

The University of Maine

DigitalCommons@UMaine

Electronic Theses and Dissertations

Fogler Library

Winter 12-2023

Surrogate Modeling With Multi-fidelity Data Sets

Yasaman Asiaee

University of Maine, yasaman.asiaee@maine.edu

Follow this and additional works at: <https://digitalcommons.library.umaine.edu/etd>



Part of the [Computer-Aided Engineering and Design Commons](#)

Recommended Citation

Asiaee, Yasaman, "Surrogate Modeling With Multi-fidelity Data Sets" (2023). *Electronic Theses and Dissertations*. 3922.

<https://digitalcommons.library.umaine.edu/etd/3922>

This Open-Access Thesis is brought to you for free and open access by DigitalCommons@UMaine. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of DigitalCommons@UMaine. For more information, please contact um.library.technical.services@maine.edu.

SURROGATE MODELING WITH MULTI-FIDELITY DATA SETS

By

Yasaman Asiaee

B.S. K. N. Toosi University of Technology, 2020

A THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

(In Mechanical Engineering)

The Graduate School

The University of Maine

December 2023

Advisory Committee:

Masoud Rais-Rohani, Richard C. Hill Professor & Chair of Mechanical Engineering, University of Maine, Advisor

Sheila Edalatpour, Libra Associate Professor of Mechanical Engineering, University of Maine

Brett Ellis, Associate Professor, Mechanical Engineering Technology & Program Coordinator, University of Maine

© 2023 Yasaman Asiaee
All Rights Reserved

SURROGATE MODELING WITH MULTI-FIDELITY DATA SETS

By Yasaman Asiaee

Thesis Advisor: Prof. Masoud Rais-Rohani

An Abstract of the Thesis
Presented
in Partial Fulfillment of the
Requirements for the Degree of
Master of Science
(In Mechanical Engineering)
December 2023

In the realm of science and engineering, there is a need for models that can accurately describe complex systems. High-fidelity models are often used for this purpose, but they tend to be costly to create and impractical to use for many applications including design optimization. On the other hand, low-fidelity models offer a more affordable alternative, but sacrifice accuracy. To strike a balance between accuracy and cost, multifidelity surrogate modeling (MFSM) has emerged as a viable solution by integrating data from sources with different levels of fidelity. This research aims to contribute to the understanding of MFSM by exploring the previous research in this field and examining the performance of two existing MFSM techniques, namely Co-Kriging and Linear Regression Multifidelity Surrogate (LR-MFS) in terms of efficiency and accuracy. Based on the recent studies, these two approaches are the ones most widely used in the MFSM domain. Co-Kriging is primarily focused on interpolation, while the LR-MFS leans towards regression analysis. As a result, Co-Kriging excels at capturing local variations accurately, whereas the LR-MFS is better suited for capturing broader global trends. Consequently, it is logical to explore the possibility of merging the strengths of both Co-Kriging and the LR-MFS to create a more efficient surrogate modeling technique. A novel approach, Co-Kriging-LR-MFS Ensemble Model based on K-means clustering method is introduced to exploit the advantages offered by both high- and low-fidelity models while detecting performance discrepancies in each region based on error analysis. The primary objective of this investigation is to advance the understanding of the benefits and limitations associated with multifidelity surrogate modeling for scientific inquiry while seeking ways to improve its overall performance in engineering. At the end, to evaluate the performance

of the Co-Kriging-LR-MFS Ensemble Model, two test cases involving the Laplace's equation and Peaks function are implemented. The results show that in both cases, the total Root Mean Square Error (RMSE) in the Co-Kriging-LR-MFS ensemble surrogate modeling approach improves by 26.27% and 27.88% as compared to the best stand-alone predictive model, respectively.

ACKNOWLEDGEMENTS

I want to express my deepest appreciation to Prof. Masoud Rais-Rohani, my dedicated advisor, whose steadfast support and invaluable guidance have been instrumental throughout my master's degree journey. Prof. Masoud Rais-Rohani's expertise, unwavering dedication, and constructive feedback have played a pivotal role in shaping the trajectory of my research and academic pursuits. I am truly grateful for the mentorship, patience, and encouragement that have not only enriched my academic experience but have also significantly contributed to my personal and professional growth. His commitment to excellence and passion for knowledge has been a constant source of inspiration, and I consider myself fortunate to have had the privilege of working under his guidance. The rewarding nature of my master's journey is a testament to Prof. Masoud Rais-Rohani's mentorship and support.

I extend my sincere thanks to prof. Sheila Edalatpour and prof. Brett Ellis for graciously agreeing to be part of my committee. Their insights and contributions are highly valued and appreciated.

In addition, I want to convey my heartfelt gratitude to my parents, Alireza and Maryam, and my sister, Elahe, for their unwavering support, which proved indispensable to my success on this educational odyssey. Finally, a special thank you to my beloved husband, Mehran, for being my rock, a source of endless encouragement, and the pillar of strength that not only made this academic endeavor possible but also incredibly rewarding.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	viii
INTRODUCTION	1
1.1. Surrogate modeling.....	1
1.2. Multifidelity models.....	1
1.2.1. Correction techniques.....	2
1.2.1.1. Additive correction	2
1.2.1.2. Multiplicative correction.....	2
1.2.2.3. Comprehensive correction.....	2
1.3. High-fidelity and low-fidelity data sets	3
1.4. Deterministic methods vs. non-deterministic methods.....	4
1.5. Basis function regression and Kriging methods	4
1.5.1. Basis function regression.....	6
1.5.2. Kriging and Co-Kriging.....	7
1.6. Thesis outline	8
SURROGATE MODELING METHODS	10
2.1. Kriging method	10
2.2. Co-Kriging method	14
2.4. Linear regression multifidelity surrogate.....	18

2.5. Co-Kriging-LR-MFS ensemble model	19
2.5.1. Co-Kriging and LR-MFS comparison	20
2.5.1.1. Data variability	20
2.5.1.2. Linear relationships	20
2.5.1.3. Nonlinear effects.....	20
2.5.1.4. Data density	20
2.5.1.5. Model assumptions	21
2.5.1.6. Noise and uncertainty	21
2.5.2. Co-Kriging-LR-MFS algorithm	21
2.5.3. K-means clustering.....	22
2.5.3.1. Centroids selection	22
2.5.3.2. Clustering based on distance	23
2.5.3.3. Updating centroids based on mean data of each cluster	23
2.5.3.4. Repeating assignment and update.....	23
2.6. Predictive model evaluation.....	23
RESULTS AND DISCUSSION OF EXAMPLE PROBLEMS	25
3.1. Example problem one: Forrester function	25
3.1.1. Co-Kriging model	25
3.1.2. LR-MFS model	26
3.2. Example problem two: Laplace's equation.....	27
3.2.1. High and low fidelity data source	27
3.2.1.1. High fidelity: Analytical solution of 2D Laplace's equation.....	27
3.2.1.2. Low fidelity: Numerical solution of 2D Laplace's equation	30
3.2.2. Co-Kriging model.....	32
3.2.3. LR-MFS model.....	32
3.2.4. Co-Kriging-LR-MFS ensemble methods	33
3.3. Example problem three: Peaks function	35
3.3.1. High and low fidelity data source	35

3.3.2. Co-Kriging methods	36
3.3.3. LR-MFS methods	36
3.3.4. Co-Kriging-linear ensemble methods.....	36
3.4. Data analytics.....	39
3.4.1. One dimensional data analysis	39
3.4.2. 2D-dimensional Data analysis.....	40
3.4.2.1. Data analysis for Kriging algorithm	40
3.4.2.2. Data analysis for LR-MFS algorithm	41
3.4.2.3. Co-Kriging-LR-MFS Ensemble modeling in optimum data points	42
CONCLUSION.....	44
APPENDICES	46
Appendix A: Stochastic process explanation.....	46
A.1. Discrete-Time Stochastic Process	46
A.2. Continuous-Time Stochastic Process	46
A.3. Stochastic process properties.....	46
A.3.1. State Space.....	46
A.3.2. Trajectories or Paths	46
A.3.3. Markov Property	47
A.3.4. Stationarity.....	47
Appendix B: Spatial Correlation explanation.....	47
B.1. Three forms of Spatial Correlation	47
B.1.1. Positive Spatial Correlation	47
B.1.2. Negative Spatial Correlation.....	47
B.1.3. No Spatial Correlation	48
B.2. Spatial Correlation importance.....	48
B.2.1. Predictive model improvement	48
B.2.2. Environmental Applications.....	48
B.2.3. Kriging Algorithm.....	48

REFERENCES	49
BIOGRAPHY OF THE AUTHOR	52

LIST OF FIGURES

Figure. 1.1	Popularity of different Multi-Fidelity Models, the chart based on 157 papers.....	5
Figure. 1.2	Summary of Techniques Investigated in This Research.....	9
Figure. 2.1	(a) Impact of changing p on Gaussian distribution. (b) Impact of changing θ on Gaussian distribution	12
Figure. 3.1	Kriging and Co-Kriging predictions on Forrester function's equations using (a) generated codes and (b) Forrester et al.....	27
Figure. 3.2	LR-MFS on Forrester function's equations.....	28
Figure. 3.3	The solution domain and boundary conditions of Laplace equation.....	30
Figure. 3.4	The solution domain discretization of Laplace equation.....	31
Figure. 3.5	(a) Analytical solution of Laplace's equation in whole domain. (b) (b) Semi-converged numerical solution of Laplace's equation in (c) whole domain.....	32
Figure. 3.6	(a) Analytical solution of Laplace equation (high-fidelity data source). (b) Semi-converged numerical solution of Laplace equation (low-fidelity data source).....	33
Figure. 3.7	Sixteen different regions for Laplace case study generated by K-means clustering where the magenta triangles represent the regions centers.....	34
Figure. 3.8	(a) Analytical solution of Laplace equation. (b) Co-Kriging- LR-MFS (b) model prediction. (c) Co-Kriging model prediction. (c) (d) LR-MFS prediction.....	36
Figure. 3.9	Peaks function (a) high-fidelity data source, (b) low- fidelity data source.....	37
Figure. 3.10	Peaks function prediction by (a) Co-Kriging model and (b) LR-MFS model. The yellow star marks the region where Co-Kriging outperforms LR-MFS model. The blue star marks the region where LR-MFS model is better than Co-Kriging model.....	38
Figure. 3.11	Nine different regions for peaks function case study generated by	

	K-means clustering where the magenta triangles represent the regions centers.....	39
Figure. 3.12	(a) Peaks function. (b) Co-Kriging-LR-MFS model prediction. (c) (b) Co-Kriging model prediction. (d)LR-MFS (c) model prediction.....	40
Figure. 3.13	(a) Variation of RMSE with Number of Data Points for Co-Kriging (b) Algorithm. (b) Predictive model using 70 data points	42
Figure. 3.14	(a) Variation of RMSE with Number of Data Points for LR-MFS (b) Algorithm. (b) Predictive model using 40 data points.....	43
Figure. 3.15	(a) Peaks function. (b) Co-Kriging-LR-MFS model prediction. (c) (b) Co-Kriging model prediction. (d) LR-MFS model prediction.....	44

CHAPTER 1

INTRODUCTION

1.1. Surrogate modeling

Within the modeling domain, a surrogate model becomes a highly valuable asset. It serves as an intermediate point between intricate and complex models by offering a simplified alternative. The purpose behind this is to replicate the behavior or predictions of the original model without incurring excessive computational costs or complexity. This thesis explores various methods for approximating functions that yield real-valued outputs. Specifically, it focuses on functions denoted as $z(\mathbf{x})$ within a defined input parameter space $Q \subset R$. These particular functions are commonly known as "Blackbox functions" in computer experiments, representing the output of computer code when provided with certain input parameters \mathbf{x} . To approximate the relationship between the input variables \mathbf{x} and the response variable $z(\mathbf{x})$, we rely on an experimental design set of n samples, i.e., $D = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}\}$ along with their corresponding outputs $\mathbf{z} = \{z(\mathbf{x}^{(1)}), z(\mathbf{x}^{(2)}), \dots, z(\mathbf{x}^{(n)})\}^T$.

However, this data set alone is insufficient for constructing a reliable surrogate model for $z(\mathbf{x})$, and it is necessary to make assumptions about the nature of the function itself. An important consideration in this process is determining the required number of observations, n . Given the increasing complexity of computer simulators, conducting a large number of simulations for tasks such as uncertainty quantification, sensitivity analysis or optimization can be impractical. Thus, the objective of this research is to develop a fast approximation method – often referred to as a *surrogate model* or *metamodel* – that utilizes only a limited number of observations. Statistical methods play a crucial role in analyzing computer experiments due to multiple sources of uncertainty involved in the process.

1.2. Multifidelity models

The concept of fidelity in modeling pertains to the degree of accuracy offered by a predictive model or simulation. Typically, high-fidelity (HF) models yield accurate predictions; however, they often require substantial computational resources. In contrast, low-fidelity (LF) models are computationally cheap, but they often lack the accuracy of HF models. Multifidelity models (MFs) combine HF and LF models to achieve rapid yet reliable predictions. These models have gained

significant traction due to their capacity to generate approximate models with remarkable accuracy while minimizing computational cost. To take advantage of both HF and LF models, a number of multifidelity surrogate models (MFSMs) based on different approximation functions (e.g., Kriging, polynomial response surface, and radial basis function) have been developed.

1.2.1. Correction techniques

In MFSMs, three primary correction techniques are employed: additive, multiplicative, and comprehensive.

1.2.1.1. Additive correction

Additive correction captures the difference between HF and LF models such that the MFSM of the HF model is expressed as

$$y_{MF}(\mathbf{x}) \approx \hat{y}_{HF}(\mathbf{x}) = y_{LF}(\mathbf{x}) + \delta(\mathbf{x}) \quad (1.1)$$

where $\hat{y}_{HF}(\mathbf{x})$ represents the multifidelity approximation of the HF model, y_{LF} represents the LF model, and $\delta(\mathbf{x})$ represents an additive correction or discrepancy function to address variations between HF and LF models. In cases where computational cost or another limitation is a hindrance, LF model can be swapped with its approximate model.

1.2.1.2. Multiplicative correction

In an alternative approach, the MFSM can be mathematically represented by incorporating a multiplicative correction factor as

$$y_{MF}(\mathbf{x}) \approx \hat{y}_{HF}(\mathbf{x}) = \rho(\mathbf{x})y_{LF}(\mathbf{x}) \quad (1.2)$$

where $\rho(\mathbf{x})$ is an approximation function incorporating the ratio of the HF and LF models at each common location given by vector \mathbf{x} . It is worth noting that if the LF model proves to be too costly, it can be replaced by an approximate model.

