2016

Locally Optimized Covariance Kriging for Non-stationary System Responses

Daniel Lee Clark Jr.
Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all

Part of the Mechanical Engineering Commons

Repository Citation
Clark, Daniel Lee Jr., "Locally Optimized Covariance Kriging for Non-stationary System Responses" (2016). Browse all Theses and Dissertations. 1502.
https://corescholar.libraries.wright.edu/etd_all/1502

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu, library-corescholar@wright.edu.
Locally Optimized Covariance Kriging for Non-Stationary System Responses

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering

by

Daniel L. Clark, Jr.
B.S.M.E., Wright State University, 2014

2016
Wright State University
I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Daniel L. Clark, Jr. ENTITLED Locally Optimized Covariance Kriging for Non-Stationary System Responses BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science in Engineering.

Ha-Rok Bae, Ph.D.
Thesis Director

George P. Huang, Ph.D.
Chair, Department of Mechanical and Materials Engineering

Committee on Final Examination

Ha-Rok Bae, Ph.D.

Ramana V. Grandhi, Ph.D.

Joseph C. Slater, Ph.D., P.E.

Robert E.W. Fyffe, Ph.D.
Dean, School of Graduate Studies
ABSTRACT

Clark, Jr., Daniel. M.S.Egr., Department of Mechanical and Materials Engineering, Wright State University, 2016. Locally Optimized Covariance Kriging for Non-Stationary System Responses.

In this thesis, the Locally-Optimized Covariance (LOC) Kriging method is developed. This method represents a flexible surrogate modeling approach for approximating a non-stationary Kriging covariance structures for deterministic responses. The non-stationary covariance structure is approximated by aggregating multiple stationary localities. The aforementioned localities are determined to be statistically significant utilizing the Non-Stationary Identification Test. This methodology is applied to various demonstration problems including simple one and two-dimensional analytical cases, a deterministic fatigue and creep life model, and a five-dimensional fluid-structural interaction problem. The practical significance of LOC-Kriging is discussed in detail and is directly compared to stationary Kriging considering computational cost and accuracy.
# Contents

1 Introduction 1

2 Background Theory 3
   2.1 Design Optimization 3
   2.2 Surrogate Models 7
      2.2.1 Design of Experiments 8
      2.2.2 Regression Analysis 11
      2.2.3 Gaussian Process Modeling 14
      2.2.4 Regression Kriging 19
   2.3 Research Contribution 21

3 Non-Stationary Surrogate Modeling 22
   3.1 Literature Review 25
      3.1.1 Geostatistics Development 25
      3.1.2 Engineering Development 27
   3.2 Summary 28

4 Locally-Optimized Covariance Kriging 29
   4.1 Formulation 29
      4.1.1 Kriging Model with Locally-Optimized Covariance 29
      4.1.2 Local Window Size Optimization 31
      4.1.3 Construction and Aggregation of Multiple Local Kriging Models 33
      4.1.4 Imposing Physics on LOC-Kriging 34
   4.2 Non-stationary Identification Test 35
   4.3 Demonstation 40
      4.3.1 One-Dimensional Example 41
      4.3.2 Two-Dimensional Example 45
      4.3.3 Ti-6242S Fatigue and Creep Coupled Numerical Experiment 50
      4.3.4 Five-Dimensional Fluid Structural Interaction Example 57
   4.4 Summary 63

5 Summary and Future Work 65
# List of Figures

2.1 Design Optimization Process ................................................. 4  
2.2 Local Optimal Solutions ..................................................... 7  
2.3 Factorial Design Schemes ................................................... 9  
2.4 Latin Hypercube Sampling ................................................... 10  
2.5 Regression Models ............................................................ 14  
2.6 Gaussian Correlation Function ............................................. 16  
2.7 Kriging Response Decomposition ........................................... 18  
2.8 Regression Kriging Response Decomposition .............................. 20  
2.9 Example Regression Kriging Response ..................................... 20  
3.1 Adaptive One-Dimensional Stationary Kriging Response ................. 24  
3.2 Even One-Dimensional Stationary Kriging Response ...................... 24  
3.3 Moving Window Kriging ...................................................... 26  
3.4 Nonlinear Mapping Approach ................................................. 28  
4.1 LOC Window Membership function ........................................ 30  
4.2 LOC Window Aggregation Membership Function .......................... 34  
4.3 Non-stationary Identification Test Outline ................................ 36  
4.4 Voronoi Test Points With Contour ......................................... 37  
4.5 Linear Regression Model ..................................................... 38  
4.6 Initial Localities ................................................................ 39  
4.7 Four Identified Localities ..................................................... 40  
4.8 1-D LOC-Kriging Predictions ............................................... 42  
4.9 1-D Comparison of Methods ................................................ 42  
4.10 1-D Convergence Histories ................................................ 45  
4.11 2-D Adaptively Collected Data ............................................. 46  
4.12 2-D Over-plot and Contour, Stationary Kriging ......................... 47  
4.13 Upper and Lower Estimated Error Bounds, Stationary .................. 47  
4.14 2-D LOC-Kriging Models .................................................... 48  
4.15 2-D Over-plot and Contour, LOC-Kriging ................................. 49  
4.16 Upper and Lower Estimated Error Bounds, LOC-Kriging ............... 50  
4.17 HTV-3X Aircraft Frame ..................................................... 52  
4.18 Loading Scenario ............................................................... 52
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.19</td>
<td>Analytical Life Model</td>
<td>54</td>
</tr>
<tr>
<td>4.20</td>
<td>Over-plot and Contour, Stationary Kriging</td>
<td>55</td>
</tr>
<tr>
<td>4.21</td>
<td>Over-plot and Contour, LOC-Kriging</td>
<td>57</td>
</tr>
<tr>
<td>4.22</td>
<td>Computational Domain and Coarse C-Grid</td>
<td>59</td>
</tr>
<tr>
<td>4.23</td>
<td>CFD Convergence Histories</td>
<td>60</td>
</tr>
<tr>
<td>4.24</td>
<td>Sting Mounted Airfoil</td>
<td>60</td>
</tr>
<tr>
<td>4.25</td>
<td>FSI Two-Dimensional True Response Surfaces</td>
<td>61</td>
</tr>
<tr>
<td>4.26</td>
<td>FSI One-Dimensional Slice</td>
<td>63</td>
</tr>
<tr>
<td>5.1</td>
<td>Modified Regression Kriging Response</td>
<td>66</td>
</tr>
</tbody>
</table>
List of Tables

4.1 1-D Performance Comparison ........................................ 44
4.2 1-D Computational Times ............................................ 44
4.3 2-D Performance Comparison ........................................ 49
4.4 2-D Computational Times ............................................ 49
4.5 Life Model Performance Comparison ............................ 56
4.6 Life Model Computational Times ................................. 56
4.7 Design Parameters and Ranges ................................... 61
4.8 FSI Performance Comparison ....................................... 62
4.9 FSI Computational Times ........................................... 63
Acknowledgments

I would like to take this opportunity to extend my thanks to Dr. Ha-Rok Bae for his mentoring and guidance throughout my graduate studies. The opportunities and experience I obtained as his graduate student has allowed me to grow as a person, researcher, and an academic. Dr. Ha-Rok Bae installed in me the confidence to pursue my Ph.D. at Wright State University.

None of the research presented herein would have been possible if it were not for the support and funding from Drs. Ravi Penmetsa and Edwin Forrester from the Air Force Research Laboratory (AFRL).

Additionally, I would like to extend my gratitude to two more individuals, Dr. Ramana Grandhi and Dr. Joseph Slater, who took the time to evaluate this thesis. I am grateful for their willingness to help and be a part of my final examination committee.

Finally, I would like to thank the members of Dr. Ramana Grandhi’s research group, CE-PRO, Ohio Center of Excellence for Product Reliability and Optimization. Their support was invaluable in both my research and course work. I specifically would like to thank Koorosh Gobal for his expertise in developing the fluid-structural interaction problem used in this thesis.
Dedicated to My Parents.
Introduction

The increasing scale and complexity of computer simulations have significantly proliferated the computational cost of performing design exploration in both academic and industrial applications. Typical explorations include design optimization, database construction, and reliability-based design optimization. To alleviate the computational burden, various surrogate models, also known as response surface models or metamodels, are considered for approximating large scale and complex engineering simulations. Generally, surrogate models are constructed with a limited number of training points [1][2][3][4][5][6]. Among the many different surrogate models, Kriging has gained its popularity due to high accuracy and flexibility representing non-linear system responses [7][8]. Kriging has many different formulations including: Regression Kriging[9], Gradient-Enhanced Kriging[10], Dynamic Kriging[11], and Stochastic Kriging[12].

Essentially, Kriging produces an optimal interpolation by combining both the trend regression and the realization of a local random process modeled by an optimized covariance structure. One of the benefits of Kriging is the expected Mean Square Error (MSE) calculated along with a response prediction at any location of interest. This unique benefit of the MSE is not present in Response Surface Method (RSM) or Radial Basis Functions (RBF) and this is why RSM and RBF are not investigated along side Kriging. The MSE is typically viewed as model uncertainty due to lack of training data points. Therefore, it is utilized in numerous global surrogate optimization strategies. In the Efficient Global Optimization (EGO) method [8], the expected improvement is defined as a function of MSE and a cur-
rent minimum response. MSE plays a key role to deploy adaptive data points sequentially in EGO. The Efficient Global Reliability Analysis (EGRA)\[13\] also uses EGO to construct an adaptive Kriging model of a limit state boundary function for efficient reliability assessment. The Multiple Surrogate Efficient Global Optimization (MSEGO) method\[14\] uses MSE from Kriging as a universal uncertainty estimate for other types of surrogate models. Thus, an accurate representation of both the estimated response and the MSE are critical in many methodologies.

To more accurately predict the response of deterministic systems the proposed Locally-Optimized Covariance (LOC) Kriging methodology approximates a non-stationary covariance matrix. The approximation is accomplished through the use of local surrogate models. The surrogate locations can either be predefined or found through a statistical significance test. Once the initial sizes and locations are established, the sizes are optimized and aggregated. The use of localities allows for a more accurate representation of local fluctuations in response. Therefore, the MSE is also represented more accurately in both situations: more data is needed or sufficient data was collected.

The remainder of this thesis is organized as follows. In Chapter 2, background theory is presented for design optimization, surrogate modeling and the primary contribution of this thesis is discussed. A brief literature review is then discussed in Chapter 3 concerning past contributions in the fields of geostatistics and engineering addressing non-stationary surrogate modeling. The Locally-Optimized Covariance Kriging methodology is then described in detail with numerical examples in Chapter 4. The thesis is concluded with a brief discussion of future work and applications.


**Background Theory**

Based on the Introduction in Chapter 1, the importance of surrogate models for large scale problems is established to decrease computational cost. However, where, when and how to implement these approximation techniques has yet to be addressed in the thesis. In this chapter, a clear indication of where and when surrogate models are utilized in design optimization is defined. From there, gradient based design optimization methodologies are discussed. This discussion is essential to establish the importance of continuity and local behaviors of a surrogate model. Then, three common stationary surrogate modeling techniques are presented including Regression Modeling, Gaussian Process Modeling, and Regression Kriging. First, the standard Design Optimization process is discussed in the context of a structural optimization problem.

