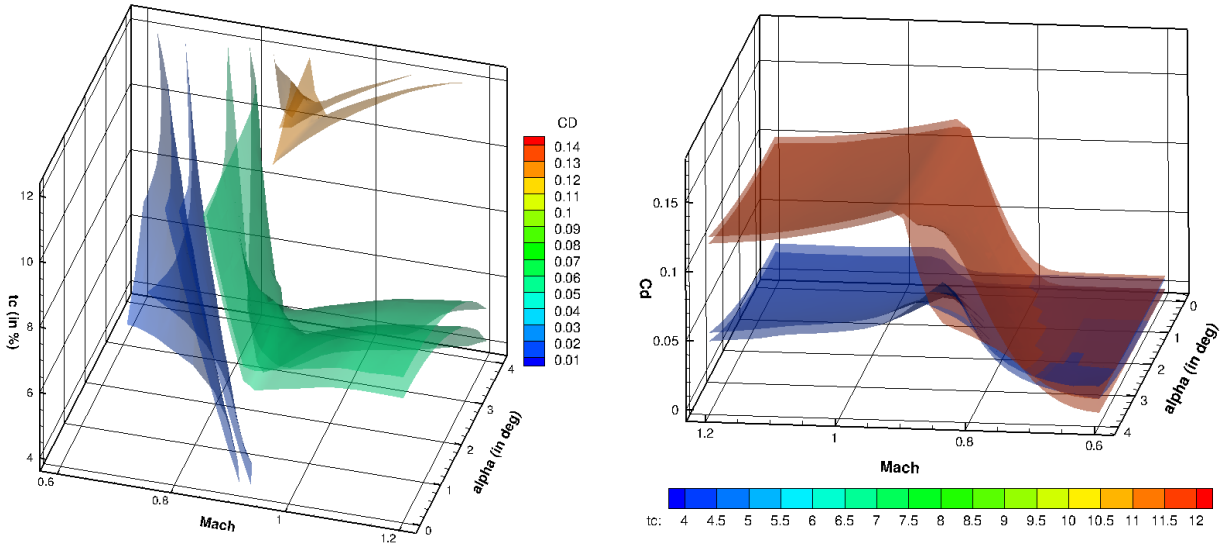

Figure 16: Left: Three isosurfaces of the two fidelity levels. Right: Contours of C_d for the lowest and highest thickness to chord ratio.

compared to using the high-fidelity data alone 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 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.

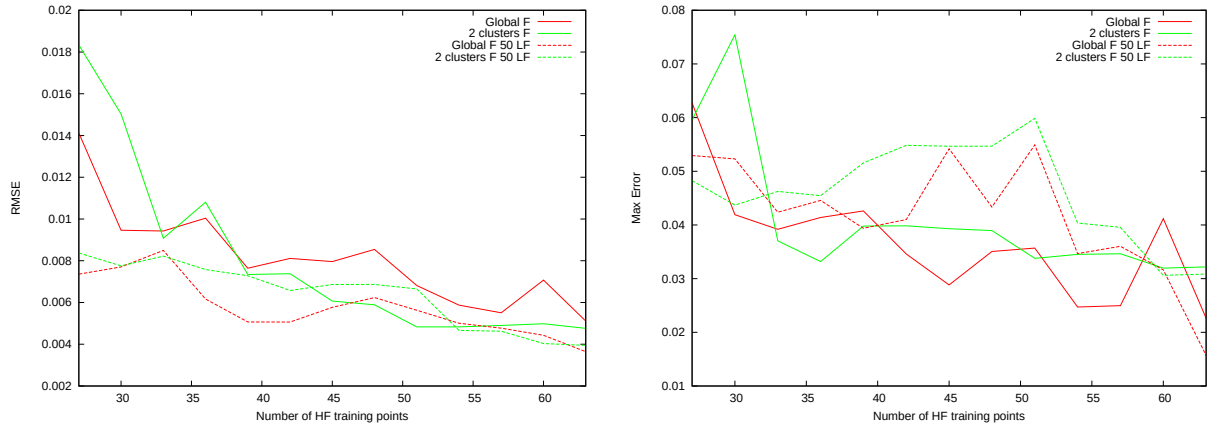

Figure 17: RMSE (left) and maximum error (right) for ALOS with two clusters and a global surrogate both enhanced with low-fidelity data as a function of number of high-fidelity training points.