1.2.2.3. Comprehensive correction

Within the context of MFSM, a comprehensive correction approach involves utilizing both additive and multiplicative adjustments within a unified framework expressed as

$$y_{MF}(\mathbf{x}) \approx \hat{y}_{HF}(\mathbf{x}) = \rho(\mathbf{x})y_{LF}(\mathbf{x}) + \delta(\mathbf{x}) \quad (1.3)$$

where the function $\rho(\mathbf{x})$ serves as a representation of the multiplicative correction function while $\delta(\mathbf{x})$ is used to represent the additive correction approximation function. An examination of previous scholarly works highlights that the preferred approach typically involves maintaining a constant value for $\rho(\mathbf{x})$ and using a surrogate model to estimate the additive correction $\delta(\mathbf{x})$.

1.3. High-fidelity and low-fidelity data sets

In the context of modeling and simulation, high- and low-fidelity data sets refer to different levels of detail or accuracy in the representation of a system. Whether a model is considered to have high or low fidelity depends on the comparison made. For instance, a full 3D simulation might be expensive as compared to a primary function calculation in a 1D simulation, but inexpensive compared to actual experiments. Additionally, various sources of data may be used to generate both high-fidelity and low-fidelity data sets, and this can vary from case to case. These sources are often classified based on various factors, including the following:

Dimensionality (3D vs. 2D):

- High-fidelity data in three-dimensional space provides a more detailed representation of the system, capturing complexities in all directions.
- Low-fidelity data in two-dimensional space may simplify the representation, neglecting variations in the third dimension for computational efficiency.

Analysis Resolution (Refined vs. Coarse):

- High-fidelity data with a refined analysis resolution contains more data points per unit dimension, offering a detailed view of the system.
- Low-fidelity data with a coarse analysis resolution reduces the number of data points per unit dimension, providing a less generalized or simplified representation.

Type of Study (Simulations vs. Experiments):

- High-fidelity data from simulations involve detailed mathematical models and algorithms that simulate the behavior of the system under various conditions.

- Low-fidelity data from experiments involve real-world observations, which may have limitations in terms of precision and control compared to simulations.

State of Flow (Transient vs. Steady):

- High-fidelity data may include information about transient states, capturing changes in the system over time.
- Low-fidelity data may focus on steady-state conditions, simplifying the analysis by assuming that the system has reached a stable condition.

Degree of Solution Convergence (Converged vs. Semi-converged):

- High-fidelity data with a fully converged solution implies that the computational model has reached a stable and accurate solution.
- Low-fidelity data with a semi-converged solution may represent a less accurate or stable state, possibly due to computational constraints or simplifications.

1.4. Deterministic methods vs. non-deterministic methods

Deterministic models assume that the parameters of the surrogate model are known with certainty. They use predefined basis functions and aim to minimize the differences between the surrogate model's predictions and actual data. They are a good choice when one is confident about the model's form and parameters. In contrast, non-deterministic models capture uncertainty. They can either assume uncertainty in the underlying function itself or in the parameters of the surrogate model. They use data or samples to estimate the model or its uncertain parameters. Non-deterministic models are preferred when dealing with uncertain or complex modeling situations where flexibility and higher accuracy are crucial. The choice between the two category of methods depends on the specific problem and the level of uncertainty present in the system.

1.5. Basis function regression and Kriging methods

In the vast landscape of multifidelity surrogate modeling, where researchers navigate through various techniques to model complex systems, two methods have captured widespread attention: “Kriging” and “Basic function regression,” with the latter referring to a broader class of regression techniques where the relationship between variables is modeled using basic functions or basis functions, which are functions used to represent the relationship between variables. These

functions can be linear, quadratic, cubic polynomials, or even more complex. Through an initial exploration of existing literature, it became evident that these two approaches serve as the backbone of multifidelity surrogate modeling, finding extensive use across different fields as shown in Figure 1.1. To delve deeper into their significance, this literature review specifically focuses on Kriging and Basis function regression approaches to unravel their unique strengths and limitations.

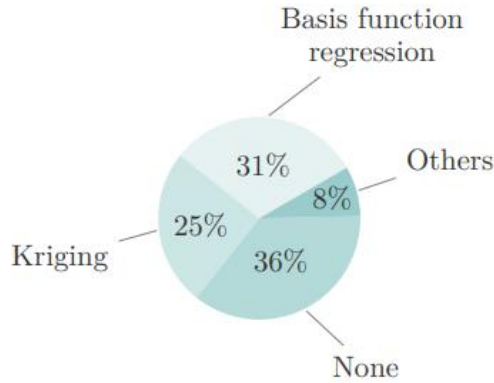


Figure 1.1 Popularity of different multifidelity models based on 157 papers surveyed by Fernández-Godino et al. [2023]

Some multifidelity models (MFs) can be created without the need for an MFSM. These types of MFs are referred to as multifidelity hierarchical models, and they fall into the category labeled as “none” in Figure 1.1. For instance, in the work of Choi et al. [2008], a hierarchical MFM approach is proposed for optimization. In this approach, HF models are selectively used to correct the deficiencies of LF models without the explicit construction of an MFs or the integration of multiple fidelities. This means that they are using different fidelity models strategically without creating an overarching model that combines them. For example, in the work of Kalivarapu and Winer [2008], an MF is used for interactive modeling of advective and diffusive contaminant transport without the construction of an MFSM. Giunta et al. [1995] and Zahir et al. [2013] are cited as other examples where this approach is followed, avoiding the need for an MFSM. These cases show that there are various ways to utilize multiple fidelity levels in modeling, and not all of them require the creation of an explicit overarching surrogate model. The group labeled as “others” comprises artificial neural networks, moving least squares, support vector machines, radial basis interpolation, and proper orthogonal decomposition, all of which have a usage rate below 1%.

1.5.1. Basis function regression

In engineering applications, single-fidelity surrogate models have long been regarded as the conventional choice. These models, such as polynomial response surface (PRS) and radial basis functions (RBF), have gained popularity over the years Myers et al. [2016], Gutmann [2001], Sun et al. [2011]. Interest in response surface-based MFSMs has been growing significantly. In an effort to reduce prediction errors of HF samples, Zhang et al. [2018] introduced a revised heuristic MFSM called linear regression multifidelity surrogate (LR-MFS), which is based on linear regression, focused on minimizing prediction errors in high-fidelity samples through optimization techniques. The approach involves approximating system behavior at a high-fidelity level by employing a linear combination of low-fidelity predictions alongside a polynomial-based discrepancy function. This algorithm involves integrating the low-fidelity model as a fundamental basis function within a multifidelity model, utilizing the scale factor as a regression coefficient. This integration forms the design matrix for least-square estimation, comprising both the low-fidelity model and the discrepancy function. The process further involves simultaneous extraction of the scale factor and coefficients of the basis functions through linear regression, ensuring a unique fitting process. Notably, the approach, termed linear regression multifidelity surrogate (LR-MFS), not only facilitates efficient parameter estimation but also holds promise for broader applicability in other regression models. The adaptability of LR-MFS lies in its capacity for straightforward substitution of the design matrix. Song et al. [2019] proposed another MFSM that utilizes polynomial regression and allows for the adjustment of judgment factors within different design spaces to enhance prediction accuracy. This work introduces an innovative approach aimed at reducing the number of evaluations required for the high-fidelity model within the optimization process. The primary objectives include accelerating optimization speed while enhancing the accuracy of the optimal solution. The proposed method focuses on a robust and computationally efficient multi-fidelity local surrogate-model optimization strategy. Leveraging the principles of response surface approximation, this method capitalizes on multifidelity coarse models and employs polynomial interpolation to create a series of localized surrogate models. Throughout the optimization process, the methodology involves iterative modeling and optimization within local regions. A critical aspect involves the introduction of a judgment factor, providing crucial information for updating the size of the local regions. An iterative refinement process is executed, culminating in the enhancement of the final local surrogate model using space mapping techniques.

This refined model facilitates the attainment of an optimal design with notably high accuracy. Rumpfkeil et al. [2019] presented an MFSM that employs the sparse polynomial chaos expansion approach as a cost-effective alternative to computationally intensive engineering analyses. Additionally, Cheng et al. [2019] developed an MFSM by combining sparse polynomial chaos expansion with recursive Gaussian process, enabling accurate response prediction for dynamic systems.

1.5.2. Kriging and Co-Kriging

Kriging, a powerful geostatistical interpolation technique, enables the estimation of unobserved spatial data points by considering the spatial correlation between known points. This method, first introduced by Matheron [1963] and further developed by Cressie [1988], relies on the assumption of spatial dependence to generate predictions and quantify uncertainties.

Kennedy and O'Hagan [2000] introduced a novel MFS model by employing a Bayesian approach along with Gaussian process techniques. Building upon this work, Forrester et al. [2007] expanded the widely-used Kriging model to create a two-level variant called the Co-Kriging model. This was achieved through the construction of a correlation matrix that incorporates both high-fidelity and low-fidelity information. The Kriging and Co-Kriging algorithms will be discussed in Chapter 2 based on this paper. Furthermore, Xiao et al. [2018] utilized orthogonal decomposition methods to generate multilevel multifidelity datasets and further extended the framework of Forrester et al. [2007] by introducing multiple levels into the Co-Kriging model.

The exploration of MFMs in this thesis has predominantly focused on two fidelities. However, it is crucial to recognize that a number of studies, including Huang et al. [2006], Forrester et al. [2007], Qian et al. [2008], Le Gratiet [2013], and Goh et al. [2013], have successfully shown the possibility of constructing MFMs with more than just two fidelities. In addition to the various adaptations and modifications of the Co-Kriging model, its applications in different industries such as aerospace, automobile, marine, and intelligent manufacturing have garnered significant attention. These recent developments have showcased the potential benefits of utilizing Co-Kriging models in these fields. For instance, a study conducted by Yong et al. [2019] demonstrated the successful implementation of the Co-Kriging model in optimizing the geometry of gas turbines. This approach resulted in substantial cost savings of 43% compared to using a single HF Kriging method. Similarly, Shi et al. [2020] applied the Co-Kriging model in designing all-electric

geostationary-orbit satellite systems. The intricate task of designing all-electric geostationary orbit (GEO) satellite systems presents a complex challenge in multidisciplinary optimization (MDO), heavily reliant on resource-intensive simulations. That study delves into the MDO issue of these satellites, employing multifidelity models. It formulates an MDO problem involving six interconnected disciplines, aiming to minimize the overall satellite system mass while adhering to various engineering limitations. To alleviate computational burdens in the multidisciplinary analysis (MDA) process, they relied on transfer dynamics models of varying fidelity and finite element analysis (FEA) models for GEO and structural disciplines, respectively. Their solution method using multi-fidelity models and an adaptive Co-Kriging-based optimization framework integrates data from high- and low-fidelity MDA processes, constructing a moderately computationally efficient Co-Kriging metamodel for optimization purposes. Additionally, an approach for refining Co-Kriging metamodels through a multi-objective adaptive infill sampling method is introduced. This method generates infill sample points based on expected improvement and probability of feasibility functions. Results from optimization trials demonstrated the framework's capability to significantly reduce total satellite system mass within a constrained computational budget. This highlights the effectiveness and applicability of employing multifidelity modeling and adaptive Co-Kriging-based optimization for the design of all-electric GEO satellite systems. The utilization of this model led to a notable reduction in total system mass while working within limited computational resources. These examples highlight how Co-Kriging models can be effectively employed to enhance efficiency and achieve favorable outcomes across various sectors. As researchers continue to explore new possibilities for applying this methodology, we can expect further advancements and improvements in diverse industrial settings.

1.6. Thesis outline

The general outline of the research conducted is shown in Figure 1.2. Chapter 2 delves into the intricate details of Kriging, Co-Kriging, and LR-MFS, elucidating their mathematical formulations step by step. Additionally, this research introduces a novel approach named "Co-Kriging-LR-MFS Ensemble Modeling" that integrates *K-means* Clustering methodology, combining the strengths of these two established methods to achieve enhanced accuracy.

Moving to Chapter 3, several case studies involving one and two-dimensional functions such as the Laplace's equation and the Peaks function are used as example problems. The outcomes of

different approaches are used to inform the subsequent introduction of the ensemble model, showcasing its results. Comparative analysis based on Root Mean Square Error (RMSE) highlight the ensemble approach's superiority, consistently yielding lower RMSE values. This pattern underscored the consistent enhancement in overall accuracy by leveraging the combined strength of multiple models over a single model.

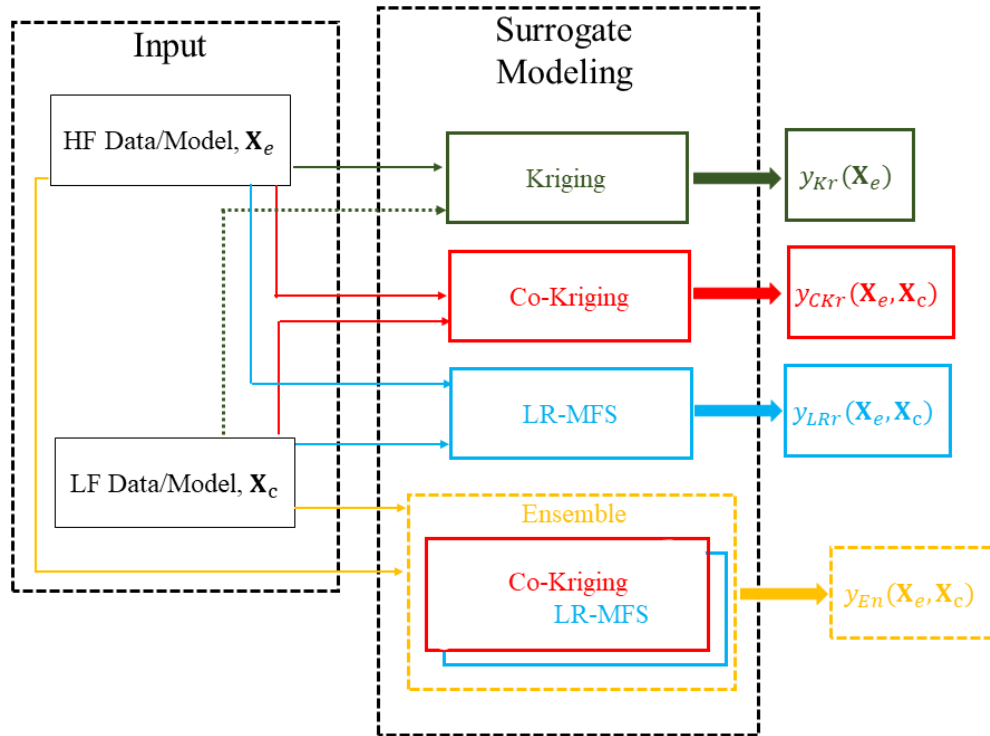


Figure 1.2 Summary of techniques investigated in this research.

CHAPTER 2

SURROGATE MODELING METHODS

2.1. Kriging

In Kriging, the objective is to model a response function as a stochastic process as noted in Appendix A. This means that given any set of input data points, the model produces a distribution encompassing potential functions that could accurately represent the response values at the data points.

