A new approach to variable-fidelity, gradient-enhanced surrogate modeling using polynomial chaos expansions is presented. An advantage of the new approach is the simultaneous determination of least squares-optimal coefficients for additive and multiplicative low-fidelity corrections as well as the model of interest. This implies that the correction type need not be chosen a priori. The new approach is compared to variable-fidelity, gradient-enhanced kriging also using a hybrid additive-multiplicative bridge. Both methods are tested using two analytic test functions as well as the generation of an aerodynamic database of a vehicle in the transonic and supersonic regimes employing Cart3D and a linearized panel method. Surrogate training points are dynamically selected to reduce model error. Kriging produces greater reductions in root-mean-square-error (RMSE) than polynomial chaos in general, except when the underlying function is an exact polynomial. Both multifidelity approaches achieve reductions in RMSE similar to their monofidelity counterparts at reduced computational expense. The polynomial chaos method benefits more from a multifidelity approach than kriging in terms of error reduction versus computational cost. Forming variable-fidelity surrogates using function values alone tends to perform better than when using gradients in the low-dimensional cases considered here. On the whole, variable-fidelity kriging outperforms variable-fidelity polynomial chaos and monofidelity kriging, and has more favorable properties of training point selection and interpolatory/extrapolatory behavior near the domain boundaries compared to polynomial chaos.

Nomenclature

\begin{align*}
    d_{ij} & \quad \text{Distance between points } i \text{ and } j \\
    f_k(x) & \quad \text{Function of fidelity } k \\
    LR & \quad \text{Low-fidelity to high-fidelity training point ratio} \\
    M & \quad \text{Design space dimension, Mach number} \\
    N & \quad \text{Number of points} \\
    P, Q, R & \quad \text{Terms in polynomial chaos expansion} \\
    R(\cdot, \cdot) & \quad \text{Kriging correlation} \\
    x & \quad \text{Design vector } (x_1, x_2, \ldots, x_M) \\
    x^{(i)} & \quad \text{Design point } i \\
    \alpha(x) & \quad \text{Multiplicative correction function} \\
    \delta(x) & \quad \text{Additive correction function} \\
    Z_k(x) & \quad \text{Stationary random process of fidelity } k \\
    \theta_m & \quad \text{Radial basis function parameter in dimension } m \\
    \mu_k & \quad \text{Mean of fidelity } k \\
    \sigma_k^2 & \quad \text{Variance of fidelity } k \\
    \Phi & \quad \text{Spatial correlation function} \\
    \Psi_k & \quad \text{Polynomial basis function}
\end{align*}

*Ph.D. Student, Dept. of Mechanical and Aerospace Engineering, brysond1@udayton.edu, Member AIAA
†Assistant Professor, Dept. of Mech. and Aerospace Engineering, Markus.Rumpfkeil@udayton.edu, Senior Member AIAA
I. Introduction

Fidelity is the degree to which a method of prediction, analysis, or experimentation accurately reproduces an effect of interest. A pervasive challenge in aerospace design is the constant balance between performing expensive analyses or experiments of high enough fidelity to accurately predict vehicle performance versus using often less accurate tools providing computational expediency. In practice, determining what is “high enough fidelity”, or even defining “fidelity” in a working sense, is difficult \textit{a priori}. The fidelity of a particular method may also vary across the design space as physical regimes (e.g., subsonic, transonic, supersonic) differ.

In some instances, data of different fidelities may be combined as an ensemble in which all sources are treated as being equally reliable. Here, datasets are used in a hierarchy in which one dataset is believed to be more trustworthy (termed high-fidelity), and less trustworthy datasets (low-fidelity) are used to augment the high-fidelity data via some correction process. Examples of high-fidelity versus low-fidelity data sources are provided in Table 1. Two of the underlying assumptions in a hierarchical multifidelity process are that the low-fidelity model is sufficiently accurate to provide useful trends, and that the cost of the low-fidelity evaluation is less than that of the high-fidelity evaluation (the collection of less accurate data at greater expense would truly be a poor choice). Thus, variable-fidelity methods seek to leverage intensively-sampled low-fidelity data to predict the trend of the response and reserve high-fidelity evaluations to correct inaccuracies in the low-fidelity data.

Table 1: Some possible combinations of variable-fidelity methods

<table>
<thead>
<tr>
<th>High-Fidelity Model</th>
<th>Low-Fidelity Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental data</td>
<td>CFD results</td>
</tr>
<tr>
<td>Navier-Stokes</td>
<td>Euler</td>
</tr>
<tr>
<td>Euler</td>
<td>Panel methods</td>
</tr>
<tr>
<td>Finer mesh CFD results</td>
<td>Coarser mesh CFD results</td>
</tr>
<tr>
<td>Fully converged solutions</td>
<td>Partly converged solutions</td>
</tr>
<tr>
<td>Detailed geometry</td>
<td>Simplified geometry</td>
</tr>
</tbody>
</table>

Corrections to the low-fidelity response may take the form of input space mappings, or, more commonly, output space mappings (commonly referred to as bridge, correction, or calibration functions). These bridge functions typically take the form of low-order polynomial additive or multiplicative corrections, or a combination of the two. A difficulty of this approach is knowing \textit{a priori} which form to use for a particular fidelity pairing and in what order to apply the corrections.

Two popular methods for building surrogate models are kriging and polynomial chaos expansion (PCE). The formulations of these methods are summarized in Subsections II.A and II.B. Both are amenable to enhancement through the addition of gradient and Hessian information and through multifidelity. For clarity, we use kriging and PCE to refer to their respective variable-fidelity, gradient-enhanced implementations rather than their ordinary forms.