### 2.1 Design Optimization

The design optimization framework that is introduced in this chapter is based on the typical process for structural optimization. An illustration of this process is given in Fig. 2.1. The process begins by selecting an initial design or configuration of the system. The system’s performance is then evaluated by an objective function and a set of constraints that the design must satisfy while minimizing the objective. The objective function and constraints are typically evaluated utilizing commercial finite element analysis (FEA) software such as Ansys, Abaqus, or Nastran. Sensitivity or gradient information can be extracted using
finite difference methods or determined analytically. However, for each gradient extracted an additional run is required, increasing computational cost at each iteration. The gradient information is then used as an input for the selected optimization algorithm to find the next design. This process is repeated until the objective function converges.

Figure 2.1: **Design Optimization Process.** Standard structural design optimization process. Figure 4.1 from reference [15]

An important consideration in the design optimization process is construction the objective and constraints. Traditionally, the mass of the structure is minimized while constraints are placed on displacements, natural frequencies, stresses, or a combination of the three. The standard mathematical optimization formulation for a structural problem [16] is
given as

\[
\begin{align*}
\text{Minimize:} & \quad F(x) \\
\text{Subject to:} & \quad g_j(x) = 0 \\
& \quad g_k(x) \geq 0 \\
\text{Design variables:} & \quad x = [x_1, x_2, x_3, x_4, \ldots, x_n]^T \\
\text{Side bounds:} & \quad x_{LB} \leq x_i \leq x_{UB}
\end{align*}
\] (2.1)

In this general representation, \( F \) is the desired minimization objective function. The objective is subject to both equality and inequality constraints \( g_i(x) \) and \( g_k(x) \) respectively. Both the objective and constraints are evaluated at the design vector, \( x \). The design space can also be subjected to side bounds, where \( LB \) and \( UB \) represent the lower and upper bounds respectively. The side bounds are decomposed into additional inequality constraints so that the final representation becomes

\[
\begin{align*}
\text{Minimize:} & \quad F(x) \\
\text{Subject to:} & \quad g_j(x) = 0 \\
& \quad g_k(x) \geq 0 \\
& \quad x_i - x_{LB} \geq 0 \\
& \quad x_{UB} - x_i \geq 0 \\
\text{Design variables:} & \quad x = [x_1, x_2, x_3, x_4, \ldots, x_n]^T
\end{align*}
\] (2.2)

As indicated by Fig. 2.1, an approximation of the system can be made for small changes to the design. This approximation can be accomplished using the current design point with gradients through a simple first-order Taylor series expansion. This method works well for linear problems or small changes to a nonlinear problem, but can be improved by considering the previous design point [17]. The slightly more complex evolu-
tion, the two-point function approximations, are advantageous during gradient based design optimization because more information is being leveraged in the iterative process where gradients are already required for the algorithm. However, the two-point approximations rely on optimally finding the nonlinearity index. This optimization may fail, causing a miss representation of the response.

Another option is to approximate the entire system or small subsections of the design space. By selecting a strategic design of experiments (DOE), the objective function, and constraints can be approximated at every untried location in the domain of interest. Depending on the desired approximation method, gradients can also be evaluated at the sample points but are not always necessary. Once the FEA model is evaluated at all of the DOE points and the approximation model built, the entire loop of Fig. 2.1 can be evaluated with approximate analytical equations. It is considered good practice to evaluate the true FEA model at the approximated optimal design variables to ensure the approximation is accurate and none of the constraints are violated. If a violation occurs, the approximation is rebuilt with the additional design point causing the optimization algorithm to avoid the previous optimal location. This process is then repeated until an optimal is found that does not violate the constraints.

Approximating the entire system and creating an approximate problem both have their merits. However, for general design exploration, approximating the entire system utilizing a surrogate model is further developed. In this case, both global and gradient optimization algorithms can be utilized. In this thesis, gradient algorithms are illustrated in further detail due to their speed and reduced computational cost in optimization.

In Fig. 2.2, a theoretical objective function, \( f(x) \), is evaluated from \( x \in [-2.5, 3] \). If this objective function is initially evaluated to the left or right of \( x = 0 \) the optimal solution will be different due to the gradient information acquired at the point. For illustration purposes, four initial points and gradients are shown in gray and red respectively. Points 1 and 2 will lead a gradient algorithm to a local optimum whereas points 3 and 4 will yield a
global optimum for the domain. Because potential situations like these exist it is important
to consider multiple initial points when optimizing. This process of selecting multiple
starting locations is indicative of a global method, section of initial points is not covered in
this thesis. However, the effect of local fluctuations of the response on the optimal solution
should be considered when selecting a surrogate model.

![Graph showing local optimal solutions](image)

**Figure 2.2: Local Optimal Solutions.** Gradient information is collected at four initial
points leading to two different optimal solutions.

Many advanced surrogate modeling techniques can introduce nonphysical changes in
the response where there is a lack of information. This can cause the optimizer to select
sub-optimal solutions. However, with the process of design optimization well defined, the
general formulation established, and possibility of local optimal solutions acknowledged,
techniques for surrogate modeling are now presented.

### 2.2 Surrogate Models

Approximating computer and experimental data is always advantageous when large quan-
tities of data are necessary to perform the analysis. This can be because of cost or time
constraints. Nevertheless, surrogate models fulfill this need. Surrogate models are char-
acterized by the method in which they attempt to fit the data they are applied to. The data is assumed to have a functional relationship between measured variables (independent variables) and predicted variables (dependent variables). Typically two categories are considered, interpolation and extrapolation methods. Interpolation methods are utilized to construct new data points within the range of collected data and should never be applied outside of the range of the original data due to inaccuracies. Extrapolation methods estimate new data points based on the perceived trend of the collected data. However, utilization of either method requires a sample data set before construction.

Before introducing Regression Analysis (extrapolation method), Gaussian Process Modeling (interpolation method) and Regression Kriging (hybrid method), the concept of Design of Experiments is presented. This method allows for the intelligent selection of sampling points for the construction of a desired surrogate model with an optimal number of samples.

2.2.1 Design of Experiments

Data collection should always be performed as efficiently as possible to avoid additional time and cost. For this reason, various Design of Experiment (DOE) methods have been established. Typically, these methods are dependent on the type of model that is fitted to the data. The advantage of DOE methods is their direct utilization of the system to obtain approximate mathematical solutions of the problem even in cases where the system equations cannot be solved easily. The following descriptions clarify the basic limitations of two basic sampling methods: Latin Hypercube Sampling (LHS) and Factorial Design. The simplest of the two DOE methods is the Factorial Design.

Factorial Designs come in many forms including fractional and full factorial designs. The most basic design is the $2^k$ Full factorial design. Where 2 indicates the number of levels and $k$, the number of design variables. $2^k$ designs are typically used early on in the design process to detect simple relationships between the inputs and the system response.
The primary limitation being the design’s fitting capabilities. With two levels, typically denoted as \([-1, 1]\), per design variable, only a simple linear model can be applied to the system.

When higher order models are desired more points are required. The simplest method is to add a center point to the design space in the form of a point \([0, 0]\) for a two-dimensional case. This can be used to check for curvature of the system, however, to approximate more coefficients a more complex design is required. For this case, a \(3^k\) design can be used to approximate the additional coefficients by mixing the center point with the previous two levels. Figure 2.3 demonstrates the evolution of the experiment from a \(2^k\) design, to a \(2^k\) design with a center point, to a \(3^k\) design. This process works for an \(N\) dimensional system. However, as the number of design variables increases, the computational cost rises exponentially due the strict structure of the samples.

![Figure 2.3: Factorial Design Schemes. Starting with a \(2^k\) design, additional points can be added in a sequential sequence to approximate.](image)

To combat the computational cost of multiple level Factual Designs and high dimensionally, Latin Hypercube Sampling (LHS) is utilized. LHS design is a stratified sampling technique, i.e. both space-filling and guarantees non-overlapping designs. For LHS designs without perturbations from the grid center locations, the total number of possible outcomes
can be defined as
\[
\left( \prod_{n=0}^{M-1} (M - n) \right)^{N-1} = (M!)^{N-1}
\] (2.3)

where \( M \) is the number of sample points and \( N \) is the dimensionality of the problem. LHS can be carried out with numerous criterion including maximization of the minimum distance between points (MaxMin) and minimization of the potential correlation between points. Both methods are illustrated in Fig. 2.4. The correlation reduction criterion for LHS designs can be shown to be the minimization of the sum of between-column squared correlations after performing ranked Gram-Schmidt step to orthogonalize the sample locations. In this case, the perturbations from the grid center locations are zero. This lack of perturbation is evident in Fig. 2.4.

![Figure 2.4: Latin Hypercube Sampling. LHS of the same domain with different criterion.](image)

(a) Reduce Correlation  
(b) Maximize the Minimum Distance

In order to perform the MaxMin case the distances between points must be defined. This is accomplished as
\[
\rho_p (s_i, s_j) = \left[ \sum_{k=1}^{N_d} |s_{i,k} - s_{j,k}|^p \right]^{\frac{1}{p}}
\] (2.4)
where $\rho_p$ is the distance measure of the $p^{th}$ order between points $s_i$ and $s_j$, $(j \neq i)$, and $N_d$ is the dimensionality of $s$. When the distance is Euclidean, $p = 2$. And $p = 1$ represents rectangular distance. The optimization formulation for determining a sampling plan is defined as

$$\text{Find: } D_{sf}$$

$$\text{maximize: } \min_{x \in D} [\rho_p (x_i - x_j)]$$

(2.5)

With two popular sampling methods defined, regression analysis can be introduced in the following section. It is important to consider the deterministic nature of the simulations investigated in this thesis. In this context, regressions are structured so that the data is actually interpolated. This requires either high order terms or structuring the regression to match the DOE.

### 2.2.2 Regression Analysis

In many engineering systems, two or more design variables are inherently related. The investigation of the functional relationship between the design variables is the fundamental idea behind regression analysis. Regression analysis can be characterized by two types: linear and nonlinear. If the relationship between the dependent variable and the independent variables can be characterized by a linear combination of some parameters, the model is referred to as linear; otherwise, it is called nonlinear. The typical form of regression analysis is known as the least squares method. This technique minimizes the sum of squares of the residuals to find the best fit.