Figure 18 shows the exact and global VFM kriging isosurfaces for three different C_d values at the end of the simulation (using 63 HF points and 86 LF points). It can be inferred that the global VFM is in good agreement with the exact model and is able to capture the transonic behavior very well, demonstrating its ability to model non-smooth functions. One can also observe that the dynamic training point selection clustered HF points in the more varying transonic region as opposed to the relatively flat supersonic region.

Figure 19 shows the domains of the two clusters and the training point distribution again at the end of the simulation. One can observe that the dynamic training point algorithm placed new training points especially in the blue cluster which covers the steep flank due to changes in the Mach number.

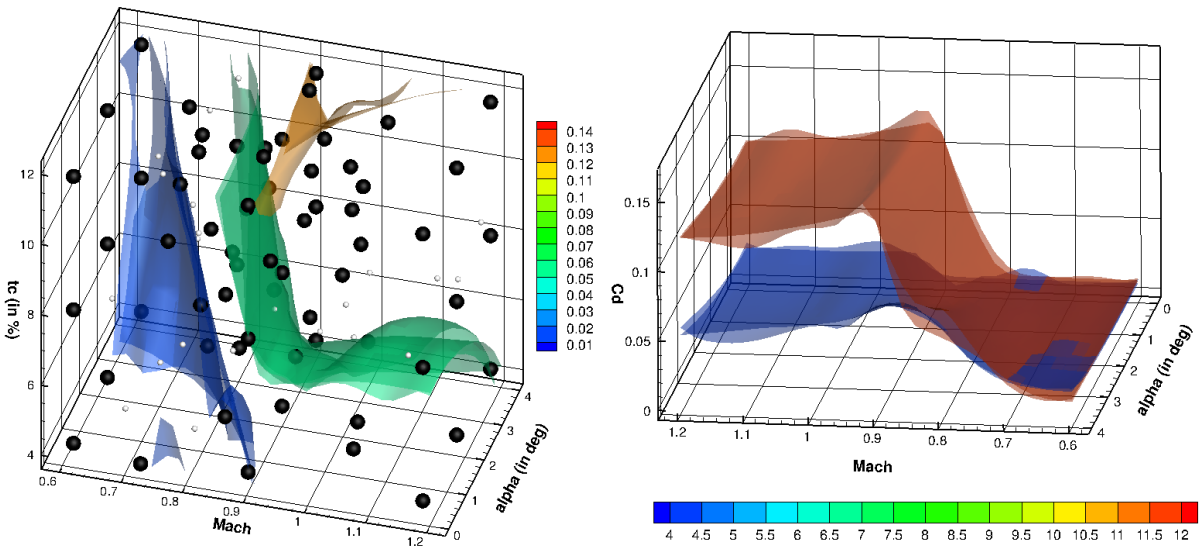

Figure 18: Left: Three isosurfaces of truth model and global VFM kriging. Right: Contours of truth model and global VFM kriging for the lowest and highest thickness to chord ratio.