With kriging, there are two general approaches to variable fidelity. The first is an all-at-once approach, in which the parameters of the (typically additive) bridge function and the surrogate model of interest are determined simultaneously, such as the method taken by Keane. Alternatively, a multi-step approach may be taken in which low-fidelity corrections are calculated in a separate step from the model of interest. Han, Görtz, and Zimmermann present an approach in which a kriging model of the low-fidelity data is generated, followed by determination of the best additive and multiplicative corrections between the low-fidelity surrogate and the high-fidelity data.

With polynomial chaos expansions, the multistep approach is generally followed. Ng and Eldred present...
a method where separate expansions are developed for the low-fidelity data as well as additive and multiplicative corrections. The three surrogates are then combined to approximate the high-fidelity response. Alternatively, Shah et al.\textsuperscript{10} determine the least squares-optimal linear multiplicative and constant additive correction to the low-fidelity data. Then, a single PCE surrogate for the high-fidelity data is created using only the transformed low-fidelity data.

In this paper, we present an all-at-once approach to generating multifidelity PCE’s with hybrid additive-multiplicative bridge functions of arbitrary order. The multifidelity approach is further augmented by the addition of gradients for both the high and low fidelities. This new approach is compared to a gradient-enhanced, multifidelity kriging approach following the method of Han, Görtz, and Zimmermann.\textsuperscript{8} Demonstrations are made using analytic test functions in two and three dimensions, and on a three-dimensional test problem combining Cartesian Euler CFD with a panel method for transonic lift coefficient prediction.

For the aerodynamic application, a tailless, lambda-wing vehicle, illustrated in Figure 1(a), is presented. This type of configuration has been the subject of many studies at the Air Force Research Laboratory\textsuperscript{11–14} and is of interest due to potential savings in weight and drag due to eliminating the empennage. The lambda planform is selected to balance subsonic and supersonic performance. The outboard wing section increases subsonic performance by increasing aspect ratio, while the inboard section provides high speed performance. Flying a tailless vehicle, however, presents challenges in trim and stability, and accurate prediction of the aerodynamic forces and moments is critical. The accurate prediction of aerodynamic performance is further exacerbated by the fact that such a vehicle must both dwell in and fly through the transonic regime, for which aerodynamic modeling is known to be difficult.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{lambda_wing.png}
\caption{Geometry of lambda wing vehicle}
\end{figure}

The airfoil used for the center body of the vehicle is shown in Figure 1(b), and was selected as a result of prior research efforts to balance vehicle performance across a variety of conditions. To provide geometric variation, the camber line of the airfoil is preserved, and the thickness-to-chord ratio is scaled from 0.06 to 0.12. The aerodynamic performance of the lambda planform in the transonic to supersonic regimes (Mach number from 0.8 to 1.6) is considered for a range of angles of attack from $-1^\circ$ to $6^\circ$.

\section{Surrogate Models}

\subsection{Multifidelity, Gradient-Enhanced Polynomial Chaos Expansion}

Polynomial chaos expansion (PCE) is a method by which responses found through stochastic processes (chaos) are represented by a series of orthogonal polynomials.\textsuperscript{9, 15} Here, however, the inputs and aerodynamic responses are treated as deterministic, and the PCE is used as a polynomial regression. A number of different polynomial bases may be used to produce optimal convergence of the surrogate to the true function, depending on the distribution of the input parameters.\textsuperscript{15} Here, Legendre polynomials form the bases, corresponding to uniform parameter distributions.\textsuperscript{15}

Boopathy and Rumpfkeil\textsuperscript{6} propose a framework for gradient- and Hessian-enhanced PCE, based on the work of Roderick, Anitescu, and Fischer\textsuperscript{16} and Li \textit{et al.}\textsuperscript{5} with dynamic training point selection to iteratively
converge an estimate of the global error. Their work is extended to admit the use of multiple analysis fidelities in a single model. Previous multifidelity PCE implementations take a multi-step approach, separating the estimation of corrective functions from the fitting of the model of interest.\textsuperscript{9,10} Typically, corrections are of the additive or multiplicative forms, and one must choose \textit{a priori} what form the correction should take or estimate a blend of both forms.\textsuperscript{9,10} The approach here differs in that only a single surrogate model is formed, simultaneously determining additive and multiplicative corrections to the low-fidelity data to best approximate the high-fidelity function in a least-squares sense.

In a traditional PCE, a function is approximated by the series

\[ \hat{f}(x) = \sum_{k=0}^{P} \beta_k \Psi_k(x), \]  

(1)

where \( \Psi_k(x) \) are the selected multidimensional bases formed by

\[ \Psi_k(x) = \prod_{i=1}^{M} \Psi_{a_k}^i(x_i). \]  

(2)

The multi-index \( a_k^{15} \) indicates the order of the one-dimensional polynomial in the \( i \)th direction for the \( k \)th term in the total expansion. The number of terms \( P \) in the expansion may not be chosen arbitrarily, but rather is a combination of the design space dimensionality, \( M \), and the polynomial order, \( p \),

\[ P = \left( M + p \right). \]  

(3)

The coefficients \( \beta_k \) may be calculated non-intrusively by the collocation method (as is done here) or by the Galerkin spectral projection method.\textsuperscript{9} The gradient and Hessian of \( \Psi(x) \) are readily available\textsuperscript{5,6} to admit the use of derivative information in the surrogate modeling procedure.