First consider the equation of a linear regression model

$$y(x) = \beta_0 + \beta_1 f_1(x) + \ldots + \beta_k f_k(x) + \epsilon$$

(2.6)
where $\beta_i, i = 0, 1, 2, \ldots, k$, are the regression coefficients and $\epsilon$ is the model error, assumed to be normally distributed with a zero mean and variance $\sigma^2_e$. Equation 2.6 is often written in matrix notation for $n$ sample values of $x$ and $y$ as

$$Y = F\hat{\beta} + \epsilon$$  \hspace{1cm} (2.7)

In expanded form Eq. 2.7 becomes

$$
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix} = 
\begin{bmatrix}
1 & f_1(x_1) & f_2(x_1) & \cdots & f_k(x_1) \\
1 & f_1(x_2) & f_2(x_2) & \cdots & f_k(x_2) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & f_1(x_n) & f_2(x_n) & \cdots & f_k(x_n)
\end{bmatrix} \begin{bmatrix}
\beta_1 \\
\beta_2 \\
\vdots \\
\beta_n
\end{bmatrix} + 
\begin{bmatrix}
\epsilon_1 \\
\epsilon_2 \\
\vdots \\
\epsilon_n
\end{bmatrix}
$$  \hspace{1cm} (2.8)

The method of least squares is used to obtain the regression coefficients as

$$\hat{\beta} = (F^T F)^{-1} F^T Y$$  \hspace{1cm} (2.9)

The fitted model is typically represented as

$$\hat{Y} = F\hat{\beta}$$  \hspace{1cm} (2.10)

with residuals

$$e = Y - \hat{Y}$$  \hspace{1cm} (2.11)

The residuals can be used to assess the quality of the regression’s fit. The nonlinear representation is not significantly different from Eq. 2.6. The linear regression of a polynomial
model of a one-dimensional case can be written as

\[ y(x) = \beta_0 p_0(x) + \beta_1 p_1(x) + \cdots + \beta_m p_m(x) + \epsilon \]  \hspace{1cm} (2.12)

where the degree of \( p_i(x) \) is \( i = 0, \cdots, m \) and the polynomial approximation is of order \( m \) for this case. The simplest polynomial model consists of monomials \( x^m \). This regression model can also be solved using Eq. 2.9. This time the \( F \) matrix is written as

\[
F = \begin{bmatrix}
p_0(x_1) & p_1(x_1) & p_2(x_1) & \cdots & p_m(x_1) \\
p_0(x_2) & p_1(x_2) & p_2(x_2) & \cdots & p_m(x_2) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
p_0(x_n) & p_1(x_n) & p_2(x_n) & \cdots & p_m(x_n)
\end{bmatrix}
\]  \hspace{1cm} (2.13)

where \( p_0(x) = 1, p_1(x) = x, \cdots, p_m(x) = x^m \). An important issue is to determine the order of polynomials. High-order polynomials can be utilized to closely fit sets of data, but the high-order model are generally less accurate between DOE points. This also means, when utilized in a gradient optimization technique, there is the potential to become stuck at a local optimum. The significance of a regression model can be accessed through the uses of ANalysis of VAriance (ANOVA). ANOVA utilizes the residuals as well as other statistical means to assess the lack of fit and confidence of the predicted response. When assessed iteratively, the model can be reduced based on the significance of the coefficients.
Figure 2.5: **Regression Models.** The models are applied to appropriate data sets with noise to illustrate their fitting behaviors.

Due to the deterministic nature of most computer simulations, regression models allow for quick and simple approximates or in some cases exact replications of the system if the underlying physics is well understood. With the combination of DOE, regression models can also act as an interpolation scheme with an optimal number of samples based on the order of the model. However, many systems have highly nonlinear responses. Given a large DOE, regression models can fit the previously mention pneumonia, but it quickly becomes computationally expensive. Therefore, more advanced methods are explored.

### 2.2.3 Gaussian Process Modeling

Unlike the regression models, Gaussian Process Models are always an interpolation model. Kriging is a special case of the Gaussian Process Model. In the standard Kriging model, the response at locations $x$ are estimated by the combination of the global trend (linear or non-linear regression) and realization of a stochastic process (fit of the regression error), Eq. 2.14.

$$\hat{y}(x) = m(x) + z(x)$$  \hspace{1cm} (2.14)
Here, the global trend function, is the same as Eq. 2.7. For completeness, the global trend is rewritten as

\[ m(x) = F\hat{\beta} \]  

(2.15)

where \( F \) is a user-selected basis function vector, and an unknown regression coefficient vector, \( \hat{\beta} \), obtained from the generalized least-squares method. The stochastic process, \( z(x) \), describes localized deviations with zero mean and a covariance structure as

\[ \text{COV} [z(s_i), z(s_j)] = \sigma^2 R(\theta_k, s_i, s_j, p_k) \]  

(2.16)

where \( \sigma^2 \) is the process variation and \( R \) is the correlation function between the two sample data points, \( s_i \) and \( s_j \), with the correlation parameter vectors, \( \theta \) and \( p \). \( \theta_k \) can be thought of as the activeness of the function whereas \( p_k \) is the function smoothness for the \( k^{th} \) design space dimension [8]. The correlation function can be selected from a collection of various function forms. The Gaussian function is often used in engineering applications and defines each element of the correlation matrix as

\[ R(\theta, p, s_i, s_j) = \text{corr}(s_i, s_j) = \prod_{k=1}^{N_d} \exp(-\theta_k d_k^{p_k}) \]  

(2.17)

where \( N_d \) is the parameter dimension and \( d_k = s_{i,k} - s_{j,k} \) is the distance between the sample points, \( s_i \) and \( s_j \) in the \( k^{th} \) dimensional direction, and \( p_k \) is defined as 2. To demonstrate the effect of the activeness parameter on the correlation function, \( \theta \), Fig. 2.6 is presented. From the plot, there is an inverse relationship between \( \theta \) and correlation length, i.e. as \( \theta \) becomes smaller the correlation length becomes larger.
Figure 2.6: **Gaussian Correlation Function.** The correlation function is shown with various activeness parameters.

With the individual elements of the correlation matrix defined and $\theta$ characterized, the correlation matrix between samples becomes

$$
R = \begin{bmatrix}
\prod_{k=1}^{Nd} e^{\exp(-\theta_k|s_{1,k} - s_{1,k}|^2)} & \cdots & \prod_{k=1}^{Nd} e^{\exp(-\theta_k|s_{1,k} - s_{n,k}|^2)} \\
\vdots & \ddots & \vdots \\
\prod_{k=1}^{Nd} e^{\exp(-\theta_k|s_{n,k} - s_{1,k}|^2)} & \cdots & \prod_{k=1}^{Nd} e^{\exp(-\theta_k|s_{n,k} - s_{n,k}|^2)}
\end{bmatrix}
$$

(2.18)

The values for $\theta$ are determined by maximizing the likelihood of the samples described by Eq. 2.19 over $\theta > 0$.

$$
L \left( \hat{\beta}, \sigma^2 \right) = \frac{1}{(2\pi)^{\frac{n}{2}} (2\sigma^2)^{\frac{n}{2}}} |R|^{\frac{1}{2}} e^{\exp \left[ -\frac{(y - \hat{\beta})^T R^{-1} (y - \hat{\beta})}{2\sigma^2} \right]}
$$

(2.19)

Based on the generalized least square regression, the regression coefficient vector is found by Eq. 2.20 while minimizing the mean squared error defined by Eq. 2.21.

$$
\hat{\beta} = (F^T R^{-1} F)^{-1} F^T R^{-1} y
$$

(2.20)
\[ \phi(s) = E \left[ (\hat{y}(s) - y(s))^2 \right] \quad (2.21) \]

Here, \( y \) and \( \hat{y} \) are the samples and predicted response vectors respectively. The estimate of the variance from the global trend is obtained by

\[ \sigma^2 = \frac{1}{n} \left( y - F\hat{\beta} \right)^T R^{-1} \left( y - F\hat{\beta} \right) \quad (2.22) \]

where, \( n \) is the number of sample points. When Eq.'s 2.20 and 2.22 are substituted into Eq. 2.19 the result is the so-called 'concentrated likelihood function'. This simplifies to Eq. 2.23. In this study, the optimization of the correlation parameter is performed by a Design of Experiments-based optimization. To obtain the optimum solution in a robust way, an advanced tailored optimization method such as generalized pattern search algorithm [18], can be used. Typically, it is assumed that the underlying function behavior is stationary within the exploration space of interest. Thus, the same correlation function is applied over the entire exploration space to predict the function behavior. However, if this assumption is violated unnecessary fluctuations and over-amplifications of both the response and expected mean squared error predictions are introduced. This violation is explored and addressed in detail in the following chapter.

\[ \text{Minimize } \Phi(\theta) = -\frac{1}{2} \ln(|R|) - \frac{n}{2} \ln(\sigma^2) \quad (2.23) \]

Now that \( \theta \) is found the Kriging prediction can be obtained at \( x \) with Eq. 2.24. Where \( r \) is the correlation between estimation points \( x \) and sample points \( s \).

\[ \hat{y} = F\hat{\beta} + r^T R^{-1} \left( y - F\hat{\beta} \right) \quad (2.24) \]

Equation 2.24 is in the same form as Eq. 2.14. The first term, \( F\hat{\beta} \), represents the regression
model, and the second term, \( r^T R^{-1} \left( y - F \hat{\beta} \right) \), is the fit of the regression error. To illustrate this, Fig. 2.7 decomposes the two processes into separate response before combining them. The line touching the second data point in 2.7a represents the error, or distance from the regression at point two. Figure 2.7b shows that the error or the distance to the points are fitted with a continuous function, resulting in 2.7c.

![Graphs](image)

(a) Regression Component

(b) Stochastic Component

(c) Total Response

**Figure 2.7:** Kriging Response Decomposition.

As demonstrated, Kriging produces an interpolation of the supplied sample points. However, some computer simulations may have a significant amount of noise due to convergence or various run time errors. Therefore, regression models may better approximate the true surface. A compromise between the traditional regression modeling techniques and Kriging is Regression Kriging.
2.2.4 Regression Kriging

To approximate the response of a noisy function, Regression Kriging offers the flexibility of Kriging with a non-interpolation scheme. This is achieved by adding a regularization constant $\lambda$ to the diagonals of the correlation matrix $R$ in Eq. 2.18. The new correlation matrix is referred to as

$$R_{\text{reg}} = R + \lambda I$$  \hspace{1cm} (2.25)

This is similar to Tikhonov regularization, also known as Ridge Regression in the field of statistics [19]. The introduction of $\lambda$ introduces a bias in the estimation of the response and ultimately causes the $z(x)$ term in Eq. 2.14 to become deviations from the mean. With the addition of the new term, the estimation of response at locations $x$ becomes

$$\hat{y} = F\hat{\beta} + r^T R_{\text{reg}}^{-1} \left( y - F\hat{\beta} \right)$$  \hspace{1cm} (2.26)

where all of the terms are defined by the equations in the previous section with $R$ replaced by $R_{\text{reg}}$ [9]. A visualization of the new response with the same underlying deterministic function as Fig. 2.7 is presented in Fig. 2.8 to demonstrate the effect of the regularization constant on the $z(x)$ term.