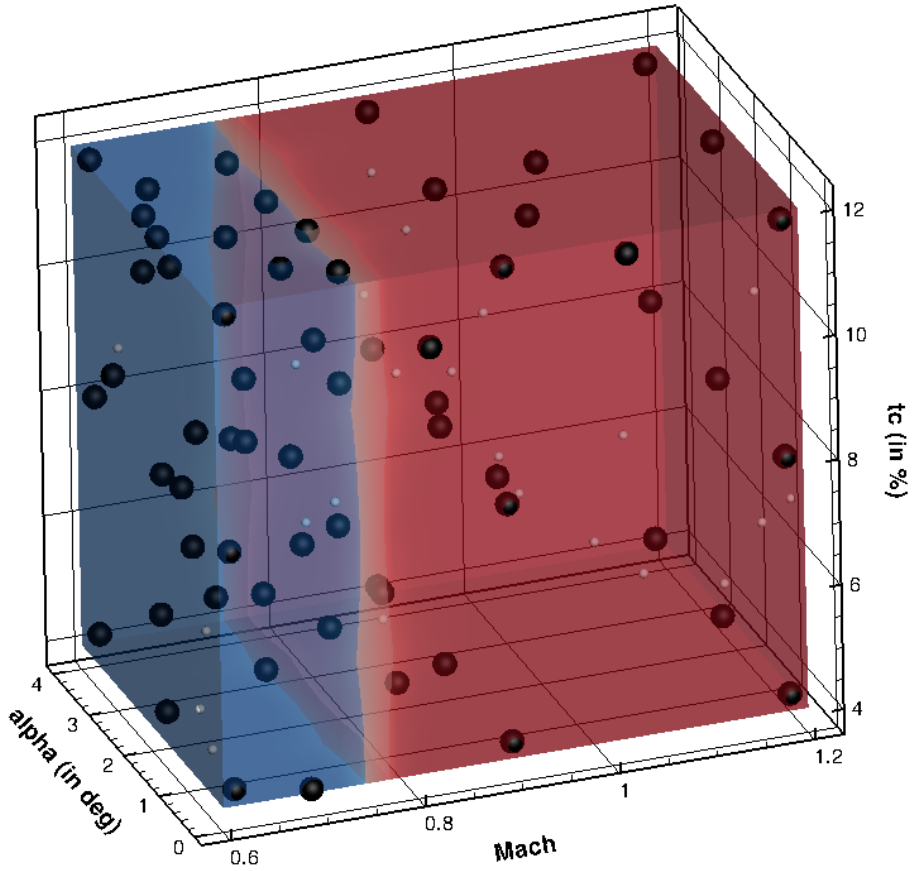

Figure 19: Plot of $\Pi_k(x)$ for the aerodynamic database using the final 63 points. Different colors correspond to different clusters. The highest color intensity within each cluster corresponds to $\Pi_k(x) = 1$.

5 Conclusion

The general goal of this work is to build highly accurate surrogate models at small overall computational cost assuming that obtaining high-fidelity training point information is the dominant cost factor. The strategy to achieve this is to enhance 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 surrogate models. Especially the latter two strategies were explained in detail in this paper and all four strategies were applied to one-, two- and three-dimensional analytic test functions demonstrating their potential. The developed dynamic variable-fidelity locally optimized surrogate models was also used for aerodynamic database creations 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 variations on the drag coefficient are of interest.

Acknowledgments

This work was supported by the US Air Force Summer Faculty Fellowship Program and the US Air Force Research Laboratory (AFRL). The second author acknowledges support of US Air Force Office of Scientific Research (grant 03VA01COR, Dr. Jean-Luc Cambier, Program Officer).

References

[1] S. A. Burton, E. J. Alyanak, and R. M. Kolonay. Efficient Supersonic Air Vehicle Analysis and Optimization Implementation using SORCER. AIAA Paper, 2012-5520, 2012.

[2] C. M. Meckstroth, E. J. Alyanak, N. J. Lindsley, M. Gabbard, and A. Hunton. Aerodynamic Modeling Techniques for Efficient Supersonic Air Vehicle Multidisciplinary Design Optimization. AIAA Paper, 2014-3254, 2014.

[3] S. Isukapalli, A. Roy, and P. Georgopoulos. Efficient sensitivity/uncertainty analysis using the combined stochastic response surface method and automated differentiation: Application to environmental and biological systems. *Risk Analysis*, 20:591–602, 2000.