For the inclusion of multifidelity data in the model fitting, it is assumed that the high-fidelity function is matched by multiplicative and additive corrections to the low-fidelity model,

\[ f_h(x) = f_l(x) + \alpha(x) f_l(x) + \delta(x). \]  

(4)

These corrections themselves may be represented by PCEs. Substituting the PCE representation from Equation (1) for the high-fidelity and corrective functions and rearranging Equation (4) yields

\[ \hat{f}_h(x) = \sum_{k=0}^{P} \beta_k \Psi_k(x), \]  

(5)

\[ \hat{f}_l(x) = \sum_{k=0}^{P} \beta_k \Psi_k(x) - f_l(x) \sum_{k=0}^{Q} \alpha_k \Psi_k(x) - \sum_{k=0}^{R} \delta_k \Psi_k(x). \]  

(6)

If the low-fidelity model correlates well with the high-fidelity model, then the polynomial order of the correction terms should be small compared to the order of the high-fidelity PCE.

Evaluating the high- and low-fidelity functions at their respective collocation points, \( x_h \) and \( x_l \), the PCE coefficients are found via least-squares regression,

\[ \begin{pmatrix} \beta_0 & \ldots & \beta_P & \alpha_0 & \ldots & \alpha_Q & \delta_0 & \ldots & \delta_R \end{pmatrix}^T = (X^T X)^{-1} X^T y, \]  

(7)

where the \((N_h + N_l)(M + 1) \times (P + Q + R)\) block matrix \( X \) is given by

\[ X = \begin{bmatrix} B^H & 0 & 0 \\ B^L & A^L & \Delta^L \end{bmatrix}, \]  

(8)
with

\[
B^H = \begin{bmatrix} X_p^{1,H} \\ \vdots \\ X_p^{N_h,H} \end{bmatrix}, \quad B^L = \begin{bmatrix} X_p^{1,L} \\ \vdots \\ X_p^{N_h,L} \end{bmatrix}, \quad A^L = \begin{bmatrix} Z_Q^{1,L} \\ \vdots \\ Z_Q^{N_h,L} \end{bmatrix}, \quad \Delta^L = \begin{bmatrix} X_R^{1,L} \\ \vdots \\ X_R^{N_h,L} \end{bmatrix},
\]

where

\[
X_j^{i,H} = \begin{bmatrix} \frac{\partial}{\partial x_1} \left( \psi_0 \left( x_h^{(i)} \right) \right) & \cdots & \frac{\partial}{\partial x_l} \left( \psi_j \left( x_h^{(i)} \right) \right) \\
\vdots & & \vdots \\
\frac{\partial}{\partial x_{M_l}} \left( \psi_0 \left( x_h^{(i)} \right) \right) & \cdots & \frac{\partial}{\partial x_{M_l}} \left( \psi_j \left( x_h^{(i)} \right) \right) 
\end{bmatrix},
\]

\[
X_j^{i,L} \text{ is similarly formed from the low-fidelity training points } x_j^{(i)}, \text{ and}
\]

\[
Z_Q^{i,L} = \begin{bmatrix} f_l \left( x_l^{(i)} \right) \psi_0 \left( x_l^{(i)} \right) & \cdots & f_l \left( x_l^{(i)} \right) \psi_Q \left( x_l^{(i)} \right) \\
\vdots & & \vdots \\
f_l \left( x_l^{(i)} \right) \psi_0 \left( x_l^{(i)} \right) & \cdots & f_l \left( x_l^{(i)} \right) \psi_Q \left( x_l^{(i)} \right) 
\end{bmatrix}.
\]

The vector of training point responses is

\[
y = \left( Y^{1,H} \quad \cdots \quad Y^{N_h,H} \quad Y^{1,L} \quad \cdots \quad Y^{N_h,L} \right)^T,
\]

\[
Y^{i,H} = \left( f_h \left( x_h^{(i)} \right) \frac{\partial}{\partial x_1} \left( f_h \left( x_h^{(i)} \right) \right) & \cdots & \frac{\partial}{\partial x_{M_l}} \left( f_h \left( x_h^{(i)} \right) \right) \right)
\]

and the vector \( Y^{i,L} \) is similarly formed from low-fidelity responses.

II.B. Multifidelity, Gradient-Enhanced Kriging

The gradient-enhanced, multifidelity kriging used here is based on the kriging formulation of Yamazaki and Mavriplis with the generalized hybrid additive-multiplicative bridge function of Han, Görtz, and Zimmermann. The derivation provided by Yamazaki and Mavriplis is summarized here; for a more thorough treatment, the reader is referred to the original source.

For “ordinary” kriging (as opposed to “universal” kriging, where the underlying mean is a low-order polynomial), an interpolatory surrogate model is represented by

\[
\hat{f}_k \left( x \right) = \mu_k + Z_k \left( x \right),
\]

where \( \mu \) is the mean of the training data and \( Z \left( x \right) \) is a stationary random process. The subscript \( k \) indicates the fidelity level, as data may be correlated between fidelities as well as within a fidelity. The random process \( Z_k \left( x \right) \) satisfies the properties

\[
E \left( Z_k \left( x \right) \right) = 0,
\]

\[
\text{Var} \left( Z_k \left( x \right) \right) = \sigma_k^2,
\]

and

\[
\text{Cov} \left( Z_k \left( x^{(i)} \right), Z_n \left( x^{(j)} \right) \right) = \sigma_k^2 \sigma_n^2 R_{kn} \left( x^{(i)}, x^{(j)} \right).
\]
The correlation matrix, $R$, is the product of spatial correlation functions $\Phi$ in all dimensions evaluated for the distance between points,

$$R \left( x^{(i)}, x^{(j)} \right) = \prod_{m=1}^{M} \Phi \left( \theta_m, d_{ij}^m \right), \quad (18)$$

$$d_{ij}^m = \left| x^{(i)}_m - x^{(j)}_m \right|. \quad (19)$$

Here, Wendland’s C4 radial basis function\(^{17}\) is implemented,

$$\Phi \left( \theta_m, d_{ij}^m \right) = \begin{cases} 
(1 - \theta_m d_{ij}^m)^6 \left( 35(\theta_m d_{ij}^m)^2 + 18 \theta_m d_{ij}^m + 3 \right) & \text{for } d_{ij}^m \leq \frac{1}{\theta_m} \\
0 & \text{otherwise}
\end{cases} \quad (20)$$