From Fig. 2.8 it is easy to see the regression component of the response exhibits similar characteristics to the original with respect to the data. However, if the vertical lines from 2.8a and 2.8b are compared, it is evident that they are no longer of equal length. Therefore, as previously described, the stochastic component no longer fits the error exactly.
Regression Kriging adds simple flexibility to the traditional Kriging formulation. However, Regression Kriging’s expected mean-squared error does not perform as one might
expect. Figure 2.9 illustrates the collapse of the MSE (gray bounds) around the predicted mean response (red line). In turn, the MSE becomes highly nonconservative of the true three standard deviation bounds (green bounds) of the true response (black line) constructed from the dotted samples. Regression Kriging also trends to the global regression outside of the data fitted by the stochastic process. Therefore, Regression Kriging’s ability to extrapolate is similar to Kriging. The implications of the Regression Kriging formation will be discussed further with a potential application in the proposed methodology in the Future Work section.

An additional shortcoming of both the typical Kriging formulation and the Regression Kriging formulation is their inabilities to predict non-stationary mean responses and data with non-stationary variation in the case of Regression Kriging. This can become detrimental in many engineering applications. In Chapter 3 non-stationary surrogate models from both the engineering and geostatistical communities are discussed in detail.

2.3 Research Contribution

To remove the stationary response assumption from the Kriging framework, a new flexible and efficient framework was developed in this research. The new framework is called Locally Optimized Covariance Kriging (LOC-Kriging). The LOC-Kriging methodology approximates a non-stationary covariance structure by using multiple stationary covariance structures optimized for local function behaviors. Eliminating the need for the stationary response assumption. This work also resulted in an original Non-Stationary Identification Test capable of identifying localities of a NS covariance structure for deterministic responses. The possibility and realization of a physics based Kriging is also explored through the uses of a weighting function and the LOC-Kriging framework. This work also resulted in two American Institute of Aeronautic and Astronautic (AIAA) conferences papers [20] [21] and one AIAA journal article [22].
Non-Stationary Surrogate Modeling

In this chapter, the various methods to address modeling of non-stationary responses with surrogate models are discussed and demonstrated. However, first it is necessary to define a stationary response. In the context of this document, a stationary response for a deterministic computer simulation is defined as a response where the frequency of fluctuations do not change drastically or suddenly over the spatial domain. For a non-deterministic response, stationarity is defined as a stochastic process whose joint probability distribution does not change with time. Kriging specifically utilizes a stationary assumption when constructing its covariance structure. Violation of this assumption is discussed in detail in the following paragraphs.

Typically, the mean-squared-error (MSE) calculation performed in Kriging is estimated with a stationary covariance structure that is a distance correlation among samples. It is expected that the function being predicted has consistent or stationary fluctuations within the domain of interest. Therefore, a trend function and a stochastic random process with a single correlation function are applied to predict the underlying function. In most engineering applications, the correlation is typically modeled with a Gaussian function. To characterize the correlation among the sample data, the hyperparameters of the correlation function are optimized using the maximum likelihood approach along each dimension. Recently, to address the challenges of the maximum likelihood approach in finding model parameters the Generalized Pattern Search algorithm [18] (GPS) and Penalized Log-Likelihood function [23] (PLL) have been developed. However, hyperparameter op-
timization will likely fail to find a true optimum for a non-stationary function; in fact, the hyperparameter, if found, will be inadequate to capture non-stationary behaviors, resulting in inaccurate predictions and amplified mean square error estimations.

When there are a large number of samples, stationary Kriging can address non-stationary behaviors with increased computational costs. However, with an insufficient number of samples, Kriging may fail to find a globally well-fitted stationary covariance structure for a non-stationary function. This malfunction can be significantly amplified when data is scattered unevenly as a result of an adaptive sampling technique that is often used in engineering design exploration. In Figs. 3.1 and 3.2, a moderately non-stationary true function [18] is approximated with 17 adaptive samples and 17 uniform samples utilizing stationary Kriging. The stationary Kriging model utilizes the same Gaussian correlation function and a second-order polynomial global trend function as Xiong's demonstration of this function [18]. The MSE with ±3σ is calculated with an optimized stationary covariance structure and over-plotted to represent an unbiased uncertainty. It is noted the Kriging expected Mean Square Error (MSE) calculated with evenly collected samples produces a type-II statistical error, shown in Fig. 3.2. The Kriging prediction would miss the true function behavior within the low input range (input < 0.3) where the true function shows a relatively high-frequency behavior.
Figure 3.1: **Adaptive One-Dimensional Stationary Kriging Response.**

Figure 3.2: **Even One-Dimensional Stationary Kriging Response.**

Typically, more samples are needed to capture the rapidly changing system response of some functions. Figure 3.1 demonstrates an adaptive sampling process where more
samples are deployed within the low input range. For this function, the hyperparameter of the stationary covariance structure is heavily influenced by the data tightly clustered in the low input range. As a result, the Kriging model produces high fluctuations of the predicted response and severe amplification of the MSE within the high input range (input > 0.4) shown in Fig. 3.1.

From this example, it is clear the stationary covariance assumption could be inadequate to capture transitional non-stationary system behavior due to its introduction of unnecessary fluctuations and over-amplifications of both the response and MSE predictions. This can be generalized for all instances where the response is non-stationary. To address this, many Non-Stationary (NS) methods were proposed by researchers, especially in the fields of geostatistics and environmental processes. Both their work and research developed in the field of engineering are presented in the following section.

### 3.1 Literature Review

Both the geostatistics and environmental processes, and engineering communities have contributed significantly to the development of Non-Stationary (NS) methods. However, due to the dimensionality of their respective problems, they face significantly different challenges. Therefore, the solutions posed by each community are presented separately. The methods developed by the geostatistics community are presented first because they were generally developed first.

#### 3.1.1 Geostatistics Development

For simple NS structures, the geostatistics community offers a variety of methods. Isaaks and Srivastaya suggested a direct implementation method accomplished by using locally varying sills [24]. This method fails to capture the non-stationary behavior of complex and
varying structures. Sampson and Guttorp [25] proposed a nonlinear transformation method to obtain an approximate stationary covariance structure. This transformation method can be sensitive to variance values and provide poor or unstable predictions especially with a complex and multimodal non-stationary system.

Haas developed Moving Window Kriging (MWK) [26]. MWK is a method in which a covariance structure is constructed within a circular neighborhood centering at an estimation point where a previous global DOE was constructed. Figure 3.3 is an illustration of MWK. As the prediction region moves though the design space an optimal local variogram or semivariogram structure is constructed and are formed under the local stationary assumption. The optimal size of the moving window is determined by model bias statistics such as root-mean-squared error (RMSE) and confidence bounds.

![Figure 3.3: Moving Window Kriging](image)

Harris [28] proposed a geographically weighted variogram (GWV) MWK to smooth the individual variograms by using a kernel function with an optimal inverse distance-weighting scheme. Since an optimal window size and a new covariance are calculated as the prediction location moves through the space, the computational cost of the method could be prohibitive for high dimensional problems. Most of the proposed methods mentioned from the geostatistics community were aimed at situations where a significant number of
samples are available within a low-dimensional design space. Therefore, they are not well suited for large-scale, high-dimensional engineering design exploration.

### 3.1.2 Engineering Development

For engineering design applications, Lin, et al.[29] proposed the Sequential Exploratory Experimental Design (SEED) in which the entries of the covariance matrix are adjusted by using the previous correlation parameter information over the course of sequential data sampling. The main focus of their work was to optimize adaptive sampling in the SEED process, but not to implement a NS-Kriging model. However, the method is notable in its ability to construct NS covariance matrices.

Ba and Joseph [30] developed the Composite Gaussian Process (CGP) model, in which two stationary Gaussian processes for global trend and local variation are used to address non-stationary system behavior. This is accomplished by defining the input regions based on the space-filling properties of data. However, the CGP model needs to optimize a vector of hyperparameters and three unknown parameters to fit both global and local processes. This optimization is accomplished in one sequence but requires relatively large amounts of data. To address the difficulties with high-dimensional engineering problems, Xiong, et al.[18] adopted the nonlinear map approach to convert a NS covariance structure into an approximated stationary structure.

Xiang’s approximation was accomplished using parameterized density function with predefined knots or local density functions. A conceptual illustration of the nonlinear map approach is shown in Fig. 3.4. In practice the continuous density function illustrated below the original space, is approximated as a piecewise density function. This method can be computationally intensive depending on the number of knots [31]. Also, the approximated univariate density functions may become a major source of error when a density function becomes complex with multimodal non-stationarity.
The one-dimensional example problem developed by Xiong, et al. [18] is used as a validation case for the method presented later in this thesis. With 8 knots, they were able to achieve a 0.0109 RMSE error. Later it is demonstrated that Locally Optimized Covariance Kriging produces a 0.0175 RMSE value with only 3 local models.

3.2 Summary

In summary, there are numerous ways to address responses where the frequency of fluctuations changes drastically or suddenly over a domain of interest also known as non-stationary responses. The methods range from coordinate transformations to modifications of surrogate covariance structures. Every method has its merits in addressing NS behavior, however, most suffer from the “curse of dimensionality”. To address this challenge, a new flexible and efficient framework of Locally Optimized Covariance Kriging (LOC-Kriging) is proposed. The proposed LOC-Kriging approximates a NS covariance structure by using multiple stationary covariance structures optimized for local function behaviors. A statistical test process using a set of model training points is proposed to identify localities of a NS covariance structure. The prediction of a NS function behavior is estimated by blending multiple LOC-Krigings based on their local membership functions. In this study, it is also discussed how the physical understanding, such as global sensitivities of the specific system behaviors, can be implemented within the proposed framework of LOC-Kriging to enhance engineering design exploration.
Locally-Optimized Covariance Kriging

4.1 Formulation

To alleviate the computational difficulties using the NS covariance structure for practical engineering application, Locally Optimized Covariance Kriging (LOC-Kriging) is proposed. In this method, multiple local stationary structures are identified and used to approximate the global non-stationary covariance structure. Unlike the aforementioned MWK method, which moves a local window along with a prediction point, LOC-Kriging constructs a finite number of local stationary structures according to the localities of function behaviors. The local window sizes are simultaneously optimized using an aggregated maximum likelihood function. The center locations of the localities can be user-defined based on prior knowledge or identified by the proposed local statistical testing method described in Section 4.2. The prediction of the LOC-Kriging is obtained by combining multiple local stationary models based on an aggregation membership function.

4.1.1 Kriging Model with Locally-Optimized Covariance

In the construction of the local Kriging model, the bell-shaped membership function shown in Fig. 4.1 is used to apply the degree of membership to samples within the identified locality windows. This function maintains the continuity of the predicted response across the finite local window boundaries. The samples within the range of full membership
have unit weightings. Over the full membership boundary, there is the transitional range defined by $\alpha x \omega$, where $\alpha$ is the scale factor of the transition range. When implemented, membership, both full and transitional, is determined by the Euclidean distance between the center of the window and the sample of interest. This is true for any dimensionality because the local window is represented as a hypersphere. In this study, the Gaussian function is used to vary the membership between unity and zero within the transition range. In practice, any transition function can be used, such as a linear, spline, or an exponential function depending on the desired smoothness.

![LOC Window Membership Function](image)

Figure 4.1: **LOC Window Membership Function.** The range from -1 one to +1 is within the window, and the transitional range extends a distance proportional to the radius.