[4] N. Kim, H. Wang, and N. Queipo. Adaptive reduction of random variables using global sensitivity in reliability-based optimisation. *International Journal of Reliability and Safety*, 1:102–119, 2006.

[5] O. Roderick, M. Anitescu, and P. Fischer. Polynomial regression approaches using derivative information for uncertainty quantification. *Nuclear Science and Engineering*, 164, No. 2:122–139, 2010.

[6] D. G. Krige. A statistical approach to some basic mine valuations problems on the Witwatersrand. *Journal of the Chemical Metallurgical and Mining Engineering Society of South Africa*, 52 (6):119–139, 1951.

[7] N. Cressie. The Origins of Kriging. *Mathematical Geology*, 22, No. 3:239–252, 1990.

[8] J. R. Koehler and A. B. Owen. Computer Experiments. In *Handbook of Statistics, Ghosh, S., Rao, C.R., (Eds.)*. pp. 261-308, 1996.

[9] W. Yamazaki, S. Mouton, and G. Carrier. Efficient Design Optimization by Physics-Based Direct Manipulation Free-Form Deformation. AIAA Paper, 2008-5953, 2008.

[10] M. P. Rumpfkeil, W. Yamazaki, and D. J. Mavriplis. Uncertainty Analysis Utilizing Gradient and Hessian Information. Sixth International Conference on Computational Fluid Dynamics, ICCFD6, St. Petersburg, Russia, July, 2010.

[11] H. S. Chung and J. J. Alonso. Using Gradients to Construct Cokriging Approximation Models for High-Dimensional Design Optimization Problems. AIAA Paper, 2002-0317, 2002.

[12] J. Laurenceau and P. Sagaut. Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging. *AIAA Journal*, 46, No. 2:498–507, 2008.

[13] J. Laurenceau and M. Meaux. Comparison of Gradient and Response Surface Based Optimization Frameworks Using Adjoint Method. AIAA Paper, 2008-1889, 2008.

[14] W. Yamazaki, M. P. Rumpfkeil, and D. J. Mavriplis. Design Optimization Utilizing Gradient/Hessian Enhanced Surrogate Model. AIAA Paper, 2010-4363, 2010.

[15] M. P. Rumpfkeil, W. Yamazaki, and D. J. Mavriplis. A Dynamic Sampling Method for Kriging and Cokriging Surrogate Models. AIAA Paper, 2011-883, 2011.

[16] K. Boopathy and M. P. Rumpfkeil. A Unified Framework for Training Point Selection and Error Estimation for Surrogate Models. *AIAA Journal*, 53:215–234, 2015.

[17] Q. Wang, P. Moin, and G. Iaccarino. A High-Order Multi-Variate Approximation Scheme for Arbitrary Data Sets. *Journal of Computational Physics*, 229, No. 18:6343–6361, 2010.

[18] P. A. Lachenbruch and M. R. Mickey. Estimation of error rates in discriminant analysis. *Technometrics*, 10:1–11, 1968.

[19] B. Efron. Estimating the error rate of a prediction rule: Some improvements on cross-validation. *Journal of the American Statistical Association*, 78:316–331, 1983.

[20] Z. H. Han, R. Zimmermann, and S. Goertz. On Improving Efficiency and Accuracy of Variable-Fidelity Surrogate Modeling in Aero-data for Loads Context. CEAS 2009 European Air and Space Conference, 2009.

[21] Z. H. Han, R. Zimmermann, and S. Goertz. A New Cokriging Method for Variable-Fidelity Surrogate Modeling of Aerodynamic Data. AIAA Paper, 2010-1225, 2010.

[22] W. Yamazaki and D. J. Mavriplis. Derivative-Enhanced Variable Fidelity Surrogate Modeling for Aerodynamic Functions. AIAA Paper, 2011-1172, 2011.

[23] Z. H. Han, S. Goertz, and R. Zimmermann. Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerospace Science and Technology*, doi:10.1016/j.ast.2012.01.006, 2012.