The correlations within a fidelity level are the same as given in Equation (18), while the correlations between fidelity levels are multiplied by a relaxation factor to improve numerical conditioning when training points between fidelities are very near each other,

$$R_{kn} = \begin{cases} 
R, & k = n \\
\rho R, & k \neq n
\end{cases} \quad (21)$$

Here, $\rho$ is taken to be 0.9999. Following Yamazaki and Mavriplis\(^4\), the means and variances of the multifidelity system may be determined analytically, but the parameters $\theta_m$ must be determined numerically. Using direct co-kriging, Yamazaki and Mavriplis\(^4\) also provide for the inclusion of gradient information in the parameter determination by taking derivatives of the correlation function.

Implementing the generalized hybrid bridge function of Han, Görtz, and Zimmerman\(^8\), the low-fidelity data may be transformed (similar to Equation (4) aside from an additional term $f_l(x)$) to match the high-fidelity by

$$f_h(x) = \alpha(x) f_l(x) + \delta(x). \quad (22)$$

In this case, $\alpha(x)$ is a low-order polynomial regression (limited to first order here), and $\delta(x)$ is an additive bridge function, itself a kriging model. A multi-step approach is taken in which a kriging model of the low-fidelity data is generated, followed by simultaneous estimation of the coefficients for the multiplicative and additive corrections. Similar to the model’s means and variances, the coefficients $\alpha(x)$ may be determined analytically, and updated with each iteration during the estimation of $\theta_m$.\(^8\) The final high-fidelity approximation is then taken to be the combination of the low-fidelity kriging, multiplicative polynomial regression, and additive kriging bridge functions.

II.C. Dynamic Training Point Selection

Following the framework of Boopathy and Rumpfkeil\(^6\), dynamic training point selection reduces the global error of the high-fidelity surrogate model. The root mean square error (RMSE) of the surrogate relative to the high-fidelity function evaluated on a Cartesian grid is the measure of global error,

$$RMSE = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} \left( f_h(x^{(i)}) - \tilde{f}_h(x^{(i)}) \right)^2}, \quad (23)$$

where $N_t$ is the total number of Cartesian points in which the truth model is evaluated. To identify training points to be added, the global surrogate model (PCE or kriging) is compared to local surrogate models generated using multivariate interpolation and regression (MIR) over a subset of the training points. MIR provides a surrogate model based on Taylor series expansions using function values and gradients.\(^{18}\) MIR typically has excellent accuracy in approximating the actual function, however, it scales poorly with the number of training points and dimensions. Hence, we restrict the use of MIR to local surrogate models for comparison. Boopathy and Rumpfkeil\(^6\) demonstrated that the discrepancy between the global model and MIR local surrogate model has similar convergence properties to the actual RMSE error.
The steps below define the iterative process of creating a surrogate model using dynamic training point selection. Slight differences arise between the processes for PCE and kriging due to the interdependence of polynomial order and number of training points for PCE. These differences are highlighted. Figure 2 illustrates several iterations of the dynamic sampling routine.

1. Initialization: The process is initialized using high-fidelity function evaluations (and gradients, if desired) at the center and corner points of the domain (i.e., five points in two-dimensions, nine points in three-dimensions).

   Kriging: In addition to the high-fidelity training points, the total desired number of low-fidelity training points is also added using Latin hypercube sampling (LHS).

   PCE: As polynomial order is iteratively increased, low-fidelity and high-fidelity training points are added in a user-specified low-fidelity ratio, \( LR = \frac{N_l}{N_h} \). In addition to the high-fidelity training points, \( N_l = LR \cdot N_h \) low-fidelity points are added.

2. Surrogate model generation: The kriging and PCE models are built based on the current sets of training points.

   Kriging: The kriging model form is fixed, and the number of training points may be chosen arbitrarily.

   PCE: The initial PCE order is chosen to be the highest order that may be determined using the available data. Note that in the multifidelity case, we have constrained the additive and multiplicative corrections to be constantly first-order (though their formulation holds for arbitrary order). For example, using an over-sampling ratio of two, a monofidelity linear model in two dimensions requires six pieces of information (function values and gradients). Thus, if only function values are used, five pieces of information are available, and the starting polynomial order is zero (requiring only two pieces of information). A similar multifidelity linear model with linear additive and multiplicative corrections requires eighteen pieces of information (six each for the model of interest and both of the corrections). Using function values only and a low-fidelity ratio of two, fifteen pieces of information are available, and the resulting model order is again zero (requiring fourteen pieces of information – two for the model of interest and six each for the corrections).

3. RMSE evaluation: The RMSE of the global surrogate models is evaluated as previously described.

4. Dynamic training point selection: On the order of \( 10^4 \) candidate training points are generated using LHS. Local MIR surrogate models are generated for each of the candidate points using \( N_{CP} \) nearest high-fidelity training points. These local models require no additional function evaluations, and are easily parallelized. Furthermore, when the number of high-fidelity training points is not greater than...
The process only requires a single MIR surrogate. The candidate points are sorted in descending order based on the absolute discrepancy between the global (PCE or kriging) and local surrogate predictions,

\[ \delta^{(i)} = \left| \hat{f}_{\text{global}} \left( \mathbf{x}^{(i)} \right) - \hat{f}_{\text{local}} \left( \mathbf{x}^{(i)} \right) \right|. \tag{24} \]

Next, the distances \( d_{ij} \) of each candidate point \( \mathbf{x}^{(i)} \) from its nearest existing training point \( \mathbf{x}^{(j)} \) are calculated,

\[ d_{ij} = \left\| \mathbf{x}^{(i)} - \mathbf{x}^{(j)} \right\|_2, \tag{25} \]

with the mean distance denoted \( \bar{d} \).