With the membership weightings applied to the samples in a local window, the locally weighted Kriging model is constructed by the generalized least square regression as

$$
\phi = E \left[ W(s) \left( \hat{y}(s) - y(s) \right)^2 \right]
$$

(4.1)

where $W$ is the diagonal matrix of the membership weights. The weighted regression
coefficient vector, $\hat{\beta}_w$ is calculated using Eq. 4.2.

$$\hat{\beta}_w = \left( F^T R_w^{-1} F \right)^{-1} F^T R_w^{-1} y$$  \hspace{1cm} (4.2)

Here $R_w^{-1} = \left( \sqrt{W} \right)^T R^{-1} \left( \sqrt{W} \right)$. The LOC-Kriging prediction, $\hat{y}_1$, at an untried location $x$, is obtained as

$$\hat{y}_1 = f \hat{\beta}_w + r_w^T R_w^{-1} \left( y - F \hat{\beta}_w \right)$$  \hspace{1cm} (4.3)

where $r_w = \left( \sqrt{W} \right)^{-1} r$. Utilizing the weighting membership function as described allows the local model to maintain their interpolation behavior within the full membership. The membership function also results in an informed extrapolation range within the scaled tails for aggregation between local models. This decreasing weight region can also be thought of as a fading interpolation. The following section explains how to optimize the local window size of multiple models.

### 4.1.2 Local Window Size Optimization

To avoid an over-parameterization in the global regression and to minimize the finite number of the local models, the local window sizes are optimized as

**Find:** $[\omega_1, \omega_2, \ldots, \omega_l]^T$

**Minimize:** $\Phi = \sum_{i=1}^{l} \phi_i(\theta)$

**Subject to:** $g_i = \omega_i - \omega_{min} \geq 0$

\hspace{1cm} $g_{l+1} = \lambda - \lambda_{min} \geq 0$

\hspace{1cm} $g_{l+2} = -\lambda + \lambda_{max} \geq 0$  \hspace{1cm} (4.4)
where $\omega$ is the local window size measured by the ratio between the current window and the entire design space; $\omega_{\text{min}}$ is the required minimum size; $\Phi$ is the aggregated likelihood function; $\lambda$ is the global coverage parameter with user-provided upper and lower bounds, $\lambda_{\text{min}}$ and $\lambda_{\text{max}}$. Nominally these quantities are selected to be 0.1 and 0.5 respectively. With a small quantity of data, less than five points per dimension, it is suggested to increase $\lambda_{\text{max}}$ accordingly. The local window size $\omega$ can be viewed as a hypersphere radius in a multidimensional problem.

In the optimization formulation, an aggregated likelihood function $\Phi$ is considered as the objective function to capture the global fitness of the combined local covariance structures. In this study, the uniform aggregation of the local likelihood measurements is used. However, any aggregation strategy can be implemented to reflect one’s understanding of the underlying function behavior with a different importance scheme. The global coverage parameter, $\lambda$, represents the amount of overlap that occurs between windows or simply put, it is a measure of double counted/claimed sample points by multiple windows. This is accomplished by

$$
\lambda = \frac{\sum_{i=1}^{l} W_i}{lN}
$$

where the numerator is the sum of the diagonal matrix of the membership weights $W_i$ for the $i^{th}$ model, $l$ is the number of models, and $N$ is the number of sample points. Like stationary Kriging, this formulation cannot ensure the entire domain is covered with full membership; this is only true if the sample points are representative of the domain. However, as long as the localities are statistically significant, the global coverage parameter guarantees overlap between localities and the transitional scale factor from the previous section ensures a smooth transition.

Any prior information or knowledge regarding the system’s local behavior can be implemented into the minimum window size requirements. The minimum window size
should also be determined by considering the order of the basis function. This ensures
the total number of available samples does not result in over fitting. The upper and lower
bounds of the global coverage parameter should be selected to yield continuous transitions
across the local windows.

In this study, the interior point optimization algorithm is used to find the solution of
the optimization problem formulated in Eq. 4.4 in a reliable and efficient manner. Note
that the optimization in which local Krigings are constructed and tested can benefit from
parallel computing to maximize the computational efficiency.

4.1.3 Construction and Aggregation of Multiple Local Kriging Models

With the expectation of mutually overlapped LOC-Krigings over the entire design domain,
it is important to maintain a smooth and continuous transition across local windows. The
aggregation method used to combine the prediction of the models is essentially a weighted
average as

\[
\hat{y}_a(x) = \frac{\sum_{i=1}^{l} \hat{y}_i(x) \gamma_i(x)}{\sum_{i=1}^{l} \gamma_i(x)}
\]  

(4.6)

where \( l \) is the number of local models; \( \hat{y}_i \) is the prediction array from the \( i^{th} \) local model,
and \( \gamma_i \) is the weight array of the local model at each prediction site. The weight of the
aggregation response is a modified version of the membership function shown in Fig. 4.1.
The full membership region is reduced by the tail factor \( \alpha \), decreasing the weight of the less
accurate tail region of the curve as shown in Fig. 4.2. This ensures the tails have
significantly less weight when aggregating multiple windows while maintaining a smooth
transition. When two or more full members are combined, Eq. 4.6 provides an average of
the windows. This same aggregation technique is applied to the estimated variance.
4.1.4 Imposing Physics on LOC-Kriging

In the typical Kriging framework covered in Chapter 2, Kriging is composed of a trend function and stochastic process, Eq. 2.14. If the general trend of the physical process is known, an appropriate regression can be selected. Many physical processes have a suggested trend line. For example, fatigue data, the base ten logarithm of cyclic life as a function of stress amplitude, is typically fitted by a linear function. However, when fitted with Kriging, the stochastic process may cause unnecessary fluctuations thus, violating the fundamental physical behavior. Other more complex interactions such as fatigue-coupled creep are understood to have a monotonically decreasing response while not having a closed form regression function. Both of these cases can be addressed by relaxing the contribution from the stochastic process causing the trend behavior to become more prominent between training points as

$$\hat{y}_1 = f\hat{\beta}_w + r_w^TR_w^{-1}(y - F\hat{\beta}_w)\delta(s, x, \zeta) \tag{4.7}$$
here, $\delta$ is the stochastic process relaxation function. This relaxation process is not possible with traditional Kriging due to a single global trend function covering the entire domain. However, the proposed framework of LOC-Kriging provides the necessary computational flexibility. When the relaxation function is applied, the stochastic deviations from the local trend for individual LOC windows are reduced. The relaxation function, $\delta$, is formulated in Eq. 4.8.

$$\delta(s, x, \zeta) = 1 - \frac{\min \{||x - x||\}}{\frac{1}{2} \max \{\min \{||s_i - s_j||\}\}} (1 - \zeta) \quad (4.8)$$

Here, the numerator of the fraction is a vector of the minimum distance between each prediction site and all of the sample points. The denominator is a scalar of the maximum of the minimum distances between sample points divided by two, where $i \neq j$, and $\zeta$ is the optimal global relaxation parameter. The optimal global relaxation parameter is determined to satisfy the fundamental behavior of the underlying function of interest, such as a monotonic behavior in a fatigue model. This additional relaxation function allows for the implementation of known physics into the Kriging framework while maintaining Kriging’s interpolation characteristic at sample points.

### 4.2 Non-stationary Identification Test

Due to the increased computational cost of approximating the non-stationary covariance structure, it is important to establish a framework to validate the additional cost. The framework is divided into five sequential steps:

1. Deploy virtual test points
2. Perform user prescribed regression
3. Calculate the root-mean-squared error (RMSE) at each local regression
4. Perform cluster analysis on the RMSE response

5. Determine the spatial statistical significance of the clusters

The fundamental idea is that the regression will produce similar model bias statistics within different local regions of similar nonlinearities. A basic outline of the framework used in the two-dimensional space is defined in Fig. 4.3. This framework is critical when considering the utilization of LOC-Kriging in a high dimensional space due to its ability to define localities. This is important because the number of localities is not dimensionally dependent, but it is dependent on both the nature of the problem and the collected samples. For example, a 1-D problem can have five localities whereas a 7-D problem can have four localities.

![Non-stationary Identification Test Outline](image)

Figure 4.3: **Non-stationary Identification Test Outline.** Illustrative preprocessor outline for the two-dimensional example with specific virtual points and regression model labeled.

The first step of the process is to deploy virtual test points between the sample points. In this study, linearly spaced points are selected for one-dimensional space. In higher
dimensional space, Voronoi vertices [32] can be selected by creating a Voronoi diagram around the sample points. The vertices can also be supplemented with another space filling method if desired. In Fig. 4.4 the Voronoi vertices are represented by the large lightly colored circles while the function evaluation points are small black circles. Within local windows centering individual virtual test points, the user prescribes the local regression models to quantify the local RMSEs.

![Figure 4.4: Voronoi Test Points With Contour. Voronoi vertices used as test sites for the illustrative example with response contour and sample points.](image)

The window of the local regression is the same for each vertex. The window size is either user-provided or adjusted to include a minimum number of sample points for the estimations of non-zero local RMSEs. If little is known about the system, it is suggested that the same order regression model is utilized as the global trend function in Kriging. As an illustrative two-dimensional example, shown in Fig. 4.5, a linear regression with the minimum four sample points is used. One of the regression models is shown as a solid red plane with respect to the meshed surface of the true response in Fig. 4.5a. A contour of the local regression with its regression data points captured by the window is shown in Fig. 4.5b.
Figure 4.5: **Linear Regression Model.** (a) Meshed view of the red regression plane and (b) Contour of the regression with window indicated by a circle.

With the local regression model selected, the RMSE is calculated and considered as the local stationary measurement at each Voronoi vertex or center. With the RMSE measurements, the K-means [33] clustering algorithm is used to determine the locality of the system response. The K-means clustering algorithm is a long-standing, robust, unsupervised learning algorithm, capable of identifying a user defined number of clusters based on sample distances. The vertices that make up the individual clusters are then averaged to create center points of potential localities. The initial number of center points was assumed to be two, indicated by stars in Fig. 4.6. The triangles and diamonds illustrate to which cluster each point belongs.
Figure 4.6: **Initial Localities.** Two black stars identify the initial localities; diamonds and triangles indicate points that belonging to each clusters.