Regarding spatial data analysis, Kriging is an application of Gaussian process regression where the input points are associated with specific spatial locations. In other words, Kriging can be seen as a special case of Gaussian process regression, specifically applied in the context of spatial interpolation, where the focus is on estimating values at unobserved locations within a spatial domain.

Kriging leverages the covariance structure between data points to make predictions at unobserved locations. It takes into account both observed data and the spatial correlation existing among these data points. For more information regarding the spatial correlation, refer to Appendix B. Considering the correlation among the training data points helps avoid duplication of information and prevents the duplication of the same data's effect, ensuring a more efficient and accurate analysis.

Typically, the Kriging correlation is modeled using either a covariance or a variogram function. The utilization of Gaussian processes allows Kriging to not only provide point estimates at unobserved locations but also quantify uncertainties associated with these estimates. Consequently, Kriging is as an influential tool for spatial prediction and interpolation across various domains such as geostatistics, environmental science, and spatial statistics.

The general or universal Kriging model, consisting of two key components, is expressed as

$$y(\mathbf{x}) = f(\mathbf{x}) + Z(\mathbf{x}) \tag{2.1}$$

where $f(\mathbf{x})$ is the global trend function, often replaced by a constant term μ as representing a mean base term, and a stationary Gaussian process, $Z(\mathbf{x})$ that quantifies the localized deviation from the global trend model or the level of uncertainty associated with the model.

The Gaussian process term is assumed to have a zero mean and a covariance matrix given as

$$\text{cov}[Z(\mathbf{x}^{(j)}), Z(\mathbf{x}^{(i)})] = \sigma^2 \psi \quad (2.2)$$

where σ^2 is the variance and $\psi = \text{corr}[y(\mathbf{x}^{(j)}), y(\mathbf{x}^{(i)})]$ is the symmetric correlation matrix (with 1's along the diagonal) that characterizes the correlation between different points in the input space (i.e., the training set). The spatial correlation quantifies how the function values at different input points are correlated, with the correlation function, defining each term in the symmetric correlation matrix, given as

$$\psi_{ij} = \exp \left(- \sum_{k=1}^m \theta_k \left\| \mathbf{x}_k^{(j)} - \mathbf{x}_k^{(i)} \right\|^{p_k} \right) \quad (2.3)$$

where vector $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]^T$, with θ_k ($0 \leq \theta_k$) and p_k ($1 \leq p_k \leq 2$) representing the hyper parameters associated with the k^{th} component of vector \mathbf{x} . The correlation function in Eq. (2.3) is shown in terms of the general exponential kernel function; other common kernel functions are given in Table 2.1.

In Kriging, the choice of hyper parameters within the kernel function can significantly impact the final prediction by influencing the model's flexibility, smoothness, and ability to capture the underlying spatial structure or correlation in the data. These hyper parameters are crucial in determining the behavior of the covariance and the kernel function, which, in turn, affects predictions in Kriging. Common kernel functions include the Radial Basis Function (RBF) or Gaussian kernel, the linear kernel, and others, as illustrated in Table 2.1, where \mathbf{x} and $\tilde{\mathbf{x}}$ represent two arbitrary data points in a multidimensional space.

Table 2.1 Common kernel functions for Kriging [Burges, 1998]

Model	Kernel function	Formula	Optimization parameters
1	Dot product	$K(\mathbf{x}, \tilde{\mathbf{x}}) = (\mathbf{x} \cdot \tilde{\mathbf{x}}) + C$	C
2	RBF or Gaussian	$K(\mathbf{x}, \tilde{\mathbf{x}}) = \exp(-\boldsymbol{\theta} \ \mathbf{x} - \tilde{\mathbf{x}}\ ^2) + C$	$\boldsymbol{\theta}, C$
3	General Exponential	$K(\mathbf{x}, \tilde{\mathbf{x}}) = \exp(-\boldsymbol{\theta} \ \mathbf{x} - \tilde{\mathbf{x}}\ ^p) + C$	$\boldsymbol{\theta}, \mathbf{p}, C$
4	Sigmoid	$K(\mathbf{x}, \tilde{\mathbf{x}}) = \tanh(\gamma(\mathbf{x} \cdot \tilde{\mathbf{x}}) + r) + C$	γ, r, C
5	Polynomial	$K(\mathbf{x}, \tilde{\mathbf{x}}) = (\gamma(\mathbf{x} \cdot \tilde{\mathbf{x}}) + r)^d + C$	γ, r, \mathbf{d}, C

Specifically, in the General Exponential kernel, which is widely used in Kriging, the hyper parameters $\boldsymbol{\theta}$ and \mathbf{p} control the width and smoothness of the function, respectively. As shown in Figure 2.1(a), when adjusting the parameter \mathbf{p} ($0.1 \leq \mathbf{p} \leq 2$) in the general exponential kernel with $\boldsymbol{\theta} = 1$, a larger value results in a smoother distribution and making the curve more spread out. Conversely, a smaller \mathbf{p} value makes the distribution sharper, accentuating the peaks and making the curve more concentrated. On the other hand, varying the parameter $\boldsymbol{\theta}$ ($0.1 \leq \boldsymbol{\theta} \leq 10$) with $\mathbf{p} = 2$ in the kernel function influences the width of the distribution: a larger $\boldsymbol{\theta}$ widens the distribution plot, making it more spread out, while a smaller $\boldsymbol{\theta}$ narrows the plot, as illustrated in Figure 2.1(b).

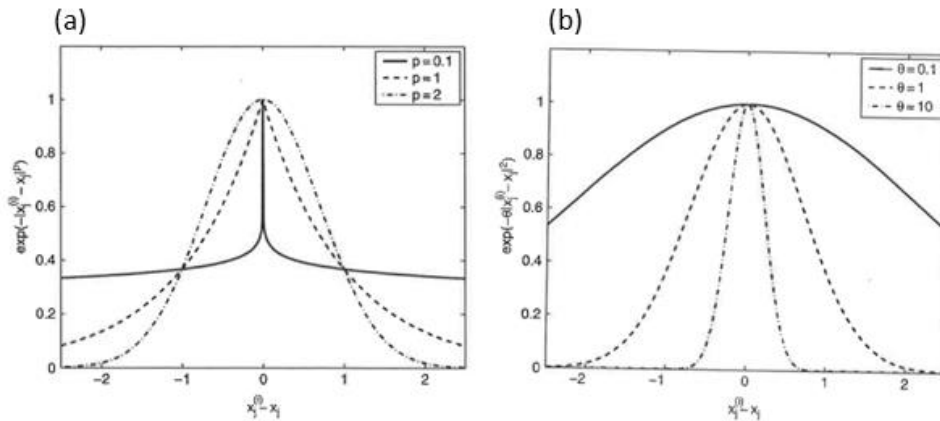


Figure 2.1 Impact of changing the values of hyper parameters (a) \mathbf{p} and (b) $\boldsymbol{\theta}$ on the general exponential kernel function. [Forrester et al., 2008]

In Kriging, the correlation vector, characterizing the correlation between the prediction point $\tilde{\mathbf{x}}$ and each training point $\mathbf{x}^{(i)}$, is given as

$$\psi^{(i)} = \text{corr}[y(\tilde{\mathbf{x}}), y(\mathbf{x}^{(i)})] = \exp \left(- \sum_{k=1}^m \theta_k \left\| \tilde{\mathbf{x}}_k - \mathbf{x}_k^{(i)} \right\|^{p_k} \right) \quad (2.4)$$

where $y(\tilde{\mathbf{x}})$ represents the unknown response at the prediction point $\tilde{\mathbf{x}}$ with its expected value, given the vector of known responses at $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(n)}$, expressed as

$$\hat{y}(\tilde{\mathbf{x}}) = E \left(y(\tilde{\mathbf{x}}) | y(\mathbf{x}^{(1)}), y(\mathbf{x}^{(2)}), \dots, y(\mathbf{x}^{(n)}) \right) \quad (2.5)$$

Introducing the mean squared error (MSE) as

$$MSE = E \left(\hat{y}(\tilde{\mathbf{x}}) - y(\tilde{\mathbf{x}}) \right)^2 \quad (2.6)$$

Minimizing MSE in Eq. (2.6) gives

$$\hat{y}(\tilde{\mathbf{x}}) = \hat{\mu} + \sum_{i=1}^n b^{(i)} \psi^{(i)} \quad (2.7)$$

where $\hat{\mu}$ is mean base term, $b^{(i)}$ is the i^{th} term of the $n \times 1$ vector $\mathbf{b} = \psi^{-1}(\mathbf{y} - \mathbf{F}\hat{\mu})$, ψ is the $n \times n$ correlation matrix with the i, j elements given by Eq. (2.3) in terms of the unknown hyper parameters θ_k and p_k , \mathbf{y} is the $n \times 1$ vector of responses at the training points, $\mathbf{F}(\mathbf{x})$ represents the regression basis function vector associated with the n sampled sites \mathbf{x} , assumed to be represented by an $n \times 1$ vector of 1's in this research, and $\psi^{(i)}$ is the correlation vector given by Eq. (2.4), which also depends on θ_k and p_k . The $\hat{}$ terms in Eq. (2.7) represent the maximum likelihood estimates.

The prevailing approach for determining the unknown hyper parameters in the Kriging surrogate involves maximizing the likelihood function given as

$$L(\mu, \sigma^2) = \frac{1}{\sqrt[n]{2\pi\sigma^2} \sqrt{|\psi|}} \exp \left[- \frac{(\mathbf{y} - \mathbf{F}\mu)^T \psi^{-1} (\mathbf{y} - \mathbf{F}\mu)}{2\sigma^2} \right] \quad (2.8)$$

The likelihood is maximal when its gradient consisting of its partial derivatives with respect to μ and σ^2 are set to zero, which gives

$$\hat{\mu} = (\mathbf{F}^T \psi^{-1} \mathbf{F})^{-1} \mathbf{F}^T \psi^{-1} \mathbf{y} \quad (2.9)$$

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{F}\hat{\mu})^T \psi^{-1} (\mathbf{y} - \mathbf{F}\hat{\mu})}{n} \quad (2.10)$$

In practical applications, it is common to utilize the natural logarithm of the likelihood function given as

$$\ln(L(\hat{\theta}_k, \hat{p}_k)) = -\frac{1}{2} [\ln(\hat{\sigma}^2) + \ln|\psi|] \quad (2.11)$$

The determination of the optimal values for the maximum likelihood estimates of the hyper parameters $\hat{\theta}_k, \hat{p}_k, k = 1, 2, \dots, m$ is achieved by solving a nonlinear optimization problem that aims to maximize the natural log of the likelihood function.

$$\max \ln(L(\hat{\theta}_k, \hat{p}_k)) \quad (2.12)$$

$$\text{s. t.} \quad \hat{\theta}_k > 0, \quad 1 \leq \hat{p}_k \leq 2, \quad k = 1, \dots, m$$

Once the optimal values of the hyper parameters are calculated, the predictive Kriging model is completely defined.

2.2. Co-Kriging

The extension of the Kriging model for systems with different levels of fidelity is called Co-Kriging and was first developed in geostatistics by Chilès and Delfiner [1999]. Also, Kennedy and O'Hagan [2000] proposed a Bayesian approach based on Co-Kriging to combine low- and high-fidelity models. This approach allows for the integration of multiple models of different levels of accuracy or fidelity. One key feature of their Co-Kriging method is the use of an autoregressive model to couple the different fidelity levels. In the context of Co-Kriging and multifidelity modeling, an autoregressive model refers to a statistical model where each fidelity level's output variable is influenced by the output of the previous fidelity level. This means that the lower-fidelity model helps inform and improve the higher-fidelity model, and the two are connected in a way that accounts for their correlations. The autoregressive structure allows for the effective use of

information from both low-fidelity and high-fidelity models, leveraging the strengths of each to make more accurate predictions or estimates.

Here, we explore the idea of merging two independent datasets, one being highly accurate but expensive (denoted as \mathbf{y}_h at points \mathbf{X}_h), while the other set is less accurate yet more affordable (referred to as \mathbf{y}_l at points \mathbf{X}_l). By concatenating these distinct datasets as

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_l \\ \mathbf{X}_h \end{pmatrix} = (\mathbf{x}_l^{(1)}, \dots, \mathbf{x}_l^{(n_l)}, \mathbf{x}_h^{(1)}, \dots, \mathbf{x}_h^{(n_h)})^T \quad (2.13)$$

a comprehensive collection of data can be formed for further analysis and insights.

It is worth mentioning that in the context of Co-Kriging, the low-fidelity values at high-fidelity locations are needed. This can be approached in two ways, either by selecting high fidelity locations, such as $\mathbf{X}_h \subset \mathbf{X}_l$, or by using low fidelity data to create a Kriging estimation and then using the predictive model to obtain the values at high fidelity locations.