Next, the candidates are cycled through from largest discrepancy to smallest. The candidate with the largest discrepancy and distance \( d_{ij} \geq c \bar{d} \), is accepted where initially \( c = 1.1 \). After a candidate is accepted, all the distances are updated to reflect the new training point, and the process is repeated until the desired number of additional training points are identified. If no candidate satisfies the distance constraint, \( c \) is reduced by 2\% until a candidate is found.

Kriging: In kriging, \( N_{\text{add}} \) high-fidelity points are added each iteration.

PCE: In PCE, the polynomial order is increased by one each iteration. The number of additional pieces of information required to form a model of that order is determined, and low- and high-fidelity points are added according to the low-fidelity ratio. Based on initial studies, performance is most reliable when some high- and low-fidelity points are coincidental for calculating the bridge functions. Thus, using the above procedure, accepted candidates are duplicated as both high- and low-fidelity points until the number of additional high-fidelity points is exhausted. Then, accepted candidate points are added as low fidelity until the required number of points is reached.

5. Iteration and termination: The function (and gradient if desired) values of the newly added training points are evaluated, and the process is repeated from step 2. The process may be terminated when an evaluation budget is exhausted, or the error or discrepancy converges to an acceptable limit.

### III. Analytic Test Functions

#### III.A. Test Functions and Fidelities

To test the multifidelity frameworks, two analytic test functions are implemented. The first is the two-dimensional Rosenbrock function,

\[ f_h(x_1, x_2) = (1 - x_1)^2 + 100 (x_2 - x_1^2)^2. \tag{26} \]

This function is characterized by a narrow, gently sloped, U-shaped region with a single minimum at \((x, y) = (1, 1)\). Globally, it is easy to approximate the overall quartic shape, but, locally, it is difficult to capture the behavior near the optimum. The second is the three-dimensional Runge function,

\[ f_h(x_1, x_2, x_3) = \frac{1}{1 + x_1^2 + x_2^2 + x_3^2}. \tag{27} \]

This function is characterized by concentric spheres about the origin.

To form low-fidelity test functions, each of the analytic functions is transformed by linear additive and multiplicative factors,

\[ f_l(x_1, \ldots, x_M) = \frac{f_h(x_1, \ldots, x_M) - 4.0 - \sum_{i=1}^{M} 0.5x_i}{3.0 + \sum_{i=1}^{M} 0.25x_i}. \tag{28} \]

In this manner, the linear corrections to the low-fidelity data should fully recover the high-fidelity function. These test functions are illustrated in Figure 3.
III.B. Comparison of Mono- and Multifidelity Surrogate Models

The RMSE’s of the PCE and kriging models for the Rosenbrock function are shown in Figure 4. For kriging, the RMSE may be plotted versus the number of high-fidelity training points alone. However, for PCE, the number of training points may not be selected independently of polynomial order (versus kriging, where the number of training points is chosen arbitrarily). Thus, a matrix of figures is provided correlating RMSE with PCE order and number of high- and low-fidelity training points. The scales of the figures are matched so the reader may easily compare kriging versus PCE from left to right. The reader may also make correlations between PCE RMSE, order, and training point count by tracing a rectangle between the four PCE plots.

For the Rosenbrock function, training points are added at a low- to high-fidelity ratio (LR) of 4:1. This ratio is selected based on preliminary experiments using the aerodynamic prediction presented in Section IV. For the Rosenbrock function, other ratios are not tested because PCE captures the exact forms of the corrective functions and high-fidelity model once a fourth-order representation is attained, regardless of the low-fidelity ratio.

Initially, using samples in only the corners and center of the domain, monofidelity PCE and kriging do a comparable job of approximating the Rosenbrock function. However, once the PCE obtains enough points to achieve a fourth-order representation, it models the function exactly (to the precision of the stored truth data). Note that the PCE of the low-fidelity model does not drop as precipitously as for the high-fidelity model. This fact is a result of the low-fidelity model being a rational function without an exact polynomial representation. Despite the quality of the approximation of the low-fidelity model, the multifidelity formulation is able to match the truth model using less than a third as many high-fidelity function evaluations. This is due to the method identifying the transformation between the high- and low-fidelity functions, and using corrected low-fidelity data to achieve a higher polynomial order than would be achievable otherwise.

In comparison, the monofidelity kriging models only achieve approximately three orders of magnitude reduction in RMSE. This is not surprising as the kriging does not implement polynomial bases, which would exactly represent the function. By adding all the low-fidelity samples at the onset of dynamic sampling, the multifidelity kriging achieves approximately two orders of magnitude in RMSE reduction over the monofidelity approach. However, with the addition of high-fidelity data, virtually no further error reduction is achieved, and the monofidelity approach overtakes the multifidelity with approximately eighteen training points.

Similar plots of RMSE are provided for the Runge function in Figure 5, using monofidelity and multifidelity approaches with low-fidelity ratios of 4:1, 2:1, and 1:1 for PCE, and 50, 150, and 300 low-fidelity training points for kriging. Overall, the kriging performed better than the PCE, achieving lower RMSE levels using fewer training points. This is likely due to the underlying correlation function of the kriging being better suited than polynomials for the contours of the function. Throughout the addition of high-fidelity training points, the multifidelity formulation converges similarly to the monofidelity model. Ultimately, us-
Figure 4: Comparison of mono- and multifidelity PCE and kriging for the Rosenbrock function (mean of 20 runs with 95% confidence intervals)

Using 50 low-fidelity points is better in terms of RMSE than using 150 or 300, and yields convergence almost identical to the monofidelity model. For a range of about 25 to 50 high-fidelity points, though, using 150 low-fidelity points produces a slightly better model.