The spatial coordinates of the initial center points are compared to the distribution of the obtained cluster center points by using hypothesized mean student t-test [34] with a 95% confidence interval. This comparison is done on a dimension-by-dimension basis, which means the \(x_i\)-coordinate of each cluster is compared to the \(x_i\)-coordinates of the samples that make up every cluster. This is then repeated for every dimension. For an example, the matrix for the \(x\) dimension is constructed with an arbitrary number of clusters \(n\), as

\[
D_{x_i} = \begin{bmatrix}
ttest(\mu_1, C_1) & \cdots & ttest(\mu_n, C_1) \\
\vdots & \ddots & \vdots \\
ttest(\mu_1, C_n) & \cdots & ttest(\mu_n, C_n)
\end{bmatrix}
\]  

(4.9)

where \(\mu_i\) is the center coordinate of the \(i^{th}\) cluster in dimension \(x\); and \(C_i\) is the \(x\)-coordinates of samples from the \(i^{th}\) cluster. The null hypothesis is: the mean of one window matches the population of another. If one fails to reject the null, the test returns a 0. If the null is rejected, i.e. the mean does not match the population of another the test returns a 1. Non-stationary is determined to be significant if 50% or more of the total test are rejected, this is true for each dimension. If significant, the cluster center points should be considered to reflect the localities of the function behavior. Otherwise, a stationary Kriging model is
sufficient and LOC-Kriging will be constructed with a single window. If only one dimension is significant, the K-means algorithm is rerun on the spatial coordinates of the clusters. This helps to ensure a space filling distribution of center points. However, it is possible that the given samples do not present any non-stationary behavior. This means the results of the proposed LOC-Kriging does depend on the given samples.

In the illustrated example in Fig. 4.6, the $X_1$ coordinates are found to be significantly different while the $X_2$ coordinates were not. Therefore, the K-means algorithm is used on the spatial coordinates of each cluster separately to ensure a space filling distribution of center points. Figure 4.7 illustrates the four local windows indicated by large circles centering the black stars. These four windows will be used as initial local windows in the proposed LOC-Kriging method.

Figure 4.7: Four Identified Localities. Four black stars identify the final localities; black and lighter diamonds, and triangles indicate points that belonging to clusters.

4.3 Demonstation

In this section, the performance of the proposed LOC-Kriging method is compared to traditional stationary Kriging through the use of representative examples of engineering applications. The basic concepts and details of LOC-Kriging are explained through the use of a simple one-dimensional mathematical problem. Then, a two-dimensional case is pre-
presented to demonstrate the potential of LOC-Kriging. A fatigue and creep-testing scenario with zero variation is presented utilizing empirical damage model equations to illustrate the flexibility of LOC-Kriging. Lastly, a five-dimensional fluid-structure interaction problem is presented to demonstrate the accuracy and applicability to a high-dimensional problem.

4.3.1 One-Dimensional Example

The simple one-dimensional mathematical example is given by Eq. 4.10 is considered first. In Fig. ??a, the true response (black dot-dash line) shows a non-stationary behavior within the $x$ range of interest. The stationary Kriging response and estimated error bounds are depicted with a solid red line and filled gray region respectively along with the adaptively collected samples. By using conventional Kriging, the error bounds are unnecessarily amplified in the second half input range as shown in Fig. ??a. This amplification is due to the assumption that the underlying function behavior is stationary. The regression of the Kriging prediction is assumed to be a second order polynomial, the same as Xiong [18].

$$y(x) = \sin(30(x - 0.9)^4) \cos(2(x - 0.9)) + \frac{x - 0.9}{2}$$

Based on the process described in the previous section: a linear regression is selected to capture the changing behavior, three K-means clusters are prescribed, and a 95% confidence is selected to compare the similarities of the spatial components of the clusters. This resulted in the identification of three local covariance structures. Their optimal range of coverage is found by solving the optimization problem formulated in Eq. 4.4 with the Interior-Point Algorithm. The LOC models were built in parallel during each optimization iteration to decrease computational time. The resulting sizes of the three LOCs are shown in Fig. 4.8. The triangles indicate the center locations of local windows, and the full membership range is identified using arrows. The three models are built using a second-order polynomial regression and Gaussian correlation function. The models are labeled as LOC-
1, LOC-2, and LOC-3 in the figure. The legend is the same as Fig. ??.

The three LOC models are weighted according to their individual aggregation membership functions.

Figure 4.8: 1-D LOC Predictions. LOC-Krigings for LOC-1, 2 and 3 with corresponding ranges signified via black arrows.

By aggregating the three LOC-Krigings, the approximated non-stationary Kriging prediction is obtained and compared to the stationary Kriging response side by side in Fig. 4.9. Figure 4.9 visibly demonstrates LOC-Kriging’s ability to produce more reasonable and consistent predictions than stationary Kriging. The plot also showcases LOC-Kriging’s error bounds that match our physical understanding within both the low and high input ranges unlike stationary Kriging. Cross-validation can be employed to access the validity of the obtained error bounds.

Figure 4.9: 1-D Comparison of Methods. (a) Stationary Kriging and (b) Aggregated LOC-Kriging.

In this study, performance is measured using three metrics: The root-mean-square
error (RMSE), the maximum standard deviation of the prediction, and the integral of the expected mean-squared error. The RMSE between the true and predicted Kriging response is calculated as

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_{true,i} - \hat{y}_i)^2}
\]  

(4.11)

where \(N\) is the total number of test data points; \(y_{true}\) is the true response, and \(\hat{y}\) is the Kriging prediction. Equation 4.12 defines the improvement between the two methods as

\[
Improvement(\%) = \frac{Stationary - LOC}{Stationary} \times 100
\]  

(4.12)

The quantitative results of the current example are summarized in Table 4.1. The RMSE was evaluated using 4000 evenly spaced testing points. The aggregated LOC-Kriging shows 74.10\% RMSE improvement in the response prediction against the stationary Kriging. There is a maximum reduction of the standard deviation of 82.40\% and most notably, the integral of the MSE is reduced by almost 100\%. This was achieved by relaxing the strong assumption of stationary covariance within the design domain and reflecting localities of function behaviors properly. Computational wall times are presented in Table 4.2 in seconds. All simulations were performed utilizing the MATLAB on desktop PC with an Intel Core i7-5820K Haswell-E 3.3GHz LGA 2011-v3 Processor and 16 GB of DDR4 2666MHz RAM.
Table 4.1: 1-D Performance Comparison. Root-mean-square error, maximum standard deviation, and integral of the MSE performance measures between Stationary Kriging and LOC-Kriging.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Stationary Kriging</th>
<th>LOC-Kriging</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>6.753E-02</td>
<td>1.749E-02</td>
<td>74.10%</td>
</tr>
<tr>
<td>Max $\sigma$</td>
<td>3.594E-01</td>
<td>6.325E-02</td>
<td>82.40%</td>
</tr>
<tr>
<td>Integral MSE</td>
<td>3.382E-02</td>
<td>5.579E-04</td>
<td>98.35%</td>
</tr>
</tbody>
</table>

Table 4.2: 1-D Computational Times. Reported wall times in seconds between Stationary Kriging and LOC-Kriging.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Stationary Kriging</th>
<th>LOC-Kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Model Build Time (s)</td>
<td>0.0426</td>
<td>0.0585</td>
</tr>
<tr>
<td>Optimal Model Build Time (s)</td>
<td>–</td>
<td>2.3820</td>
</tr>
<tr>
<td>Prediction Time (s)</td>
<td>0.0203</td>
<td>0.0727</td>
</tr>
</tbody>
</table>

The convergence plots for LOC-Kriging’s window optimizations objective history and window size history is presented in Fig. 4.10. The convergence histories show there were a required 45 parallel model builds to reach a converged solution. This translates to LOC-Kriging taking 2.38 seconds to arrive at an optimal model. However, this cost is dependent on the convergence criteria of the optimization loop, initial window sizes and the number of localities. In this example, the time required for LOC-Kriging to build three parallel models is 0.0585 seconds, this is almost identical to the single stationary Kriging build time. Depending on the problem it is possible to use the initial windows to generate a prediction making computational times very similar. Now, if large-scale problems are considered, the matrix inversion times for a single Stationary model can become computationally expensive. Large-scale stationary Kriging models often suffer from numerical instabilities as
well. Whereas the local models in LOC-Kriging are constructed in parallel and are significantly smaller. Therefore, as problems become more complex LOC-Kriging will become more advantageous compared to stationary Kriging. LOC-Kriging can also be further optimized. For example, in Fig. 4.10b, LOC-3s window size does not change after model build 20, but in the current formulation the model is rebuilt at each iteration. Nevertheless, considering the high computational costs of generating samples by running engineering simulations such as Finite Element Analysis (FEA) or Fluid Structural Interaction (FSI), the costs of building LOC-Kriging models could be marginal.

![Objective History](image1)
![Window Size History](image2)

**Figure 4.10: 1-D Convergence Histories.** (a) Objective function histories and (b) Window size as a diameter for all three windows.

### 4.3.2 Two-Dimensional Example

To demonstrate the performance of the proposed LOC-Kriging, a two-dimensional mathematical example is presented. The mathematical function has polynomial and trigonometric terms, and interactions between $x_1$ and $x_2$ as shown in Eq. 4.13.

$$y(x_1, x_2) = \sin(21(x_1 - 0.9)^4) \cos(2(x_1 - 0.9)) + \frac{x - 0.7}{2} + 2x_2^2 \sin(x_1 x_2) \quad (4.13)$$
The true function shows distinctive non-stationary behavior with respect to the input variable $x_1$ as shown in Fig. 4.11. A similar behavior can be observed when plotting aerodynamic drag of an airfoil as a function of angle of attack and Mach number when considering the transitional Mach regime. For this example, twenty-four samples are collected adaptively focusing on the region of fluctuating response. It is suspected Ordinary Kriging, or Kriging with a Gaussian correlation function and zero order regression, will have difficulties finding a single optimum covariance structure that covers the entire domain. With the optimum set of stationary correlation parameters, Fig. 4.12 shows an over-plot of the meshed surface of the stationary Kriging prediction against the solid red surface of the true response.

Figure 4.11: 2-D Adaptively Collected Data.
Figure 4.12: **2-D Adaptively Collected Data.** (a) Kriging prediction (meshed) against the true surface (red) and (b) contour of the stationary Kriging prediction.

Figure 4.13 shows the estimated ±3σ variation bounds that are unnecessarily amplified due to the stationary requirement on the covariance structure. The variation bounds are much larger with respect to the response; they are likely to be misleading if utilized in an iterative design exploration process. With the unnecessarily amplified bounds, one may seek a minimum in the high $x_1$ and low $x_2$ range of the domain since the lower MSE bound in the range is the minimum in the entire domain.

Figure 4.13: **Upper and Lower Estimated Error Bounds, Stationary.** Upper and lower estimated error bounds ($\pm3\sigma$, meshed surfaces) by stationary Kriging.

With the given set of samples collected unevenly, the non-stationary identification test
was utilized with two initial windows. As illustrated in Section 4.2 in detail, four LOC-Krigings are obtained centering the points marked by black stars as shown in Fig. 4.14. The optimum coverages and correlation parameters of the LOC-Krigings are obtained by solving the optimization problem as described in Eq. 4.4.

Figure 4.14: 2-D LOC-Kriging models. Four LOC-Kriging models and their optimum full membership coverage over top of the true contour.