The mathematical algorithm used here is that presented by Forrester et al. [2007]. Similar to the Kriging technique, in Co-Kriging, the value at a specific location within the set \mathbf{X} is considered as an observation of a Gaussian random variable. In this model, we assume that $cov[y_h(\mathbf{x}^{(i)}), y_l(\mathbf{x}) | y_l(\mathbf{x}^{(i)})] = 0, \forall \mathbf{x} \neq \mathbf{x}^{(i)}$. In other words, if the value of the high-fidelity function y_h (assumed to be the true function value) at location $\mathbf{x}^{(i)}$ is known, no additional information can be gained about its value from the LF model y_l . By employing the auto-regressive model, the expensive model is approximated as

$$Z_h(\mathbf{x}) = \rho Z_l(\mathbf{x}) + Z_d(\mathbf{x}) \quad (2.14)$$

where ρ is a scaling factor and $Z_d(\mathbf{x})$ is a Gaussian process that captures the difference between $\rho Z_l(\mathbf{x})$ and $Z_h(\mathbf{x})$. Because of using two sets of data, three separate covariance matrices need to be calculated as

$$cov[\mathbf{y}_l(\mathbf{X}_l), \mathbf{y}_l(\mathbf{X}_l)] = cov[Z_l(\mathbf{X}_l), Z_l(\mathbf{X}_l)] = \sigma_l^2 \psi_l(\mathbf{X}_l, \mathbf{X}_l) \quad (2.15)$$

$$cov[\mathbf{y}_h(\mathbf{X}_h), \mathbf{y}_l(\mathbf{X}_l)] = cov[\rho Z_l(\mathbf{X}_l) + Z_d(\mathbf{X}_h), Z_l(\mathbf{X}_h)] = \rho \sigma_l^2 \psi_l(\mathbf{X}_l, \mathbf{X}_h) \quad (2.16)$$

$$\begin{aligned}
\text{cov}[\mathbf{y}_h(\mathbf{X}_h), \mathbf{y}_h(\mathbf{X}_h)] &= \text{cov}[\rho Z_l(\mathbf{X}_h) + Z_d(\mathbf{X}_h), \rho Z_l(\mathbf{X}_h) + Z_d(\mathbf{X}_h)] \\
&= \rho^2 \text{cov}[Z_l(\mathbf{X}_h), Z_l(\mathbf{X}_h)] + \text{cov}[Z_d(\mathbf{X}_h), Z_d(\mathbf{X}_h)] \\
&= \rho^2 \sigma_l^2 \psi_l(\mathbf{X}_h, \mathbf{X}_h) + \sigma_d^2 \psi_d(\mathbf{X}_h, \mathbf{X}_h)
\end{aligned} \tag{2.17}$$

As shown by Forrester et al. [2007], the complete correlation matrix is written as

$$C = \begin{bmatrix} \sigma_l^2 \psi_l(\mathbf{X}_l, \mathbf{X}_l) & \rho \sigma_l^2 \psi_l(\mathbf{X}_l, \mathbf{X}_h) \\ \rho \sigma_l^2 \psi_l(\mathbf{X}_h, \mathbf{X}_l) & \rho^2 \sigma_l^2 \psi_l(\mathbf{X}_h, \mathbf{X}_h) + \sigma_d^2 \psi_d(\mathbf{X}_h, \mathbf{X}_h) \end{bmatrix} \tag{2.18}$$

where the correlation matrix ψ_l is a function of unknown hyper parameters $\boldsymbol{\theta}_l$ and \mathbf{p}_l whereas ψ_d is a function of unknown hyper parameters $\boldsymbol{\theta}_d$ and \mathbf{p}_d with ρ as the last unknown parameter. The same process as in Kriging is followed in Co-Kriging for optimizing the hyper parameters. The natural log of the likelihood equation for low-fidelity data is given as

$$\ln(L) = -\frac{n_l}{2} \ln(\sigma_l^2) - \frac{1}{2} \ln(|\psi_l(\mathbf{X}_l, \mathbf{X}_l)|) - \frac{(\mathbf{y}_l - \mathbf{F}\mu_l)^T \psi_l(\mathbf{X}_l, \mathbf{X}_l)^{-1} (\mathbf{y}_l - \mathbf{F}\mu_l)}{2\sigma_l^2} \tag{2.19}$$

When we take the partial derivatives of $\ln(L)$ with respect to μ_l and σ_l^2 , and set them equal to zero, we obtain the Maximum Likelihood Estimates (MLEs) of these parameters as

$$\hat{\mu}_l = (\mathbf{F}^T \psi_l(\mathbf{X}_l, \mathbf{X}_l)^{-1} \mathbf{F})^{-1} \mathbf{F}^T \psi_l(\mathbf{X}_l, \mathbf{X}_l)^{-1} \mathbf{y}_l \tag{2.20}$$

$$\hat{\sigma}_l^2 = \frac{(\mathbf{y}_l - \mathbf{F}\hat{\mu}_l)^T \psi_l(\mathbf{X}_l, \mathbf{X}_l)^{-1} (\mathbf{y}_l - \mathbf{F}\hat{\mu}_l)}{n_l} \tag{2.21}$$

where n_l is the number of low-fidelity data. In order to construct the Co-Kriging metamodel for the LF model, it is necessary to obtain the MLEs of the hyper parameters $\hat{\boldsymbol{\theta}}_l$ and $\hat{\mathbf{p}}_l$. This can be achieved by maximizing Eq. (2.19).

$$\ln(L(\hat{\boldsymbol{\theta}}_l, \hat{\mathbf{p}}_l)) = -\frac{n_l}{2} \ln(\hat{\sigma}_l^2) - \frac{1}{2} \ln(|\psi_l(\mathbf{X}_l, \mathbf{X}_l)|) \tag{2.22}$$

The hyper parameters ρ , and θ_d and p_d are estimated using MLE in order to build the Co-Kriging surrogate model. This metamodel captures the discrepancy between the high- and low-fidelity models, denoted as

$$\mathbf{d} = \mathbf{y}_h - \rho \mathbf{y}_l(\mathbf{X}_h) \quad (2.23)$$

The natural log of the likelihood equation for HF data is given as

$$\begin{aligned} \ln(L) = & -\frac{n_l}{2} \ln(\sigma_d^2) - \frac{1}{2} \ln(|\psi_d(\mathbf{X}_h, \mathbf{X}_h)|) \\ & - \frac{(\mathbf{d} - \mathbf{F}\mu_d)^T \psi_d(\mathbf{X}_h, \mathbf{X}_h)^{-1} (\mathbf{d} - \mathbf{F}\mu_d)}{2\sigma_d^2} \end{aligned} \quad (2.24)$$

MLEs of $\hat{\mu}_d$ and $\hat{\sigma}_d^2$ are obtained as

$$\hat{\mu}_d = (\mathbf{F}^T \psi_d(\mathbf{X}_h, \mathbf{X}_h)^{-1} \mathbf{F})^{-1} \mathbf{F}^T \psi_d(\mathbf{X}_h, \mathbf{X}_h)^{-1} \mathbf{d} \quad (2.25)$$

$$\hat{\sigma}_d^2 = \frac{(\mathbf{d} - \mathbf{F}\hat{\mu}_d)^T \psi_d(\mathbf{X}_h, \mathbf{X}_h)^{-1} (\mathbf{d} - \mathbf{F}\hat{\mu}_d)}{n_h} \quad (2.26)$$

where n_h is the number of high-fidelity data, with $\hat{\theta}_h$, \hat{p}_h and ρ found by maximizing

$$\ln(L(\hat{\theta}_d, \hat{p}_d)) = -\frac{n_h}{2} \ln(\hat{\sigma}_d^2) - \frac{1}{2} \ln(|\psi_d(\mathbf{X}_h, \mathbf{X}_h)|) \quad (2.27)$$

To optimize Eqs. (2.29) and (2.34), it is necessary to employ a numerical approach like genetic algorithm (GA), which ensures global search capabilities. This method allows for the maximization of the mentioned equations through an efficient and effective process. The Co-Kriging prediction of the HF model is given by Eq. (2.28).

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_l \\ \mathbf{y}_h \end{pmatrix} = (y_l^{(1)}, \dots, y_l^{(n_l)}, y_h^{(1)}, \dots, y_h^{(n_h)})^T \quad (2.28)$$

$$\hat{\mathbf{y}}(\tilde{\mathbf{x}}) = \hat{\mu} + \mathbf{c}^T \mathbf{C}^{-1} (\mathbf{y} - \mathbf{F}\hat{\mu}) \quad (2.29)$$

$$\mathbf{c} = \begin{pmatrix} \hat{\rho}\hat{\sigma}_l(\mathbf{X}_l, \tilde{\mathbf{x}}) \\ \hat{\rho}^2\hat{\sigma}_l^2\psi_l(\mathbf{X}_h, \tilde{\mathbf{x}}) + \hat{\sigma}_d^2\psi_d(\mathbf{X}_h, \tilde{\mathbf{x}}) \end{pmatrix} \quad (2.30)$$

$$\hat{\mu} = (\mathbf{F}^T \mathbf{C}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{C}^{-1} \mathbf{y} \quad (2.31)$$

2.4. Linear regression multifidelity surrogate

LR-MFS modeling is an innovative approach explored by Zhang et al. [2018]. LR-MFS has been formulated using the fundamental Eq. (2.32). However, in this particular algorithm, the disparity function can be replaced by a matrix, which is depicted as a polynomial response surface approximation as shown in Eq. (2.33).

$$\hat{y}_h(\mathbf{x}) = \rho y_l(\mathbf{x}) + \hat{\delta}(\mathbf{x}) \quad (2.32)$$

$$\hat{\delta}(\mathbf{x}) = \sum_{j=1}^p \xi_j(\mathbf{x}) b_j \quad (2.33)$$

where $\xi_j(\mathbf{x})$ represents the monomial basis, b_j is the corresponding unknown coefficient, and p is the degree of polynomial. This implies that it is possible to estimate the discrepancy matrix through a polynomial function. If a higher-degree polynomial is used, the approximation of the entire system improves, however, the computational cost increases as well. In the case of a one-dimensional problem which will be further elaborated on later, a second-degree polynomial suffices. However, for higher-dimensional cases, because the overall system is more complex, a higher-degree polynomial may be needed Zhang et al. [2018].

Based on this framework, the error is defined as

$$e^{(i)} = y_h^{(i)}(\mathbf{x}_h^{(i)}) - \hat{y}_h(\mathbf{x}_h^{(i)}) \quad (2.34)$$

$$= y_h^{(i)} - \rho y_l(\mathbf{x}_h^{(i)}) - \sum_{j=1}^p \xi_j(\mathbf{x}_h^{(i)}) b_j, i = 1, \dots, n_h$$

Equation (2.34), written as a vector looks like this

$$\mathbf{e} = \mathbf{Y} - \mathbf{XB} \quad (2.35)$$

$$\mathbf{e} = \begin{Bmatrix} e^{(1)} \\ \vdots \\ e^{(n_h)} \end{Bmatrix}, \mathbf{Y} = \begin{Bmatrix} y_h^{(1)} \\ \vdots \\ y_h^{(n_h)} \end{Bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} y_l(\mathbf{x}_l^{(1)}) & \xi_1(\mathbf{x}_h^{(1)}) & \dots & \xi_p(\mathbf{x}_h^{(1)}) \\ \vdots & \vdots & \ddots & \vdots \\ y_l(\mathbf{x}_h^{(n_h)}) & \xi_1(\mathbf{x}_h^{(n_h)}) & \dots & \xi_p(\mathbf{x}_h^{(n_h)}) \end{bmatrix}, \mathbf{B} = \begin{Bmatrix} \rho \\ b_1 \\ \vdots \\ b_p \end{Bmatrix}$$

To determine the unknown coefficient vector \mathbf{B} , we can employ the technique of minimizing the sum of squared errors. Within this framework, ρ serves as a scaling factor. As previously noted, when ρ is excessively high, it indicates a substantial disparity between low- and high-fidelity models. The unknown parameters in LR-MFS are obtained by using the standard regression least-squares technique given as

$$\mathbf{B} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2.36)$$

2.5. Co-Kriging-LR-MFS ensemble model

Sections 2.1 and 2.4 demonstrated that Co-Kriging is primarily focused on interpolation, while LR-MFS leans towards regression analysis. As a result, Co-Kriging excels at capturing local variations accurately, whereas LR-MFS is better suited for capturing broader global trends Rumpfkeil et al. [2022]. Interpolation involves the surrogate passing through all sample points, encompassing radial basis functions and Kriging. Conversely, the regression model can create a smoother model and alleviate the overfitting issue of the surrogate model. Typically, it doesn't necessitate the prediction of the sample point to be identical to the actual response. Typical regression surrogate models comprise the polynomial response surface, support vector regression, and moving least squares. Liu [2022] Moreover, extensive research has demonstrated that there is

no universal single-fidelity surrogate model that can outperform others in all scenarios. [Goel et al. 2007] Consequently, selecting an appropriate single-fidelity surrogate model beforehand poses a considerable challenge for practitioners. Thus, by merging the strengths of both Co-Kriging and LR-MFS it might be possible to create a more efficient MFSM technique.

2.5.1. Co-Kriging and LR-MFS comparison

The variability in the performance of Co-Kriging and LR-MFS across different regions can be attributed to a multitude of factors that are related to the data characteristics and the underlying system being modeled. By exploring various factors, we can gain a better understanding of this phenomenon.

2.5.1.1. Data variability

Co-Kriging is a statistical method that takes into account the spatial correlation between data points. If the data exhibit significant spatial variability or spatial autocorrelation, Co-Kriging can capture and exploit this information effectively. In regions with strong spatial correlations, Co-Kriging is likely to perform better.

2.5.1.2. Linear relationships

Linear regression assumes that the relationship between input variables and the response variable is linear. If the underlying system exhibits predominantly linear behavior in certain regions, the linear modeling approach is more appropriate and can capture the relationships accurately.

2.5.1.3. Nonlinear effects

Some regions of the modeling space may exhibit nonlinear behavior or interactions that cannot be effectively captured by a linear model. In such cases, Co-Kriging, which can capture nonlinear spatial correlations, may perform better.

2.5.1.4. Data density

The density of data points in different regions can significantly impact the model's performance. [Rahman et al. 2013] If there are more data points in regions where Co-Kriging performs well, the model may have more information to make accurate predictions. In contrast, if data in some regions are sparse, where linear modeling is better, the LR-MFS can perform better due to a simpler model structure.

2.5.1.5. Model assumptions

Co-Kriging and linear regression make different assumptions about the data and the underlying system. For example, Co-Kriging assumes that the data have a stationary covariance structure, while linear regression assumes a linear relationship. These assumptions may be more valid in certain regions of the problem space.

2.5.1.6. Noise and uncertainty

The presence of noise or uncertainty in the data can also impact the choice of modeling approach. Co-Kriging can help to reduce the effects of noise in regions with strong spatial correlation, while LR-MFS may struggle to handle noisy data.

Table 2.2 provides a comprehensive overview of the key differences between Co-Kriging and LR-MFS methodologies. In the case where the surrogate model has stochastic features, such as Co-Kriging, it will undergo multiple iterations to ensure accuracy. The resulting average root mean square error (RMSE) and its corresponding standard deviation will then be documented. Conversely, for a deterministic surrogate model like LR-MFS, only one iteration is necessary to obtain reliable results.

Table 2.2 Comparison of Co-Kriging and LR-MFS models

Criteria	Co-Kriging	LR-MFS
Computational cost	Expensive	Cheap
Low fidelity data	Use all low-fidelity data at all locations (may not necessarily coincide with high-fidelity data locations) with $\mathbf{X}_h \subset \mathbf{X}_l$ condition	Only use low fidelity data at high fidelity locations
Relation	Can handle nonlinear spatial correlations	Better at finding linear relations
Data density	Better for dense regions	Better for sparse regions
Algorithm type	Involving optimization algorithm and probability	No probability

2.5.2. Co-Kriging-LR-MFS algorithm

Based on the reasons mentioned earlier, we hypothesize that the combination of Co-Kriging and the LR-MFS approach would lead to a better model in terms of accuracy.

We first divide the available high-fidelity data into two groups. The first group is used for training the Co-Kriging and LR-MFS models. For the LR-MFS, we use low-fidelity data at the points that are co-located at the high-fidelity data points. In Co-Kriging, the low-fidelity points are at different locations, thus, providing additional information to the system. The second group of data is used to evaluate the error in the results of the two algorithms, verifying which algorithm performs better in a specific region. For this purpose, we employ the *K-means* clustering technique, [Na 2010] to divide the data space into specific regions with assigned boundaries. In Section 2.5.3., we will discuss the details of this algorithm.

In the following step, we calculate the mean absolute error between the responses predicted by each algorithm and the true value of responses in each region. If the majority of points in a specific region have a lower error in one algorithm, we choose that algorithm as the predictive model for that region. This process is repeated for each region to find the appropriate predictive model.