The RMSE using PCE indicates that, similar to the Rosenbrock function, it is advantageous to use corrected low-fidelity data to achieve higher polynomial orders with fewer high-fidelity evaluations. Using a low-fidelity ratio of 4:1, the multifidelity approach achieves RMSE levels similar to the monofidelity approach in 80% fewer high-fidelity evaluations. Comparing reductions in RMSE versus polynomial order indicates that for a particular order, adding low-fidelity data increases RMSE. However, the advantage is not in reducing error at a particular order, but rather in achieving that order using less expensive computations. Note that the PCE with a low-fidelity ratio of 1:1 exhibits no reduction, and frequently an increase, in RMSE. This is likely due to the fact that in the dynamic sampling process for PCE, low- and high-fidelity collocation points are coincidental. For larger low-fidelity ratios, only a subset of points are coincidental. However, when all the low-fidelity points are coincidental with the high-fidelity, the linear system becomes poorly conditioned. Contours and training points for PCE and kriging models of the three-dimensional Runge function are provided in Figure 6.

Based on the results of studying the Runge function with varying low-fidelity ratios, the impact of adding gradient information is evaluated for both PCE using a low-fidelity ratio of 4:1 and kriging with 50 low-fidelity points. The results of this study are shown in Figure 7. For the purposes of comparison, a training point with gradient information is considered to have the equivalent cost of two training points without gradients. This scaling is based on the relative cost of obtaining gradients via the adjoint method.

Using kriging, adding gradient information did not provide error reduction over the multifidelity approach using function values alone. The initial greater cost of the gradient case is due to evaluating the same initial training points with gradients at twice the cost of a function evaluation (assuming an adjoint approach). Initially, the gradient case exhibits a steeper convergence profile than using function values alone. However,
at approximately forty training points, the surrogate model saturates, and no further reduction in RMSE is obtained by adding points. While the addition of gradients did not provide any improvement here, Boopathy and Rumpfkeil\textsuperscript{6} showed that utilizing gradient information can be beneficial for higher-dimensional problems.

For PCE, adding gradients to the monofidelity model yields no improvement in terms of RMSE reduction versus number of equivalent function evaluations. In terms of RMSE versus polynomial order, the surrogate with gradient information actually performs slightly worse. This is likely due to performing a regression with more information at fewer points. Local to the training points, the surrogate may be more accurate, but globally the surrogate is worse than the case with more training points. Using multifidelity PCE with function values alone for both low- and high-fidelity training points outperforms the cases using gradient information. Again, this may be attributed to having more points distributed across the design space in the regression process. The case of using function and gradients for both fidelities yields faster reduction of RMSE compared to using gradient information only for the high-fidelity evaluations. This result indicates that in some situations, the addition of gradient information is advantageous.

IV. Aerodynamic Performance Application

IV.A. High-Fidelity Aerodynamic Simulation

The high-fidelity aerodynamic analysis tool employed here is Cart3D.\textsuperscript{19–21} Cart3D is a second-order, cell-centered, finite-volume Cartesian Euler solver using a time-marching procedure to advance the steady solution to convergence. Here, a five-stage temporal scheme of van Leer\textsuperscript{22} is utilized with a CFL number of 1.15. The van Leer flux limiter is utilized due to the presence of a shock in almost all the cases of interest.

Cart3D has two particular features that make it attractive. First, mesh generation is simplified by its Cartesian meshing strategy. The user provides a surface triangulation of the vehicle, the distance to the far field, and the number of nodes in the initial far field mesh. The geometry in this case is generated using
Engineering Sketch Pad (ESP), a parametric modeler based on the OpenCASCADE geometry system, automatically generates a hexahedral mesh, splitting large cells into eight smaller cells in regions near the body where more refinement is needed. The slip wall boundary condition is applied within the split cells that intersect the body. Based on this meshing strategy, traditional mesh refinement may be carried out by increasing the initial number of far field nodes, or increasing the number of refinement levels the mesher uses near the body.

The second reason Cart3D is attractive is because it implements adjoint-based mesh adaptation, which is utilized here. After solving for the requested flow condition, the code will analyze the sensitivity of the final result to errors in each cell in the domain. Using this error map, Cart3D is able to refine the mesh only in regions that have high impact on the solution, rather than refining globally. This selective refinement leads to grid-converged solutions with fewer cells, and thus faster run times.

In order to reach a grid-independent solution, the far field mesh size and number of mesh adaptation cycles were varied to find a balance between accuracy and computational cost. Two such studies are presented in Figure 8 – one with all design variables given in Subsection IV.C at their lower bounds, and the other at the upper bounds. The number of far field nodes in each direction was repeatedly doubled, from \( N = 4 \) to \( N = 32 \). The number of spanwise far field nodes is half of those on the chordwise and thickness directions, as a half-span model with symmetry plane is used. Each mesh used six levels of refinement initially, followed by nine levels of adjoint-based adaptation. The cell counts for each mesh are tabulated in Table 2.