The LOC-Kriging predictions are aggregated based on the weighting function is given by Eq. 4.6. The following figure shows the over-plot of the aggregated LOC-Kriging predictions represented with the meshed surface against the true responses indicated with the solid red surface. The performance metrics for the two-dimensional function are evaluated on a $120 \times 120$ point grid to ensure the rapidly changing region is captured. The performance metric comparisons between LOC-Kriging and Ordinary Kriging are shown in Table 4.3. The computational wall times are presented in Table 4.4 and are similar to the times of example 1. The four LOC-Kriging models are constructed in about 0.046 seconds in each iteration, this is almost the same time it takes to create a single stationary Kriging model, and the LOC-Kriging takes slightly longer than stationary Kriging to make a prediction. The model build time for LOC-Kriging represents 39 model builds to arrive at a converged solution in 2.8 seconds.
Figure 4.15: **2-D Over-plot and Contour, LOC-Kriging.** (a) Over-plot of the LOC-Kriging prediction (meshed) against the true surface (red) and (b) Contour plot of the LOC-Kriging prediction.

Table 4.3: **2-D Performance Comparison.** Root-mean-square error, maximum standard deviation, and integral of the MSE performance measures between Stationary Kriging and LOC-Kriging.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Stationary Kriging</th>
<th>LOC-Kriging</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.2896</td>
<td>0.0842</td>
<td>70.93%</td>
</tr>
<tr>
<td>Max (\sigma)</td>
<td>0.5693</td>
<td>0.2781</td>
<td>51.15%</td>
</tr>
<tr>
<td>Integral MSE</td>
<td>0.0629</td>
<td>0.0149</td>
<td>76.31%</td>
</tr>
</tbody>
</table>

Table 4.4: **2-D Computational Times.** Reported wall times in seconds between Stationary Kriging and LOC-Kriging.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Stationary Kriging</th>
<th>LOC-Kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Model Build Time (s)</td>
<td>0.0417</td>
<td>0.0460</td>
</tr>
<tr>
<td>Optimal Model Build Time (s)</td>
<td>–</td>
<td>2.7983</td>
</tr>
<tr>
<td>Prediction Time (s)</td>
<td>0.0208</td>
<td>0.0619</td>
</tr>
</tbody>
</table>

As a result, the study shows LOC-Kriging produces a 70.93% more accurate result.
than stationary Kriging. More importantly, the estimated $\pm 3\sigma$ variation bounds from LOC-Kriging provide a more accurate representation of the uncertainty based on the collected samples. This will be meaningful especially in a design exploration with an adaptive sampling scheme. This is accomplished by eliminating the stationary covariance assumption and is shown in Fig. 16. The estimated maximum error from LOC-Kriging is reduced by 51.15% and the integral of the MSE is reduced by 76.31% compared to that from the stationary Kriging.

![Figure 4.16: Upper and Lower Estimated Error Bounds, LOC-Kriging.](image)

Upper and lower estimated error bounds ($\pm 3\sigma$, meshed surfaces) by LOC-Kriging.

### 4.3.3 Ti-6242S Fatigue and Creep Coupled Numerical Experiment

Hypersonic aircraft provide the strategic potential to perform an aerospace mission anywhere in the world within one hour. The United States Air Force has invested considerable efforts to develop and demonstrate the effectiveness of hypersonic technologies. Lockheed Martin Aerospace recently entered a 3-phase project with the objective of evaluating current predictive capabilities for hypersonic structural response and life prediction. Phase I and II have been approved for public release [35], [36]. Lockheed Martin investigated the usage of the DARPA Falcon program developed HTV-3X loft Fig. 4.17, for its translation into an operational vehicle with the hypothetical cruise profile of 30 minutes at Mach 5.2.
From a structural science and technology perspective, fatigue and creep predictions under the repetitive thermo-mechanical acoustic loadings from the extreme harsh flight conditions are essential to assess flight critical damage states with high confidence. Therefore, collecting high-quality material data from physical experiments is of the utmost importance to achieve high confidence. If LOC-Kriging can reduce the amount of physical experiments by even a small fraction, the savings would be sizable. Panel 3, for example, is constructed from sheets of Ti-6Al-2Sn-4Zr-2Mo-0.1Si and a Ti-BETA21S core in Fig. 4.17. Both materials are relatively expensive to produce but more importantly, the amount of material information necessary for only a single composite component can become enormous when considering a large variety of loading conditions at varying temperature conditions.

The proposed LOC-Kriging will be used to construct the fatigue-creep model and to assess the model fitness for an optimal data collection plan. The experimental procedure is an idealized uniaxial fatigue and creep coupled testing scenario. The material being tested is Ti-6Al-2Sn-4Zr-2Mo-0.1Si. The test is conducted utilizing empirical equations, a damage model, and limited sample points at various temperatures. The loading scenario involves one fully reversible fatigue cycle followed by an equally long creep period at the maximum stress amplitude of the cycle. The ranges investigated are 72 to 900 °F and 15 to 160 ksi. An illustrative load example is shown in Fig. 4.18 with unit stress amplitude.
Figure 4.17: **HTV-3X Aircraft Frame** The panels are labeled one through four, panel three is the panel of interest. Figure 3.0.1 from Reference [36].

Figure 4.18: **Loading Scenario.** Fatigue cycle followed by creep

The Manson and Halford [37] empirical model represents fatigue life under various strains in Eq. 4.14. Here, $\Delta \epsilon$ is the total strain (in/in); $\epsilon'_f$ is a fatigue ductility coefficient (in/in); $\sigma'_f$ is a fatigue strength coefficient (ksi); $\sigma_m$ is mean stress (ksi); $E$ is the elastic modulus (ksi), and $b$ and $c$ are material constants.

$$\frac{\Delta \epsilon}{2} = \frac{\sigma'_f - \sigma_m}{E} \left(2N_i^{(fat)}\right)^b + \epsilon'_f \left(\frac{\sigma'_f - \sigma_m}{\sigma'_f}\right)^c \left(2N_i^{(fat)}\right)^c$$  

(4.14)

The Robinson rule for cumulative creep and a creep rupture rule are utilized as the empirical
model for creep life as

$$\frac{1}{N_i^{(cr)}} = t_{cyc} e^{\exp \left( C - \frac{1000}{T + 273.15} LMP \right)}$$

(4.15)

where the Larson-Miller Parameter (LMP) is exploited to represent viable creep rupture rules across the applicable temperature range; $T$ is in units $F$; $t_{cyc}$ is the cycle time in hours; and $C$ is a material constant. The various material parameters for both Eq. 4.14 and 4.15 are presented in detail by Gordon [38]. To combine the two curves, Palgren-Miner rule for cumulative damage is implemented in Eq. 4.16.

$$D_{total} = \frac{1}{N_i^{(cr)}} + \frac{1}{N_i^{(fat)}}$$

(4.16)

The combined analytical life model is represented in Fig. 4.19, where $X_1$ is temperature and $X_2$ is stress amplitude. The life model will be assumed as the true model for comparison. Now, consider a situation where a life model is constructed with the 19 available samples depicted by the black dots in Fig. 4.19. The samples from the true model are assumed as statistical percentile life measurements after accounting sampling randomness in data. Since the true surface is highly nonlinear, a Kriging model is used to capture the combined life response.
Figure 4.19: **Analytical Life Model.** Log10 life prediction at various combinations of temperature and stress amplitude.

Figure 4.20 represents the conventional stationary Kriging response with a first order trend function and a Gaussian correlation function. Conventional Kriging can be detrimental when utilized in engineering design exploration causing the trend to miss represent the fundamental physics. Again, the surface of the Kriging prediction is the meshed surface and the true surface is indicated by solid red in the normalized design space. From the contour, highlighted by a bold red line in Fig 4.20, it is observed that the predicted fatigue and creep life from the stationary Kriging decreases, increases, and decreases again as stress amplitude increases. Based on fundamental physics, fatigue and creep should show monotonically decreasing behavior as temperature or stress amplitude increase.
To capture the fatigue life behavior based on our understanding of fundamental physics, four LOC windows are identified and optimized by the proposed process as shown in Fig 4.21. From the new contour it is immediately evident the LOC methodology produces a more accurate prediction utilizing the same local linear regressions and Gaussian correlation functions. However, the contour still shows the discrepancy from the expected monotonic behavior. To address this issue, the weighting function, $\delta$ is used to relax the stochastic process as illustrated in Section 4.1.4. The relaxation is used to generate a Physics-Informed (PI) LOC-Kriging model shown in the bottom of Fig. 4.21.

The performance comparison of Physics-Informed LOC-Kriging and stationary Kriging is shown in Table 4.5. The relaxed LOC-Kriging method not only improves all three-performance measures, but it also produces a more physically meaningful prediction than stationary Kriging. The computational time comparisons are shown in Table 4.6. Again, LOC-Kriging’s model build time is represented by 35 model builds in 2.7 seconds to obtain a converged solution and stationary Kriging’s time of 0.038 seconds is a single build. For this particular example, LOC-Kriging constructs the four parallel windows in 0.045 seconds, only 0.0073 seconds slower than stationary Kriging. In this example the underlying physics of the fatigue-creep interaction is imposed on the surrogate utilizing an additional
weighting function. The proposed function is kept general so that it can be implemented in many different aerospace applications where the physical trend is well understood and a regression model is inadequate.

Table 4.5: **Life Model Performance Comparison.** Root-mean-square error, maximum standard deviation, and integral of the MSE performance measures between Stationary Kriging and PI-LOC-Kriging.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Stationary Kriging</th>
<th>PI-LOC-Kriging</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>1.1468</td>
<td>0.5818</td>
<td>49.27%</td>
</tr>
<tr>
<td>Max σ</td>
<td>0.6357</td>
<td>0.4667</td>
<td>26.58%</td>
</tr>
<tr>
<td>Integral MSE</td>
<td>0.1939</td>
<td>0.1286</td>
<td>33.68%</td>
</tr>
</tbody>
</table>

Table 4.6: **Life Model Computational Times.** Reported wall times in seconds between Stationary Kriging and PI-LOC-Kriging.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Stationary Kriging</th>
<th>PI-LOC-Kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Model Build Time (s)</td>
<td>0.0377</td>
<td>0.0450</td>
</tr>
<tr>
<td>Optimal Model Build Time (s)</td>
<td>–</td>
<td>2.6940</td>
</tr>
<tr>
<td>Prediction Time (s)</td>
<td>0.0927</td>
<td>0.2266</td>
</tr>
</tbody>
</table>
4.3.4 Five-Dimensional Fluid Structural Interaction Example

Understanding the interaction of immersed elastic structure with surrounding fluid has applications in many fields of engineering such as: the stability and response of aircraft wings and the vibration of turbine blades. To understand these phenomena it is necessary to model both the structure and the fluid.

Aeroelasticity is a specific class of fluid-structure interaction problems. It studies the effect of aerodynamic forces on elastic bodies. The aeroelasticity coupling mechanism can be explained as follows. The aerodynamic forces acting on an aircraft depend on the orientation of its lifting surfaces with respect to the flow. The orientation of the aircraft...
depends on its elastic deformations. Hence, the magnitude of the aerodynamic forces cannot be known until the elastic deformations they induce are first determined. It follows that the external loads cannot be evaluated until the coupled aeroelastic problem is solved [39]. The ability to accurately predict the coupled aeroelastic response of aircraft is essential for developing high performance, safe designs. There are a number of computational methods for achieving this objective. Many computational methods such as doublet-lattice method and piston theory are based on the linear aeroelastic assumptions [40]. However, in designing such systems, the large shape change of the lifting surfaces and deformation of structure can produce noticeable changes in the aeroelastic behavior. This behavior can be accounted for only by using a rigorous nonlinear analysis [41].