In the end, we have an ensemble approach that can choose the surrogate model in each region of the data space based on the calculated error in that region. This approach helps save computational time, considering that the Co-Kriging approach is a computationally expensive algorithm, and we make use of the majority of available data for assigning model to each region. Although the LR-MFS model is fast, it cannot capture nonlinear trends and needs information from the Co-Kriging model to improve its performance. Thus, the final result is a fast algorithm with higher accuracy, leveraging the advantages of both models.

2.5.3. K-means clustering

K-means clustering is a widely utilized algorithm in the field of unsupervised Machine Learning. Its purpose is to cluster data points based on their similarities or proximities, effectively grouping them into distinct clusters. The main objective of K-means clustering is to divide a dataset into non-overlapping clusters, with the number of desired clusters determined by the user-defined parameter K . The process of K-means clustering involves several steps as described below.

2.5.3.1. Centroids selection

First, a set of K initial centroids is selected to represent the cluster centers. The choice of these initial centroids can impact the outcome of the algorithm, and there are different approaches for their initialization, such as random selection or more advanced techniques.

2.5.3.2. Clustering based on distance

Next, each data point in the dataset is assigned to the nearest centroid based on a distance metric, typically Euclidean distance. Euclidean distance is commonly used as a distance metric to measure the similarity or dissimilarity between two data points in a multi-dimensional space. This assignment step places each data point in a cluster based on its distance to the associated centroid.

2.5.3.3. Updating centroids based on mean data of each cluster

Afterwards, the centroids are updated by calculating the mean (average) position of all the data points assigned to each cluster. By moving the centroids towards the center of their respective clusters, they better represent their members.

2.5.3.4. Repeating assignment and update

The assignment and update steps are repeated iteratively until certain convergence criteria are met. Convergence can be defined by reaching a maximum number of iterations or when there is minimal change in the positions of cluster centroids in consecutive iterations. Other stopping conditions may also be employed. Once convergence is achieved, we obtain final centroids that serve as representatives for each cluster. Additionally, every data point becomes associated with the cluster corresponding to its nearest centroid within this resulting configuration.

2.6. Predictive model evaluation

Different metrics can be used to evaluate the performance of regression models, and the choice of metric depends on the specific problem and goals of the analysis. Table 2.3 shows some common error metrics and their applications. The RMSE evaluation metric is used in this thesis for determining the accuracy of different models.

Table 2.3 Error metrics with advantages and disadvantages

Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^n y_i - \hat{y}_i }{n}$ <p>y_i: true value \hat{y}_i: predicted value</p>	<p>Advantages:</p> <ul style="list-style-type: none"> • Easy to understand and interpret. • Resistant to outliers since it uses the absolute differences between predicted and actual values. • Produces a metric in the same units as the target variable, which is easy to interpret. <p>Disadvantages:</p> <ul style="list-style-type: none"> • All errors are treated equally; it does not penalize large errors more than small errors.
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$ <p>y_i: true value \hat{y}_i: predicted value</p>	<p>Advantages:</p> <ul style="list-style-type: none"> • Gives higher weight to larger errors, making it more sensitive to outliers. • It considers the overall or average error across the entire dataset, providing a comprehensive measure of the model's performance. <p>Disadvantages:</p> <ul style="list-style-type: none"> • Sensitive to outliers, which can distort the evaluation if outliers are present. • Square root makes the interpretation less intuitive compared to MAE.
Relative Mean Error (RME)	$RME = \frac{1}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{ y_i } \times 100$ <p>y_i: true value \hat{y}_i: predicted value</p>	<p>Advantages:</p> <ul style="list-style-type: none"> • Expresses the error as a percentage of the actual values, making it more interpretable and easier to compare across different datasets. • Emphasizes the relative size of the errors. <p>Disadvantages:</p> <ul style="list-style-type: none"> • It can become problematic when actual values are close to zero, leading to division by small numbers or zero.
Mean Squared Log Error (MSLE)	$MSLE = \frac{1}{n} \sum_{i=0}^n (\log(y_i + 1) - \log(\hat{y}_i + 1))^2$ <p>y_i: true value \hat{y}_i: predicted value</p>	<p>Advantages:</p> <ul style="list-style-type: none"> • Treats small differences between small actual and predicted values same as big differences between large actual and predicted values. <p>Disadvantages:</p> <ul style="list-style-type: none"> • Penalizes underestimates more than the overestimates.
R ²	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$ <p>y_i: true value \hat{y}_i: predicted value \bar{y}: mean value of true responses</p>	<p>Advantages:</p> <ul style="list-style-type: none"> • R² is easy to interpret. It provides a value between 0 and 1, where 0 indicates that the model does not explain any of the variance in the dependent variable, and 1 indicates that the model explains all of the variance. <p>Disadvantages:</p> <ul style="list-style-type: none"> • Inadequate for nonlinear models [Spiess et al. 2010]

CHAPTER 3

RESULTS AND DISCUSSION OF EXAMPLE PROBLEMS

This chapter presents the results obtained from applying the Co-Kriging model, Linear Regression Multifidelity Surrogate (LR-MFS) model, and the innovative Co-Kriging-LR-MFS ensemble model to various test cases. The chapter begins by analyzing the performance of these models in a one-dimensional case, namely the Forrester function.

Following this initial evaluation, the focus shifts to more complex scenarios in two dimensions. The first case involves the application of these models to solve the Laplace equation, while the second example involves the Peaks function. For both 2D cases, the Co-Kriging, LR-MFS, and the newly introduced Co-Kriging-LR-MFS ensemble model are implemented and thoroughly compared based on several performance metrics.

In the Data Analysis section, we will investigate the effect of the number of training data on the final system prediction.

3.1. Example problem one: Forrester function

3.1.1. Co-Kriging model

To validate the code for the Kriging and Co-Kriging approaches, we utilized Equations 3.1 and 3.2 from the Forrester et al. [2007] as sources of HF and LF data sets, respectively, and given as

$$f_h(x) = (6x - 2)^2 \sin(12x - 4) \quad (3.1)$$

$$f_l(x) = 0.5f_h(x) + 10(x - 0.5) - 5 \quad (3.2)$$

The HF data locations are denoted as:

$$\mathbf{x}_h = \{0, 0.4, 0.6, 1\}^T$$

while the LF data locations are represented by:

$$\mathbf{x}_l = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}^T$$

To verify the developed code against the original article, the results are shown in Figure 3.1. Observing Figure 3.1, when employing Kriging solely based on high-fidelity locations, an

interpolative prediction is obtained through those points, and displays a rather large error in comparison to the curve for $f_h(x)$. However, by incorporating LF data as an additional source and applying Co-Kriging the prediction is enhanced, resulting in an RMSE of 0.0588.

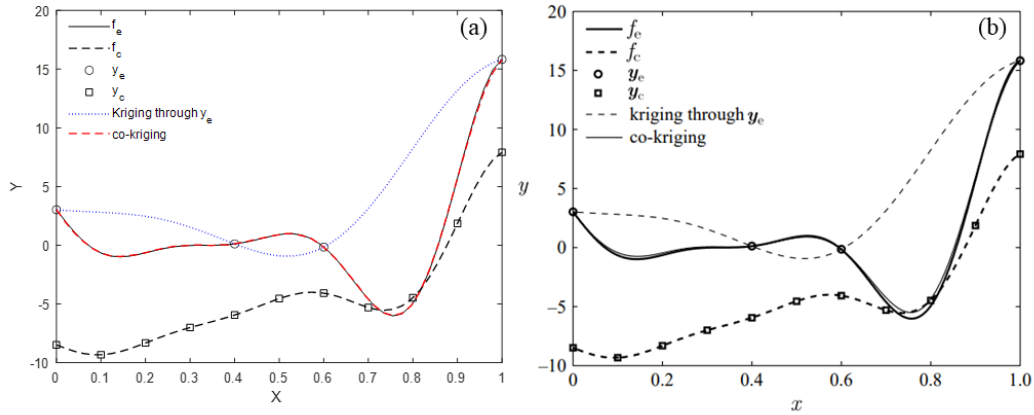


Figure 3.1 Kriging and Co-Kriging predictions on Forrester function's equations using (a) generated codes and (b) [Forrester et al., 2007]

3.1.2. LR-MFS model

Figure 3.2 illustrates the utilization of the same dataset for the LR-MFS model. One disadvantage of the LR-MFS model is that it only allows the use of low-fidelity data points in high-fidelity locations. Consequently, we cannot utilize all of the available low-fidelity data points, but only those present at the high-fidelity locations. Therefore, when employing the same high-fidelity and low-fidelity data sources while encountering this limitation, it becomes evident that the LR-MFS struggles to produce accurate predictions, resulting in a significantly large RMSE of 4.536. In contrast, Co-Kriging outperforms LR-MFS remarkably in this scenario by leveraging the capability to use low-fidelity data in various locations beyond those at high-fidelity locations.

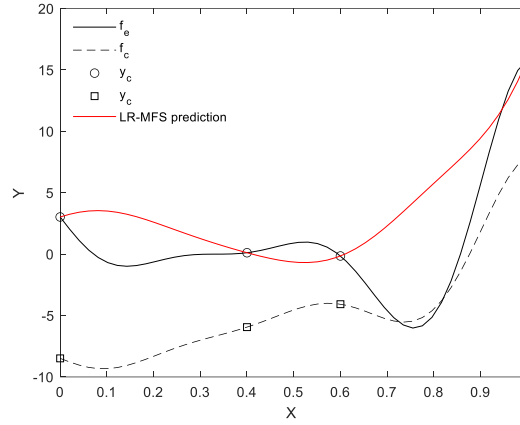


Figure 3.2 LR-MFS prediction on Forrester function's equations.

3.2. Example problem two: Laplace's equation

Various physical phenomena (e.g., 2D steady-state heat conduction, the static deflection of a membrane, and electrostatic potential) give rise to Laplace's equation. In the realm of heat conduction, Laplace's equation emerges when analyzing two-dimensional systems in a state of thermal equilibrium. This equation governs the distribution of temperature within these systems, providing valuable insights into how heat is transferred and dissipated. The static deflection of a membrane also falls under the purview of Laplace's equation. When examining thin membranes subjected to external forces or pressures, this equation describes their deformed shapes. Understanding these deformations is crucial for various applications involving structures like drumheads or stretched fabrics. Moreover, Laplace's equation finds relevance in studying electrostatic potential distributions. It arises when investigating electric fields generated by stationary charges in empty space. By solving this equation, one can determine the electric potential at any point surrounding these charges, enabling precise calculations and predictions regarding electrical interactions. In this example problem, the analytical solution for the Laplace's equation is treated as the high-fidelity source of data, and a numerical solution is considered as the low-fidelity source of data.

3.2.1. High and low fidelity data sources

3.2.1.1. High fidelity: Analytical solution of 2D Laplace's equation

Laplace's equation in Cartesian coordinates can be written as

$$\nabla^2 u(x, y) = \frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = 0 \quad (3.3)$$

where $u(x, y)$ represents the independent variable, e.g., electrostatic potential.

Assuming $u(x, y)$ can be factorized as a product of two independent one-dimensional functions in the form

$$u(x, y) = v(x)\omega(y) \quad (3.4)$$

Eq. (3.1) can then be written as

$$v_{xx}\omega + v\omega_{yy} = 0 \quad (3.5)$$

where $v_{xx} = \partial^2 v(x)/\partial x^2$ and $\omega_{yy} = \partial^2 \omega(y)/\partial y^2$. Through separation of variables, Eq. (3.3) can be expressed as

$$\frac{v_{xx}}{v} = -\frac{\omega_{yy}}{\omega} = \lambda \quad (3.6)$$

Now we have two independent differential equations

$$\begin{cases} v_{xx} - \lambda v = 0 \\ \omega_{yy} + \lambda \omega = 0 \end{cases} \quad (3.7)$$

The solutions for different conditions are summarized in Table 3.1 for negative, zero, and positive values of λ .

Table 3.1 General solutions for Laplace's equation.

λ	$v(x)$	$\omega(y)$	$u(x, y)$ $= v(x)\omega(y)$
$\alpha^2 > 0$	$\cos(\alpha x), \sin(\alpha x)$	$\cosh(\alpha y), \sinh(\alpha y)$	$\cos(\alpha x) \cosh(\alpha y),$ $\sin(\alpha x) \sinh(\alpha y),$ $\cos(\alpha x) \sinh(\alpha y),$ $\sin(\alpha x) \cosh(\alpha y)$
0	$x, 1$	$1, y$	$1, x, y, xy$
$\alpha^2 < 0$	$\cosh(\alpha x), \sinh(\alpha x)$	$\cos(\alpha y), \sin(\alpha y)$	$\cosh(\alpha x) \sin(\alpha y),$ $\sinh(\alpha x) \sin(\alpha y),$ $\cosh(\alpha x) \cos(\alpha y),$ $\sinh(\alpha x) \cos(\alpha y)$

For a specific solution, the two-dimensional domain and the associated boundary conditions need to be identified. For a rectangular domain, one possible set of boundary conditions is shown in Figure 3.3, where a is equal to π and b equal to $\frac{\pi}{3}$. Also, 20 grid points are used in x and y direction for discretization purposes.

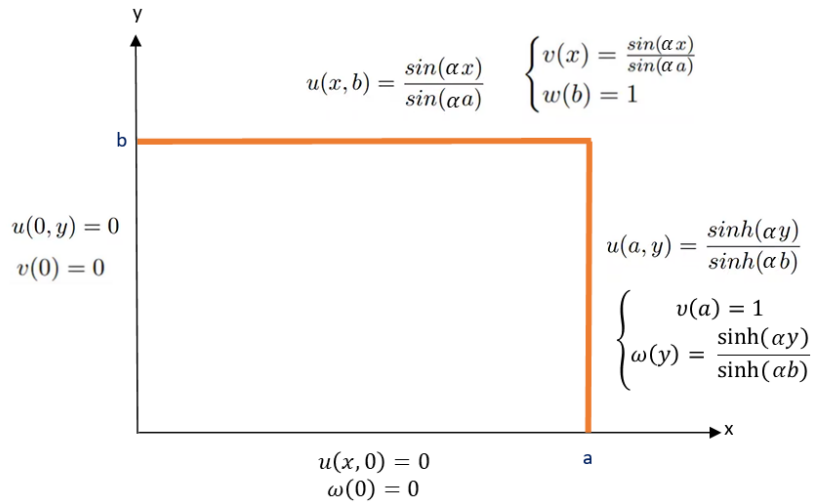


Figure 3.3 An example of the solution domain and boundary conditions of Laplace's equation.