The objectives provided to the mesh adaptation are to minimize errors in lift and drag, and the variations in lift and drag coefficients versus mesh size are plotted in Figure 8. For the higher Mach number cases, the force coefficients converge to the same value within five to nine mesh adaptations, depending on the number of far field nodes. For the lower Mach number cases at small angle of attack, \( \alpha = -1^\circ \), the forces are more difficult to converge. For lift, at least 16 far field nodes in each direction and 7 mesh adaptations are
Figure 7: Comparison of mono- and multifidelity PCE and Kriging with gradient information for the 3D Runge function (mean of 20 runs with 95% confidence intervals)

Table 2: Initial and final mesh sizes (half-span model) for grid convergence study at upper bounds of design space

<table>
<thead>
<tr>
<th>Far Field Nodes ($N_x = 2N_y = N_z$)</th>
<th>Initial Refinement Level</th>
<th>Initial Cell Count</th>
<th>Final Cell Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>6</td>
<td>4512</td>
<td>336814</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>7704</td>
<td>534569</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>14017</td>
<td>962085</td>
</tr>
<tr>
<td>32</td>
<td>6</td>
<td>45137</td>
<td>3114228</td>
</tr>
</tbody>
</table>

required to achieve an acceptably converged solution. Similarly for drag, larger far field meshes and more mesh adaptations are required to turn the knee in the curve. To make a final selection of mesh size and adaptation, the processor time required for each solution is assessed, and is shown in Figure 8. Ultimately, a far field mesh size of $N = 16$ and 9 levels of mesh adaptation were selected for a balance between accuracy and solution time. Typical convergence histories of the residuals and forces are provided in Figure 9.

IV.B. Low-Fidelity Aerodynamic Simulations

For the low-fidelity prediction of lift, the Wingdes code\textsuperscript{24} is used. Wingdes provides subsonic and supersonic linearized solutions for forces on twisted, cambered, zero-thickness lifting surfaces. In addition to providing lift and induced drag, Wingdes also implements empirical corrections to account for leading edge suction, the thrust arising from the low pressure of the flow around a finite-thickness leading edge from the stagnation point, and for vortex forces due to leading-edge separation. While Wingdes has the capability to design the mean camber surface to minimize induced drag, the code is only used in analysis mode here. Analysis inputs
IV.C. Comparison of Multifidelity to Monofidelity

To assess the quality of the kriging and PCE models, a Cartesian database of Cart3D results was created. The mesh is defined for a domain of angle of attack from $-1^\circ \leq \alpha \leq 6^\circ$ in one degree increments, Mach number from $0.8 \leq M \leq 1.6$ in increments of 0.05, and thickness-to-chord ratio from $0.06 \leq \frac{t}{c} \leq 0.12$ in increments of 0.1, yielding a total of 952 samples. This same database is used to verify the suitability of Wingdes in Figure 10. For comparison, the high-fidelity database required on the order of four days on three 16-core nodes of a computational cluster to complete, while the low-fidelity database required on the order of 10 minutes on a single comparable processor core.

Qualitatively, Wingdes captures the trends predicted by Cart3D reasonably well. Cart3D predicts peak lift coefficient near Mach 0.9, while Wingdes predicts the peak at Mach 1.0, corresponding to the switch-over from subsonic to supersonic models. Aside from the shift in peak location, the variation in lift with Mach number agrees well between the two codes. The expected variation of lift with angle of attack is comparable, and Wingdes does an excellent job of predicting zero-lift angle of attack. Cart3D predicts slight variation in lift with thickness-to-chord ratio, with thinner airfoils generating higher lift. Wingdes, which is based on a zero-thickness formulation, does not exhibit any such variation.

The RMSE’s for the surrogate models of lift coefficient are presented in Figure 11. Overall, kriging performs better than PCE, achieving almost an order of magnitude lower RMSE for both mono- and multifidelity formulations. Provided the initial nine training points, PCE forms a better linear model than the initial kriging model. However, using approximately twenty training points, kriging surpasses PCE. The
ability of PCE to produce significant reduction in error is likely due to the over-fitting of a challenging surface with polynomials of increasingly higher order. The actual high- and low-fidelity surfaces being fit are illustrated in Figure 10.

![Surrogate Models](image)

Figure 10: Monofidelity surrogate models of predicted lift coefficient; meshed surfaces are truth models for comparison

The high-fidelity aerodynamic response is difficult to fit for both kriging and PCE due to the variation in behavior between the transonic and supersonic regimes and from low angle of attack to high. The transonic response is characterized by a pulse centered about Mach 0.9. The higher Mach numbers, on the other hand, present a very flat response that may be easily captured by a linear model. As angle of attack decreases, the pulse amplitude attenuates until it vanishes at the zero-lift angle of attack. It is challenging for the PCE to fit a high-order polynomial that exhibits the single pulse flanked by a linear response. Similarly, kriging must find a single correlation length that captures both phenomena; ideally, kriging would use a short correlation length for the pulse, and a longer one for the linear response. Compared to the Cart3D response, the low-fidelity Wingdes response is easier for both methods to fit, resembling more of an absolute value function scaled by angle of attack.

Subfigure 11a indicates that using fewer low-fidelity training points versus more is beneficial for kriging. Using only 50 low-fidelity points achieves the best convergence between both mono- and multifidelity approaches, whereas using 150 low-fidelity points yields the worst result. Our conjecture is that using too many low-fidelity points may lead to over-fitting of the response, resulting in oscillatory behavior that the bridge function must compensate for. This over-fitting is exacerbated by the lack of a distance constraint in the Latin hypercube sampling of the low-fidelity data. With more training data, the likelihood of points in close proximity to each other (which leads to model oscillations) increases.

Subfigure 11b shows that for PCE, increasing the low-fidelity ratio from two to four has virtually no impact. This indicates that using more or less low-fidelity data does not change the ability to determine the additive and multiplicative corrections in this case. Rather, the RMSE of the surrogate model is dominated by the difficulty in resolving the transonic pulse and by the oscillatory Runge phenomenon near the domain boundary.
Figure 11: Comparison of mono- and multifidelity PCE and Kriging for lift coefficient prediction (mean of 5 runs with 95% confidence intervals)

boundaries. These difficulties are highlighted in Figure 12. Using a sixth-order, monofidelity PCE (168 training points), the surrogate is able to approximate the transonic pulse, though it cannot exactly match it, and the high-order polynomial suffers from the Runge phenomenon at the extremes. Reducing the expansion to fourth order (70 training points), the oscillatory errors are alleviated, but the model loses important degrees of freedom to model the pulse. Adding 276 low-fidelity training points (low-fidelity ratio of four) enables PCE to produce an eighth-order approximation with a reduced high-fidelity budget. This addition recovers some of the ability to model both the pulse and linear regions, though the surrogate still suffers near the boundaries.