One of the most powerful computational tools available for aerodynamic analysis is the computational fluid dynamic (CFD) model for solving the Navier-Stokes equations. Such model is often desirable for advanced aerospace applications since it makes the fewest assumptions about the characteristics of the flow field and is capable of accurately predicting its complex response. However, the computational resources required to treat the Navier-Stokes equations are significant. In this section, the proposed LOC-Kriging is used to explore the design space of different wing configuration based on limited data points. The response being investigated is lift over drag.

The shape of the lifting surface is parameterized using the Joukowski transform. This is defined in the complex plane and maps the image of a circle passing through $z_1 = 1$ and containing the point $z_2 = -1$ to an airfoil using the following equation.

$$J(z) = z + \frac{1}{z}, \quad z = x + iy$$ (4.17)

The shape of the airfoil is defined using the coordinates of circle’s center coordinate. The x-coordinate of the center defines the thickness and y-coordinate defines the chamber of the airfoil. The computational domain, Fig. 4.22a, extends 20 chord lengths downstream
and 15 chord lengths to the sides and upstream. The size ensures the boundary effects are not felt near the surface of the airfoil. A two-dimensional structured C-Grid domain is used to discretize the domain. A coarse grid of 29000 points is shown in Fig. 4.22b to illustrate the mesh structure around an airfoil with a thickness parameter of -0.1, and chamber of 0.1. As shown in the figure, the mesh is refined near the airfoil to capture the effect of the viscous boundary layer.

Figure 4.22: **Computational Domain and Coarse C-Grid.** (a) Illustration of the computational domain around the airfoil. (b) Structured C-Grid of 29000 points with a highlighted region used to check convergence.

OpenFOAM CFD solver is used to model the steady viscous flow around the airfoil. The flow is modeled as an incompressible fluid with turbulence modeled using the Spalart-Allmaras model. The mesh convergence study is accomplished using three levels of grid cells, coarse (29000), medium (145000), and fine (232000). The flow results between these different grids are compared by plotting the pressure and velocity magnitude on the red line shown in Fig. 23b. As shown in Fig. 4.23, the difference between the results for the medium and fine grid are negligible. Therefore, the fine grid is selected for the rest of the simulations.

The wing’s structure is modeled as a cantilever beam of 1 meter with square cross section (0.04 m × 0.04 m). The modulus of elasticity of the beam is selected as 200GPa. A simple visual representation of the model is shown in Fig. 4.24. The model is similar to
the Jointed Wing mounted airfoil presented by Liu and Canfield [42]. The only difference is the lack of the strut, causing this model to be interpreted as a traditional wing.

Figure 4.23: CFD Convergence Histories.

Figure 4.24: Sting Mounted Airfoil. Joukowski airfoil mounted to an Euler-Bernoulli Beam.

Both a stationary and LOC-Kriging model are fitted to the five-dimensional design.
space using 161 points, Gaussian correlation functions, and linear global/local regressions. The design parameters and their corresponding ranges are listed in Table 4.7.

Table 4.7: **Design Parameters and Ranges.** Parameters and ranges were selected to represent a low speed aircraft.

<table>
<thead>
<tr>
<th>Design Parameters</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thickness</td>
<td>$X_0$ -0.3 -0.01</td>
</tr>
<tr>
<td>Chamber</td>
<td>$Y_0$ 0 0.2667</td>
</tr>
<tr>
<td>Velocity</td>
<td>$U_\infty$ 40m/s 70m/s</td>
</tr>
<tr>
<td>Angle of Attack</td>
<td>AoA 1° 7°</td>
</tr>
<tr>
<td>Mounting Point</td>
<td>$X_m$ -0.25m 0m</td>
</tr>
</tbody>
</table>

The thickness and chamber airfoil shape parameters were determined to have the most significant effect with respect to the lift over drag response. Therefore, 23 points are evaluated adaptively between thickness and chamber. This process is repeated by varying the other three design parameters within the high and low ranges. This approximation will act as an early stage model exploration. In Fig. 4.25, the true Lift/Drag response is shown as a function of two variables, thickness ($X_0$) and chamber ($Y_0$), while holding: $U_\infty = 50$ meters per second, AoA = 1 degree, and $X_m = -0.125$ meters. Figure 4.25b demonstrates how much the response varies by changing the angle of attack to 7 degrees and the red highlighted cross section is shown in detail in Fig. 4.26.

Figure 4.25: **FSI Two-Dimensional True Response Surfaces.** (a) and (b) true meshed surface of the FSI evaluation points.
The performance results of the two methods are found in Tables 4.8 and 4.9. Due to the high computational cost of the FSI problem the RMSE was evaluated using 540 additional model evaluations and the prediction time is also based on 540 sites. Again, the LOC-Kriging methodology is more computationally expensive, taking 0.37 seconds to build six parallel models when stationary Kriging takes 0.30 seconds to build one model. LOC-Kriging’s model build time includes 35 model builds to find a converged solution in 10.02 seconds. With this additional cost, the LOC-Kriging method outperforms the stationary method with respect to all three metrics. The LOC-Kriging method also produces fewer fluctuations due to its non-stationary covariance structure, shown in Fig. 4.26. Figure 4.26 is a one-dimensional slice of the design domain where: thickness is -0.1711, velocity is 50 meters per second, angle of attack is 7 degrees, mounting point is negative 0.25 meters, and chamber is varied from 0 to 0.2667 in the normalized space. The dots are the discrete RMSE evaluation sites where the diamonds are two samples of the 161 samples used to construct the approximations. The reduction in fluctuations from the stationary Kriging model to the LOC-Kriging’s predicted response is important when considering an optimization algorithm. Gradient-based methods have a difficult time coping with large fluctuations and will likely select one of the valleys as a local optimum. This means LOC-Kriging will not only provided the user a more accurate surface, but a surface more favorable for optimization.

Table 4.8: FSI Performance Comparison. Root-mean-square error, maximum standard deviation, and integral of the MSE performance measures between Stationary Kriging and LOC-Kriging.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Stationary Kriging</th>
<th>LOC-Kriging</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE</td>
<td>0.7938</td>
<td>0.6678</td>
<td>15.88%</td>
</tr>
<tr>
<td>Max σ</td>
<td>5.5750</td>
<td>4.9815</td>
<td>10.64%</td>
</tr>
<tr>
<td>Integral MSE</td>
<td>18.3185</td>
<td>10.1291</td>
<td>44.70%</td>
</tr>
</tbody>
</table>
Table 4.9: **FSI Computational Times.** Reported wall times in seconds between Stationary Kriging and LOC-Kriging.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Stationary Kriging</th>
<th>LOC-Kriging</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Model Build Time (s)</td>
<td>0.3041</td>
<td>0.3700</td>
</tr>
<tr>
<td>Optimal Model Build Time (s)</td>
<td>–</td>
<td>10.0156</td>
</tr>
<tr>
<td>Prediction Time (s)</td>
<td>0.0139</td>
<td>0.3356</td>
</tr>
</tbody>
</table>

Figure 4.26: **FSI One-Dimensional Slice.** Thickness, velocity, angle of attack, and mounting point are all held constant as chamber is varied between its max and min in the normalized space.

### 4.4 Summary

In summary, the proposed locally optimized covariance Kriging method was shown to be more accurate than stationary Kriging thorough the use of four examples: a simple one-dimensional analytical expression, page 45; a complex two-dimensional analytical expression, page 45; a analytical material life model, page 50 and an explicit fluid structural interaction problem, page 57. In all of the examples LOC-Kriging also demonstrated its ability to proved an efficient and flexible computational framework capable of capturing transitional system behaviors. In the analytical material life model example LOC-Kriging
demonstrated additional flexibility by imposing physical understanding in the surrogate construction process.
Summary and Future Work

To address the non-stationary covariance structure with unevenly distributed adaptive samples, the locally optimized covariance Kriging (LOC-Kriging) method is proposed. The proposed method begins with the non-stationary identification process to capture potential localities of an underlying function behavior. With the localities identified, an optimization problem is formulated to determine the optimal sizes of individual local windows by considering the balance between local stationarity and global trend behavior representations. Multiple Krigings with locally optimized covariance structures are aggregated by using membership-weighting function to capture non-stationary behaviors over the entire design domain.

The proposed LOC method demonstrates several potential benefits. First, segmentation of the design space through the uses of membership sets is equivalent to a divide-and-conquer strategy, which can take advantage of parallel computing; therefore, reducing the computational cost significantly. Second, using multiple covariance structures, LOC-Kriging effectively removes the unnecessary amplified modeling errors and provides more meaningful MSE bounds that can be useful in a follow-up process, such as the estimation of expected improvement for an adaptive sampling or a design exploration. Third, physical understanding of the system can be implemented by selecting a locally representative regression and relaxing the stochastic process with a weighting function to preserve interpolation qualities.

In this study, it was assumed an initial DOE was constructed, in future work, the ob-
jective will be to explore an adaptive sampling method integrated with LOC-Kriging to capture the nature of the underlying model in an efficient and accurate manner. This work also was limited to the Universal Kriging and Ordinary Kriging models; however, other advanced Kriging models, such as Dynamic Kriging [11] and Stochastic Kriging [12] can be adopted to obtain furthermore performance improvement for computer simulation and physical experiments respectively. Dynamic-Kriging specifically would allow for optimal selection between discrete basis functions. In future studies, the selection of a covariance structure will be addressed by posing an optimization problem with multiple candidate covariance structures [6] including Wendland, Gaussian, Matern, etc., within local windows, further improving the accuracy. With the flexible computational framework of the proposed LOC-Kriging, model uncertainty due to lack of samples and statistical randomness in samples can be addressed separately for a new effective design exploration process, which is a subject of current research.

The current status of LOC-Kriging representing model uncertainty leverages the Regression Kriging formulation. However, instead of adding the regularization constant to the diagonals of the correlation matrix as in Eq. 2.25, the correlation matrix is scaled by the constant. This alteration allows the MSE of modified Regression Kriging model to represent a constant three standard deviation confidence bound, Fig. 5.1.

![Figure 5.1: Modified Regression Kriging Response.](image)

(a) 20 Sample Points  
(b) 200 Sample Points
The plots above are constructed using a nonlinear mean response and a constant normal standard deviation of 0.05. The green fill represents the true $\pm 3\sigma$ error bounds and the gray fill represents the variation predicted by Kriging. With this in mind, it is evident that as more sample points are added the approximation of both the mean and standard deviation are improved. In the future, this methodology will be coupled with LOC-Kriging to create a framework capable of predicting both a non-stationary mean and variation. This will enable swift identification of key Validation and Verification factors. Once coupled with an adaptive sampling technique, this methodology will be capable of maximizing the confidence of improvement for additional samples.