[24] N. M. Alexandrov, R. M. Lewis, C. R. Gumbert, L. L. Green, and P. A. Newman. Approximation and Model Management in Aerodynamic Optimization with Variable-Fidelity Models. *Journal of Aircraft*, 38(6):1093–1101, 2001.

[25] S. Choi, J. J. Alonso, I. M. Kroo, and M. Wintzer. Multi-fidelity design optimization of low-boom supersonic business jets. AIAA Paper, 2004-1530, 2004.

[26] S. G. Nash R. M. Lewis. A multigrid approach to the optimization of systems governed by differential equations. AIAA Paper, 2000-4890, 2000.

[27] S. Gano, J. E. Renaud, and B. Sanders. Variable fidelity optimization using a kriging based scaling function. AIAA Paper, 2004-4460, 2004.

[28] M. S. Eldred, A. A. Giunta, S. S. Collis, N. A. Alexandrov, and R.M. Lewis. Second-order corrections for surrogate-based optimization with model hierarchies. AIAA Paper, 2004-4457, 2004.

[29] F. Palacios, M. R. Colonno, A. C. Aranake, A. Campos, S. R. Copeland, T. D. Economon, A. K. Lonkar, T. W. Lukaczyk, T. W. R. Taylor, and J. J. Alonso. Stanford University Unstructured (SU2): An open-source integrated computational environment for multi-physics simulation and design. AIAA Paper, 2013-0287, 2013.

[30] F. Palacios, T. D. Economon, A. C. Aranake, S. R. Copeland, A. K. Lonkar, T. W. Lukaczyk, D. E. Manosalvas, K. R. Naik, A. S. Padron, B. Tracey, A. Variyar, and J. J. Alonso. Stanford University Unstructured (SU2): An open-source analysis and design technology for turbulent flows. AIAA Paper, 2014-0243, 2014.

[31] ZONA TECHNOLOGY, INC. ZEUS Version 3.3L User Manual. Technical report, 2013.

[32] D. L. Clark and H.-R. Bae. Engineering Design Exploration Utilizing Locally-Optimized Covariance Kriging. AIAA Paper, 2015-1372, 2015.

[33] D. L. Clark, H.-R. Bae, K. Gobal, and R. Penmetsa. Enineering Design Exploration utilizing Locally Optimized Covariance Kriging. *AIAA Journal*, 2016.

[34] R. P. Liem, C. A. Mader, and J. R. R. A. Martins. Surrogate models and mixtures of experts in aerodynamic performance prediction for aircraft mission analysis. *Aerospace Science and Technology*, 43:126–151, 2015.

[35] B. Tang, M. I. Heywood, and M. Shepherd. Input partitioning to mixture of experts. in: Proceedings of the 2002 International Joint Conference on Neural Networks, vol. 1, pp. 227-232, 2002.

[36] D. Reynolds. *Gaussian mixture models, in: Encyclopedia of Biometric Recognition*. Springer, 2008.

[37] S. Masoudnia and R. Ebrahimpour. Mixture of experts: a literature survey. *Artificial Intelligence Review*, pages 1–19, 2012.

[38] A. K. Jain. Data clustering: 50 years beyond K-means. *Pattern Recognition Letter*, 31:651–666, 2010.

[39] G. McLachlan and D. Peel. The EMMIX Algorithm for the Fitting of Normal and t-Components. *Journal of Statistical Software*, 4(2), 1999.

[40] R.T. Biedron J. Carlson, J.M. Derlaga, P.A. Gnoffo, D.P. Hammond, W.T. Jones, B. Kleb, E. Lee-Rausch, E.J. Nielsen, M.A. Park, C.L. Rumsey, J.L. Thomas, and W.A. Wood. FUN3D Manual: 12.7. NASA Technical Memorandum 218761, 2015.

[41] P. R. Spalart and S. R. Allmaras. A One-Equation Turbulence Model for Aerodynamic Flows. AIAA Paper, 92-0439, 1992.