By applying the boundary conditions in Figure 3.3, we find

$$v(x) = \begin{cases} A\cos(\alpha x) + B\sin(\alpha x) & \lambda < 0 \\ Cx + D & \lambda = 0 \\ E\cosh(\alpha x) + F\sinh(\alpha x) & \lambda > 0 \end{cases} \quad (3.8)$$

$$v(0) = 0 \rightarrow A = D = E = 0 \quad (3.9)$$

$$v(x) = \frac{\sin(\alpha x)}{\sin(\alpha a)} \rightarrow C = F = 0 \text{ and } B = \frac{1}{\sin(\alpha a)} \quad (3.10)$$

Based on the boundary conditions we conclude that $v(x) = \frac{\sin(\alpha x)}{\sin(\alpha a)}$ is a valid solution. We need to apply the other boundary conditions to find the valid solutions for $\omega(y)$; based on the last part, this is the only acceptable solution for $\omega(y)$:

$$\omega(y) = A \cosh(\alpha y) + B \sinh(\alpha y) \quad (3.11)$$

$$\omega(0) = 0 \rightarrow A = 0 \text{ and } B = \frac{1}{\sinh(\alpha b)} \quad (3.12)$$

This is the overall analytical solution for Laplace's equation with assigned boundary conditions:

$$u(x, y) = v(x)\omega(y) = \frac{\sin(\alpha x) \sinh(\alpha y)}{\sin(\alpha a) \sinh(\alpha b)} \quad (3.13)$$

3.2.1.2. Low fidelity: Numerical solution of 2D Laplace's equation

The domain in Figure 3.3 is discretized as shown in Figure 3.4. The second partial derivatives in Eq. (3.1) can be estimated by using a central finite difference scheme on red dot locations as

$$\frac{\partial^2 u}{\partial x^2} \approx \frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{(\Delta x)^2} \quad (3.14)$$

$$\frac{\partial^2 u}{\partial y^2} \approx \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{(\Delta y)^2} \quad (3.15)$$

where Δx and Δy are the finite distances in x and y directions, respectively.

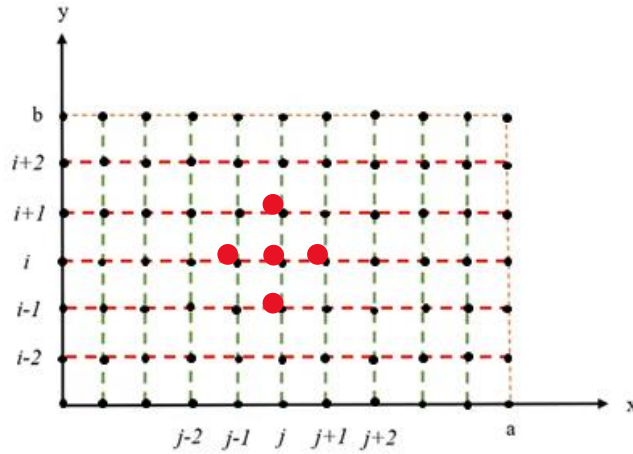


Figure 3.4 The solution domain discretization for Laplace's equation.

If we substitute the above derivatives into the main equation, we have

$$\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{(\Delta x)^2} + \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{(\Delta y)^2} = 0 \quad (3.16)$$

Based on above equation, the value of function $u(x, y)$ at the discretized location i, j can be approximated as

$$u_{i,j} = \frac{(\Delta x)^2(u_{i+1,j} + u_{i-1,j}) + (\Delta y)^2(u_{i,j+1} + u_{i,j-1})}{2((\Delta x)^2 + (\Delta y)^2)} \quad (3.17)$$

Figure 3.5 (a) and (b) show the analytical and numerical solutions for the entire domain, respectively. However, since the boundary condition varies at the top of the domain (i.e., $y = b$), the most significant information resides in this section. The left and bottom domains maintain a constant value of 0. Consequently, based on this information, we have omitted the bottom section of the domain and focused solely on considering nodal values ranging from 0.471 to $\frac{\pi}{3}$ for this problem. As a result, the final domain is depicted in Figure 3.6.

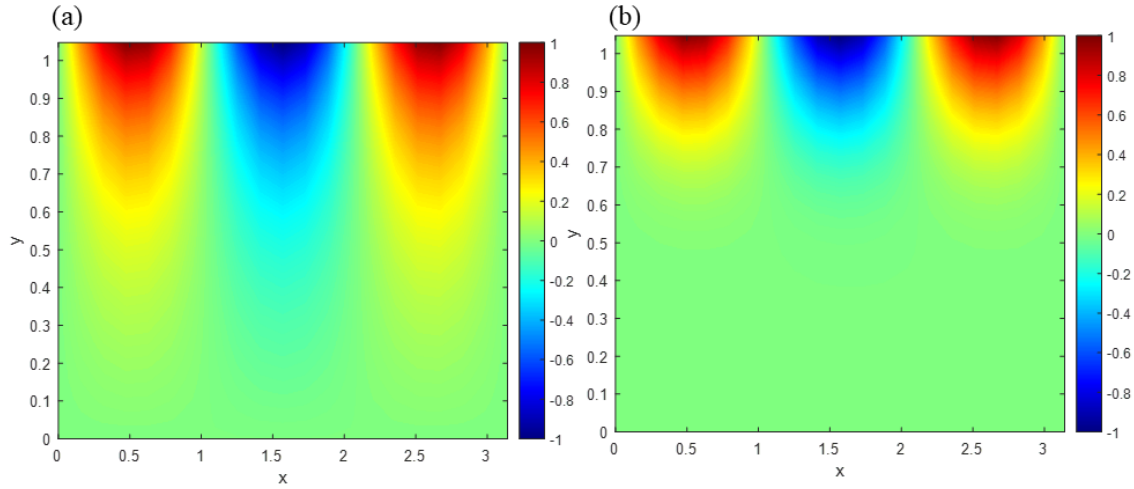


Figure 3.5 (a) Analytical solution of Laplace's equation in whole domain. (b) Semi-converged numerical solution of Laplace's equation in whole domain.

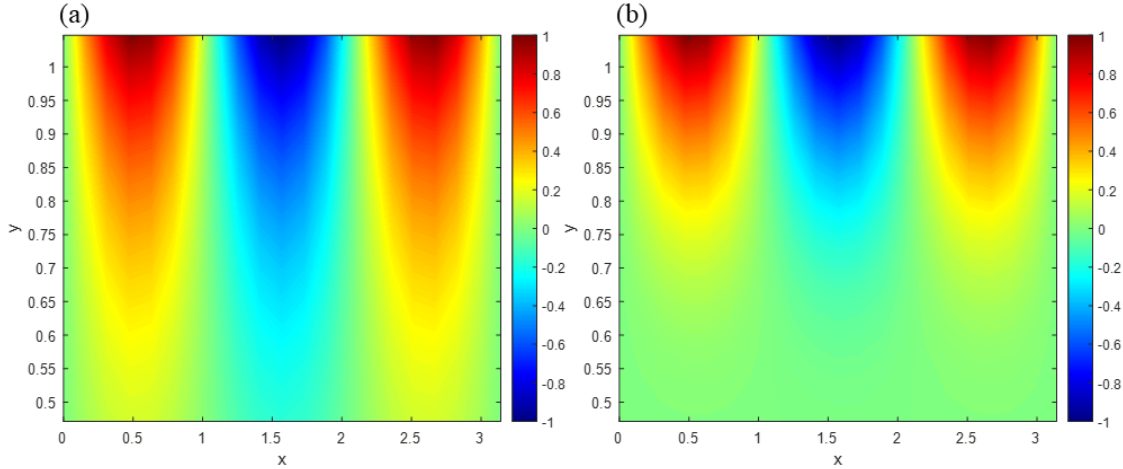


Figure 3.6 (a) Analytical solution of Laplace's equation (high-fidelity data source). (b) Semi-converged numerical solution of Laplace's equation (low-fidelity data source).

As described in section 1.3, one source for LF data is based on the degree of solution convergence. In this example problem, the analytical solution to the Laplace's equation is used as the HF data source and the semi-converged numerical solution is used as the LF data source, as shown in Figure 3.5. This numerical solution needs 100 iterations to converge; however, the solution with 15 iterations is used as a LF data source.

3.2.2. Co-Kriging model

The rectangular domain in Figure 3.4 is used for identifying 20 locations for collecting data using the Latin Hypercube Sampling (LHS) approach. LHS is a stratified sampling method that divides the sample space into equally probable intervals. In each interval, one and only one sample point is selected. The key feature of LHS is that it ensures a more even and representative coverage of the input parameter space compared to simple random sampling. This feature can be especially advantageous in simulations and experiments where a thorough exploration of the input space is essential. The same 20 locations are used for obtaining responses (observations) from the high- and low-fidelity solutions.

3.2.3. LR-MFS model

The same 20 data points used in the Co-Kriging model were also used for training the LR-MFS model. Eq. 2.28 implies that it is possible to estimate the discrepancy matrix through a polynomial function. For this case the 4th degree polynomial has the lowest RMSE error.

3.2.4. Co-Kriging-LR-MFS ensemble methods

In the third approach, we combined the two models to create an ensemble model, aiming to leverage the strengths of both approaches. We utilized 60 data points, generated by LHS method in design domain, to assess the absolute error for each point, enabling the identification of the model with superior performance in specific regions. By considering the majority of points with lower error, we selected a model with lower absolute error in those particular regions. Figure 3.7 illustrates a region clustering using the K-means approach, as discussed in Section 2.5.3. The points for each region are identified with a unique color. The magenta triangle shows the center of each cluster. Table 3.2 shows the selected method in each region for the Laplace case study.

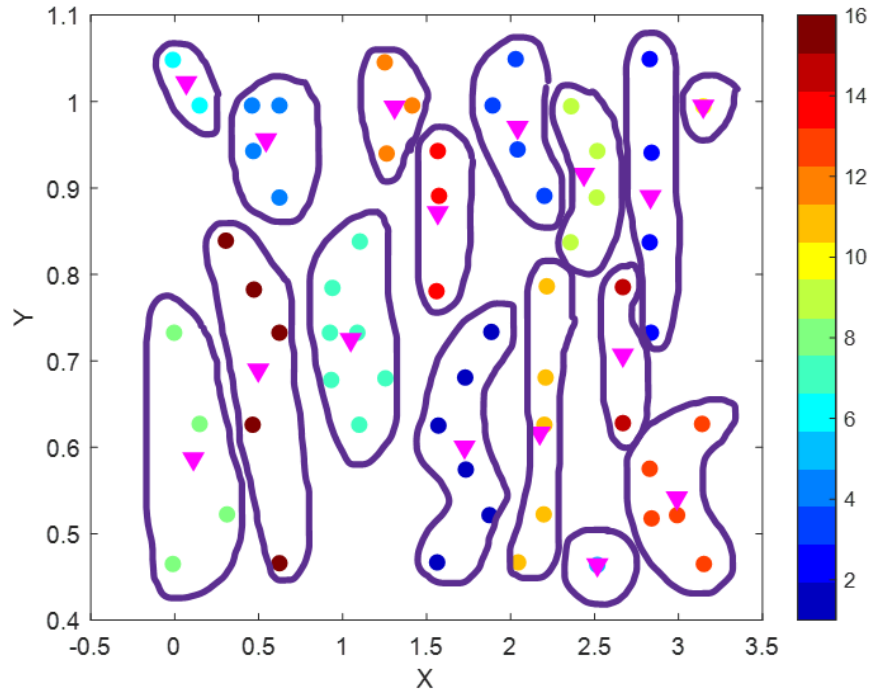


Figure 3.7 Sixteen different regions for Laplace case study generated by K-means clustering where the magenta triangles represent the regions' centers. The color bar shows cluster number based on assigned color.

Table 3.2 Best model in each region for Laplace case study.

Region number	Model
1	Co-Kriging
2	LR-MFS
3	LR-MFS
4	LR-MFS
5	Co-Kriging
6	LR-MFS
7	LR-MFS
8	Co-Kriging
9	LR-MFS
10	LR-MFS
11	Co-Kriging
12	LR-MFS
13	Co-Kriging
14	LR-MFS
15	LR-MFS
16	Co-Kriging

3.2.5. Results

Figure 3.8(c) illustrates the final results based on the Co-Kriging model which was discussed in Section 2.3. The overall RMSE of 0.12 for this model is relatively high compared to the LR-MFS. However, in some local locations, the error is lower compared to the LR-MFS. One reason for the high RMSE in comparison to the LR-MFS is that the high- and low-fidelity models exhibit completely linear relationships. Therefore, the LR-MFS is better at capturing these relationships. Figure 3.8(d) shows the prediction based on LR-MFS. As mentioned earlier, in LR-MFS, it is necessary to map a polynomial. However, the degree of the polynomial must be determined through trial and error based on the data. For example, in this case, the minimum RMSE is achieved with $n = 4$. Also, it is obvious that, the RMSE of 0.067 is considerably lower than that of the Co-Kriging model. Figure 3.8(b) displays the results of the Co-Kriging-LR-MFS ensemble model. The RMSE decreased to 0.0494, which is lower than the best predictive model, LR-MFS. This result indicates that although the Co-Kriging model has a much higher RMSE compared to the LR-MFS model, it still outperforms in certain local regions, contributing to an overall improvement in the performance of the LR-MFS model.

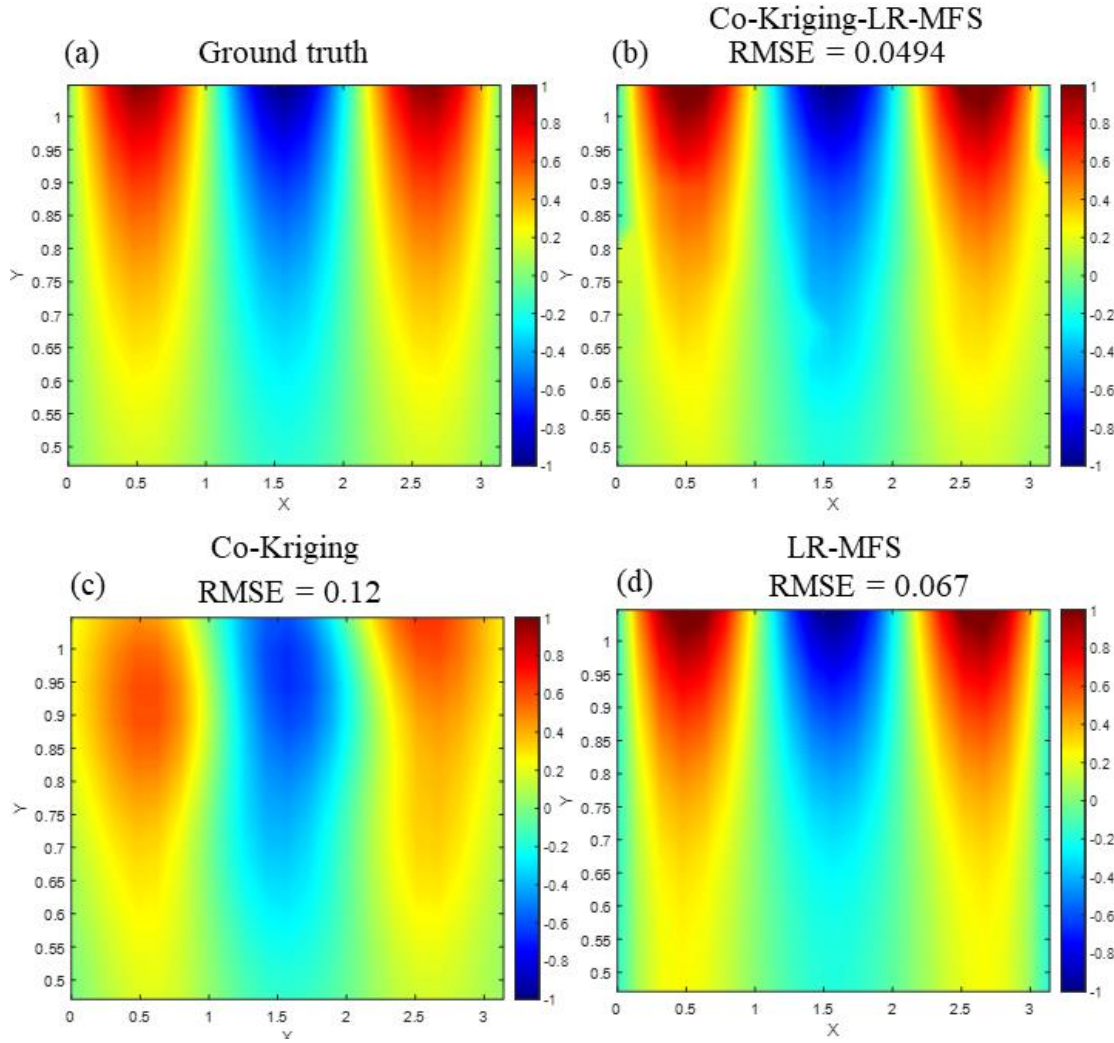


Figure 3.8 (a) Analytical solution of Laplace equation. (b) Co-Kriging-Linear model prediction. (c) Co-Kriging model prediction. (d) Linear model prediction.