A similar progression of kriging models is provided in Figure 13. The high-fidelity model using 167 training points does qualitatively an excellent job of capturing the aerodynamic response, though the flat regions at low angles of attack and high Mach numbers exhibit small oscillations with respect to Mach number. These oscillations are due to the extension of the short correlation length required to model the pulse across the entire domain. Reducing the number of high-fidelity training points to 89, the predictive capability is maintained at high angles of attack, though the prediction of zero lift is poor. Adding just 50 low-fidelity points, the predictive capability at small angles of attack is restored.

The loss of predictive capability by kriging at small angles of attack with reduced training points may be attributed in part to the dynamic sampling routine. Subfigure 13e highlights that the dynamic sampling focused training point additions on the peak in lift near Mach 0.9 at higher angles of attack. The short correlation length combined with a limited number of data points resulted in poor prediction of the zero-lift surface. However, had the dynamic sampling not focused on the area of peak lift, the errors isolated to low angles of attack would likely spread across the entire domain.

A comparison of training point distribution for PCE in Figure 12 and kriging in Figure 13 indicates another significant difference between the two methods. In Figure 12, training points are distributed relatively evenly throughout the domain. An even distribution of points is required to quell peaks in high-order
Figure 12: Comparison of mono- and multifidelity PCE models using fewer high-fidelity points

Figure 13: Comparison of mono- and multifidelity kriging models using fewer high-fidelity points
polynomials where they are not desired. Kriging, in contrast, requires training points in areas where trends differ from neighboring regions. Thus, the dynamic sampling in Figure 13 concentrates on the pulse region rather than the flat region on which PCE focuses.

Another distinction between multifidelity kriging and PCE is in the implementation of the additive correction. Here, the kriging additive bridge is itself a kriging model, whereas the PCE correction is limited to first-order polynomials. This gives the kriging additional freedom to correct the low-fidelity model. Providing this additional freedom to the PCE poses severe model management challenges. First is the curse of dimensionality. The number of training points required in PCE scales with the factorial of polynomial order. Increasing the order of the additive correction along with that of the model of interest doubles the points required. Second, the effect of over-fitting and the Runge phenomenon is compounded. To investigate this effect, training points for an eighth-order PCE were used to fit sixth-order models for lift prediction and additive correction. The resulting model bore little resemblance to the truth model, and the calculated RMSE was worse than that of a first-order PCE. Thus, determination of when to increase the polynomial order of the corrective functions is an area requiring further research.

V. Conclusions

A new approach for constructing multifidelity and gradient-enhanced polynomial chaos expansions is presented. This approach has the advantage of simultaneously determining least squares-optimal additive and multiplicative corrections of arbitrary order along with the model of interest. A key disadvantage, common to all polynomial regression approaches, is that the number of training points may not be selected independent of polynomial order. This difficulty is exacerbated in the multifidelity approach by the inclusion of corrective polynomials for which additional coefficients must be determined. The multifidelity strategy, however, partially addresses both of these problems. Through the addition of cheaper low-fidelity data, additional data for the determination of coefficients is gleaned. Thus, a certain model order may be achieved more cheaply with multifidelity versus the monofidelity approach.

The PCE approach is compared to a multifidelity kriging approach. Kriging has the advantage of arbitrariness in the number of training points. Along with this comes added freedom in developing the additive bridge function compared to PCE. Near the domain boundaries and in extrapolation, kriging tends to perform better than PCE, which suffers from the oscillatory Runge phenomenon. In cases where polynomials are known to be good approximate models of the true function, PCE is an obvious choice. However, for non-polynomial functions, such as the Runge function and the presented aerodynamic database, kriging outperformed PCE in terms of RMSE reduction. PCE, on the other hand, converged more consistently between runs, thus having narrower confidence intervals than kriging. PCE also has the advantage of having coefficients that may be used to track convergence and identify critical terms.

When the corrective functions accurately map the low- to high-fidelity data, the multifidelity PCE approach has a clear advantage over monofidelity. This advantage is achieved by attaining higher polynomial orders with fewer high-fidelity evaluations rather than by increasing accuracy of the regression at a particular order. Multifidelity kriging also performs better than monofidelity. Here, the advantage comes from reducing RMSE in the first iterations. In contrast to PCE, kriging appears to benefit from using fewer low-fidelity training points rather than more. In subsequent iterations, RMSE reduction with the addition of training points is similar for both mono- and multifidelity approaches. Multifidelity modeling with gradients is implemented for both PCE and kriging. In both cases, using function values alone provided better RMSE reduction than when using gradients. However, why the gradient information did not provide more improvement is an area for further research. Gradient information may prove more beneficial in higher-dimensional domains, as the amount of information added per cost of evaluation is proportional to dimensionality.

The dynamic sampling approach utilized is also successful in addressing the differing needs of kriging and PCE. In kriging, the dynamic sampling concentrates training points in regions of varying trends, better capturing the region of interest. To the contrary, in PCE the training points are more dispersed, providing control over the multiple extrema inherent to higher-order polynomials.
Acknowledgments

This effort was supported in part by the Air Force Research Laboratory, Aerospace Systems Directorate, under the Civilian Academic Degree Payment Program.

References