3.3. Example problem three: Peaks function

3.3.1. High and low fidelity data sources

The "peaks" function is widely employed in the field of mathematics and numerical optimization as a standard test function to assess the effectiveness of different optimization algorithms. It serves as a reference point for evaluating the performance of various techniques utilized in solving optimization problems. The "peaks" function, being two-dimensional in nature, produces a surface plot characterized by an arrangement of peaks and valleys, which is why it's called the "peaks" function. The standard "peaks" function, typically denoted as $f_h(x_1, x_2)$, is defined as follows and is utilized as a high-fidelity data source for this case study, x_1 and x_2 varies between -3 to 3 :

$$f_h(x_1, x_2) = 3(1 - x_1)^2 e^{-(x_1^2) - (x_2+1)^2} - 10\left(\frac{x}{5} - x_1^3 - x_2^5\right) e^{-x_1^2 - x_2^2} - \frac{1}{3} e^{-(x_1+1)^2 - x_2} \quad (3.18)$$

In order to ensure a data source with reduced fidelity, certain terms have been excluded from the general form of the equation. Additionally, the power of the variable x_1 is changed from 3 to 2, resulting in a non-linear relation between the two sources of fidelity, which is defined as follows:

$$f_l(x_1, x_2) = -10\left(\frac{x}{5} - x_1^2 - x_2^5\right) e^{-x_1^2 - x_2^2} \quad (3.19)$$

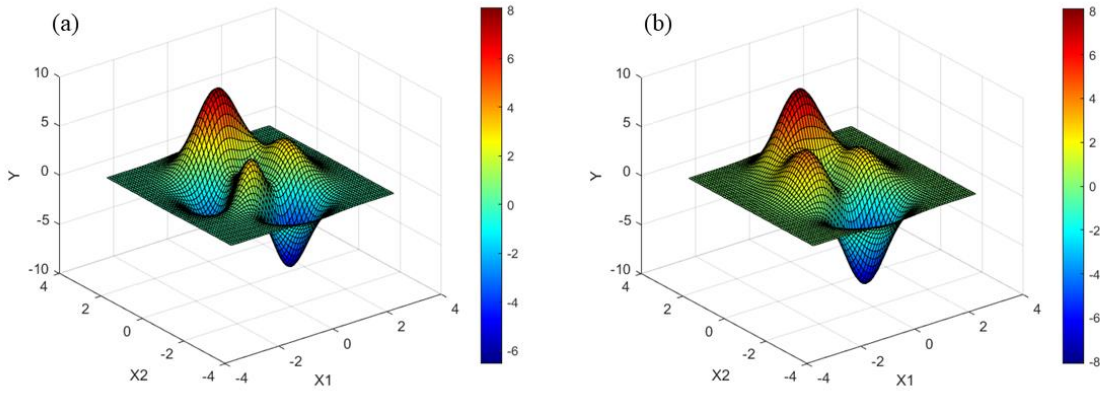


Figure 3.9 Peaks function (a) high-fidelity data source, (b) low-fidelity data source.

3.3.2. Co-Kriging methods

25 high and low-fidelity samples were used at the same locations for training the Co-Kriging model.

3.3.3. LR-MFS methods

The same 25 data points used in LR-MFS at identical locations were employed for training the Co-Kriging model. In this case, the minimum RMSE is achieved with a 3-degree polynomial.

3.3.4. Co-Kriging-linear ensemble methods

In the third approach, we combined the two models to create an ensemble model, aiming to leverage the strengths of both approaches. Figure 3.10 shows the clustering of 80 data points based on the K-means algorithm. As we can see, there are 9 regions, each indicated with its identical center point. We calculated the mean absolute error in each region and assigned the model with better performance based on the majority of points with lower error to the specific region. Table

3.3 shows the selective method in each region. For example, in region 1, the LR-MFS has better performance; however, in region 2, the Co-Kriging method is better.

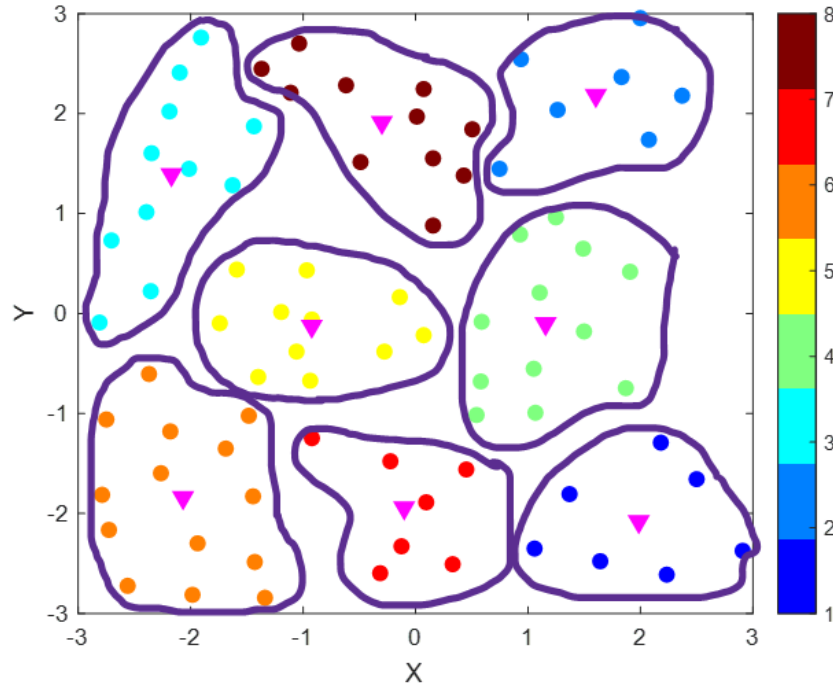


Figure 3.10 Nine different regions for peaks function case study generated by K-means clustering where the magenta triangles represent the regions centers. The color bar shows cluster number based on assigned color.

Table 3.3 Best model in each region for peaks function case study.

Region number	Model
1	LR-MFS
2	Co-Kriging
3	LR-MFS
4	LR-MFS
5	Co-Kriging
6	Co-Kriging
7	LR-MFS
8	LR-MFS
9	Co-Kriging

3.3.5. Results

Figure 3.9(a) and (b) show the 3D prediction models of Co-Kriging and LR-MFS, respectively. The results show that Co-Kriging demonstrates good performance in certain regions. For example, it excels in the region marked with a yellow star in Figure 3.9(a). On the other hand, the LR-MFS exhibits superior performance at the global maximum location, marked with a blue star in Figure 3.9(b). As we can see in Figure 3.11, the results show that the ensemble model has an RMSE of 0.765, which is a lower error compared to both the Co-Kriging and LR-MFS. In other word, we combined the two models to create an ensemble model, aiming to leverage the strengths of both approaches. However, we observe that in some regions, for example, near the boundary, the LR-MFS is more accurate; One way to decrease these types of errors is to shrink the regions by having more clusters, enabling the capture of a more accurate predictive model in each region. In the end, this leads to a better overall model with a lower RMSE.

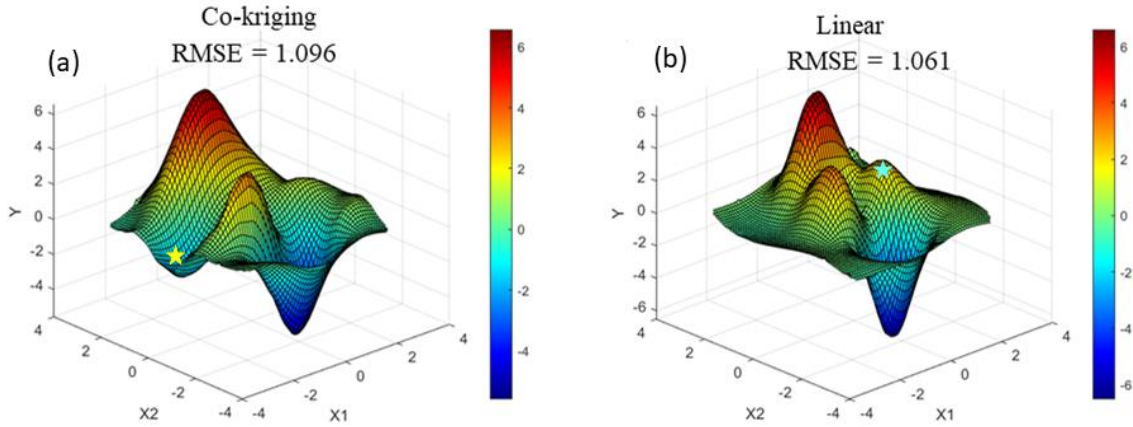


Figure 3.11 Peaks function prediction by (a) Co-Kriging model and (b) LR-MFS. The yellow star marks the region where Co-Kriging outperforms Linear model. The blue star marks the region where LR-MFS is better than Co-Kriging model.

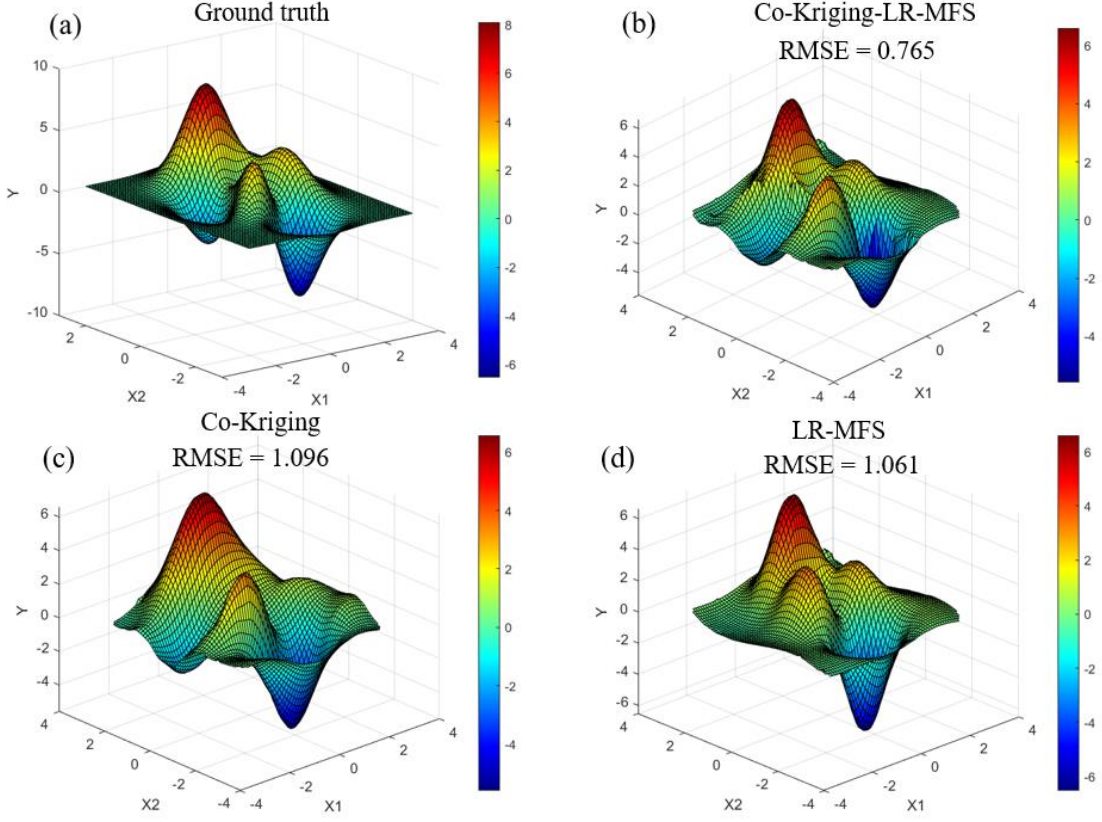


Figure 3.12 (a) Peaks function. (b) Co-Kriging-Linear model prediction. (c) Co-Kriging model prediction. (d) Linear model prediction.

3.4. Data analytics

In this section, data analysis has been conducted to gain a better perspective on determining the sufficient amount of data. For instance, the second case study, the Peaks function, was selected.

3.4.1. One dimensional data analysis

To further investigate data points for the Co-Kriging algorithm, Eqs. 3.20 and 3.21 were utilized as HF and LF data sources, respectively.

$$yf_h(x) = \cos(3.5\pi x)\exp(-1.4x) \quad (3.20)$$

$$f_l(x) = \cos(3.5\pi x)\exp(-1.4x) + 0.75x^2 \quad (3.21)$$

The locations and quantity of LF data $X_l = \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ remained consistent across all these cases. However, in the initial study, three HF points were used, each in different locations. Table 3.4 demonstrates that the ultimate RMSE of the predictive model, relying on the Co-Kriging algorithm, significantly hinges on the training dataset. In case one, we obtained an RMSE of 0.43. Conversely, in case four, utilizing the same LF dataset but different HF datasets, we achieved an RMSE of 0.186. Remarkably, the latter RMSE is substantially lower than that of the first case.

Table 3.4 one-dimensional case with different HF locations using 3 data points

Case number	X_h	RMSE
Case 1	$\{0, 0.5, 1\}$	0.435
Case 2	$\{0.18, 0.37, 0.62\}$	0.222
Case 3	$\{0.25, 0.5, 0.75\}$	0.193
Case 4	$\{0.26, 0.55, 0.77\}$	0.186

Table 3.5 displays the RMSE obtained using four high-fidelity data points situated at various locations. The data points $X_l = \{0, 0.5, 1\}$ remain consistent across all cases, while the fourth data point is positioned differently. Each scenario resulted in a distinct RMSE value. It can be inferred that the placement of the fourth point significantly impacts the overall RMSE.

Table 3.5 one-dimension case with different high-fidelity locations using 4 data points

Case number	X_h	RMSE
Case 1	$\{0, 0.25, 0.5, 1\}$	0.014
Case 2	$\{0, 0.1, 0.5, 1\}$	0.013
Case 3	$\{0, 0.5, 0.6, 1\}$	0.026

3.4.2. 2D-dimensional Data analysis

3.4.2.1. Data analysis for Kriging algorithm

In the case study, we used 20 data points from both high and low fidelity locations. As we progressed in this section, the number of data points increased in both high and low fidelity areas. Figure 3.12 (a) demonstrates that as we added more data, the RMSE decreased as expected.

However, when we reached 30 data points, the RMSE was higher compared to having 25 data points. This proves that more data doesn't always assist in constructing a better predictive model. Sometimes, the additional data might act as noise, not contributing any useful information to decrease the RMSE. With 50 and 60 data points, the RMSE values were 0.509 and 0.4118 respectively, showing a decreasing trend, which makes sense. Generally, more data leads to lower RMSE. However, as we increased the number of data points to 70, although the RMSE decreased to 0.308, the predictive model's performance on test data became nonsensical, It also causes some local minima in the boundaries, which are not aligned with the ground truth function, as seen in Figure 3.13 (b).

In conclusion, the optimal number of data points to achieve the minimum RMSE using the Co-Kriging algorithm appears to be 60. This balance allows for improved accuracy without overfitting the model to the training data.

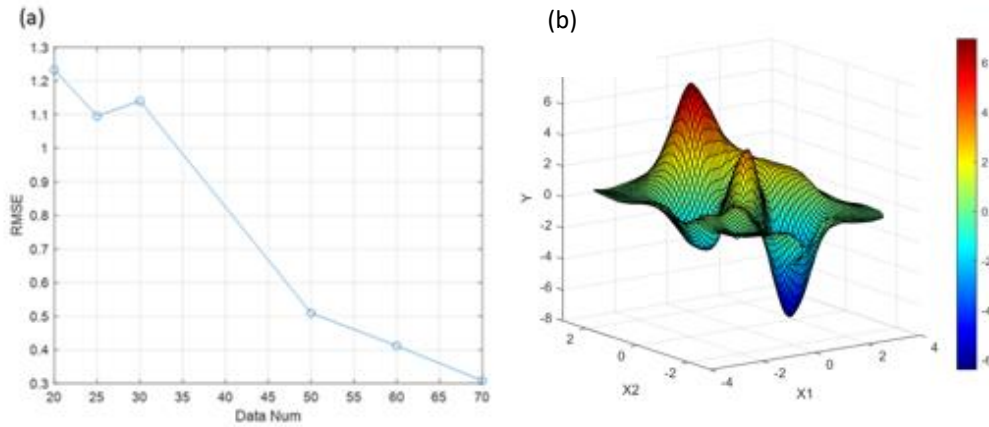


Figure 3.13 (a) Variation of RMSE with Number of Data Points for Co-Kriging Algorithm. (b) Predictive model using 70 data points

3.4.2.2. Data analysis for LR-MFS algorithm

This section presents the data analysis conducted for the LR-MFS. Figure 3.14(a) illustrates that as the number of data points increases to 25 and 30, the RMSE decreases. However, beyond 30 data points, the RMSE begins to rise. When we visualize the results of the predictive model in Figure 3.14(b), it becomes evident that the model exhibits signs of overfitting and fails to capture the underlying trend. Consequently, the optimal number of data points for this case study using the LR-MFS appears to be 30.

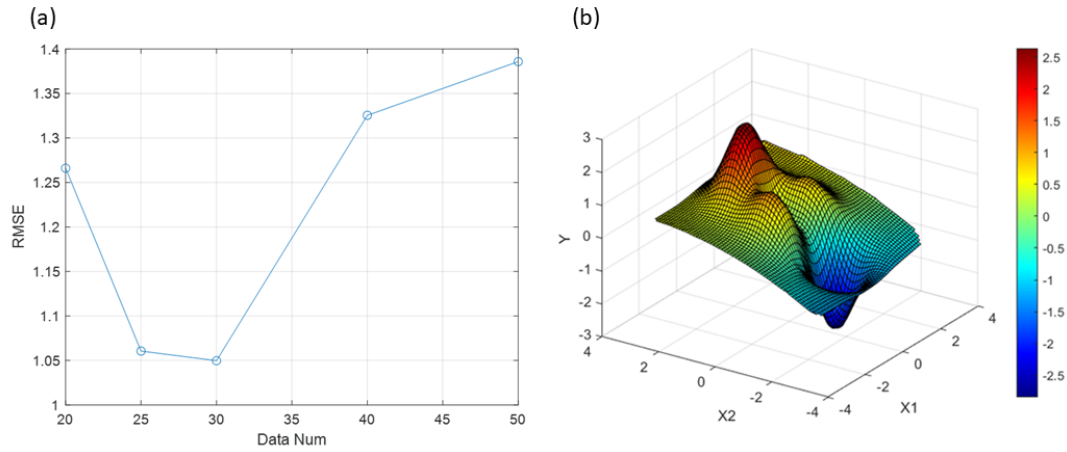


Figure 3.14 (a) Variation of RMSE with Number of Data Points for LR-MFS Algorithm. (b) Predictive model using 40 data points

3.4.2.3. Co-Kriging-LR-MFS Ensemble modeling in optimum data points

In sections 3.4.2.1 and 3.4.2.2, it was concluded that the optimal number of data points for the Co-Kriging and LR-MFS is 60 and 30 points, respectively. This section focuses on constructing the Co-Kriging- LR-MFS ensemble model, utilizing the optimal number of data points determined for both Co-Kriging and LR-MFS algorithms. Figure 3.15(a) and (b) illustrates a notable decrease in RMSE from 0.4118 to 0.266 when employing the Co-Kriging- LR-MFS ensemble modeling approach. This demonstrates that while utilizing Co-Kriging with 60 data points for training resulted in an RMSE decrease to 0.4118, the LR-MFS with an RMSE of 1.0497 still contributes valuable information to the system. Integrating this LR-MFS enhances the performance of the ensemble model, leading to a decrease in RMSE and an overall improvement in the system's predictive capabilities.

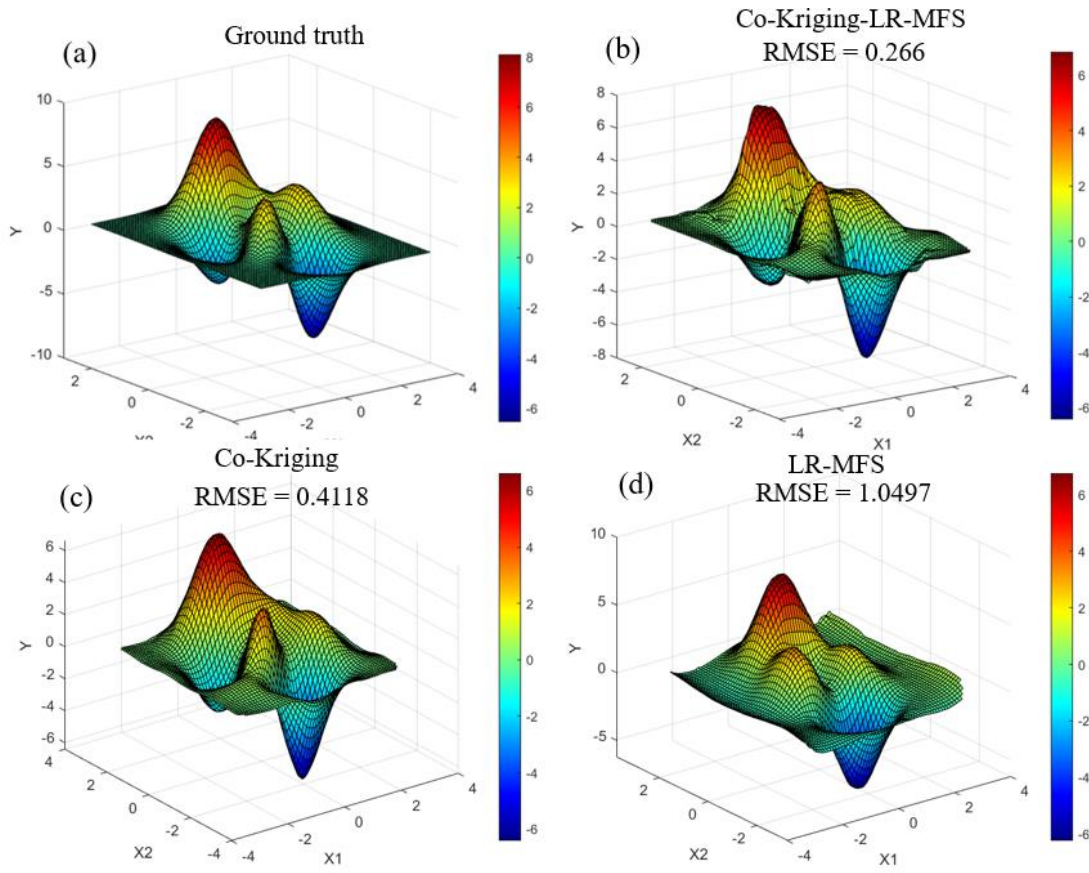


Figure 3.15 (a) Peaks function. (b) Co-Kriging-Linear model prediction. (c) Co-Kriging model prediction. (d) Linear model prediction.

CHAPTER 4

CONCLUSION

This thesis research focused on the implementation and exploration of the Co-Kriging model and the Linear Regression Multifidelity Surrogate (LR-MFS) model, both renowned algorithms in multifidelity systems. Additionally, an ensemble model was introduced by combining these two approaches using K-means clustering, a novel and unique approach within this domain. Beyond a literature review that delved into prior research, the groundwork for the study was established by exploring the intricate mathematical underpinnings of Kriging, Co-Kriging, and LR-MFS. the performance of Co-Kriging and LR-MFS approaches were examined in one-dimensional and two-dimensional cases, specifically investigating the results obtained from the Co-Kriging-LR-MFS ensemble model. This innovative model has the strengths of both approaches, showing promising potential.

The research demonstrated that the selection between Co-Kriging and linear models is dependent on the specific characteristics of the system. For instance, in cases like the Laplace's example, where the relationships between high- and low-fidelity data follow a linear pattern, the linear model exhibited exceptional performance. Conversely, Co-Kriging stood out due to its utilization of interpolation and Gaussian processes, making it adept at capturing nonlinear relations and local trends. By integrating these two methodologies through ensemble modeling, the synergistic effects led to a significant reduction in RMSE. Moreover, considering these disparities are region-specific, employing K-means clustering proved effective in identifying the optimal model for each region. Notably, the new approach offered several advantages:

- Utilizing low-fidelity data from various locations alongside high-fidelity data, which is an advantage of the Co-Kriging method.
- Ability to capture both linear and nonlinear trends in the function.
- Efficiency in dealing with high-order datasets, optimizing computational time by utilizing data points for model evaluation rather than training, which typically demands substantial computational resources.

An opportunity for further exploration involves refining the K-means clustering aspect. Developing algorithms or codes that can automatically adjust the initial regions could significantly improve the overall performance of the Co-Kriging- LR-MFS ensemble modeling approach. Since the selection of the model in each region markedly impacts the model's overall performance, optimizing the regioning process could enable more efficient utilization of this methodology, and could lead to using this algorithm for higher-dimensional problems easily.

APPENDICES

Appendix A: Stochastic process explanation

A random process, also referred to as a stochastic process, is a mathematical concept utilized for modeling and explaining the progression of variables that are subject to randomness over time or space. It serves as a fundamental tool in probability theory and statistics and finds applications across various disciplines such as finance, physics, engineering, among others. Stochastic processes can be broadly classified into two categories: discrete-time and continuous-time processes. Let's delve into an overview of these stochastic processes.

A.1. Discrete-Time Stochastic Process

In discrete-time processes, random variables are observed at specific points in time or space rather than continuously. These observations are typically denoted as $\{x_1, x_2, x_3, \dots\}$, where each x_t represents a distinct random variable. For instance, consider a simple example involving coin flips where each x_t signifies the outcome (heads or tails) at time t .

A.2. Continuous-Time Stochastic Process

Contrarily, continuous-time processes involve observing random variables over an uninterrupted interval of time or space. These observations are commonly represented as $\{x(t): t \geq 0\}$, with $x(t)$ representing the value of the random variable at time t . To illustrate this point further, let's examine the movement of stock prices where $x(t)$ denotes the price of a particular stock at time t .

A.3. Stochastic process properties

A.3.1. State Space

Stochastic processes operate within what is known as a state space - encompassing all possible values that a given random variable can assume. The state space may adopt either a discrete or continuous nature or even incorporate elements from both realms.

A.3.2. Trajectories or Paths

The term trajectory or path refers to the specific realization of a stochastic process; it entails the sequence of values assumed by the associated random variables over time or space. Trajectories can be perceived as sample paths drawn from the underlying stochastic process.

A.3.3. Markov Property

Many stochastic processes exhibit what is termed as the Markov property - indicating that future behavior solely relies on the present state and not the entire historical context. This property simplifies both modeling and analysis procedures for such processes.

A.3.4. Stationarity

Certain stochastic processes possess stationarity, which implies that their statistical properties (e.g., mean, variance, correlation) remain constant across time. This characteristic facilitates analysis by enabling a focus on the invariant statistical properties inherent to the process.

Appendix B: Spatial Correlation explanation

Spatial correlation, also known as spatial autocorrelation, is a statistical concept used to describe the degree to which data points or observations in a geographical or spatial context are related to or dependent on one another based on their spatial locations. It is particularly relevant in fields such as geography, geostatistics, ecology, environmental science, and geosciences. Spatial correlation refers to the idea that nearby locations tend to have similar values or exhibit similar characteristics, while locations that are farther apart may be less similar or even unrelated.

B.1. Three forms of Spatial Correlation

B.1.1. Positive Spatial Correlation

This occurs when nearby observations are more similar than what would be expected by random chance. In other words, locations close to each other tend to have similar values. For example, in a temperature dataset, if neighboring weather stations record similar temperatures, it indicates positive spatial correlation.

B.1.2. Negative Spatial Correlation

Negative spatial correlation, or spatial anticorrelation, occurs when nearby observations are less similar than expected by random chance. In such cases, locations close to each other tend to have dissimilar values. An example might be, soil nutrient levels, where neighboring soil samples have contrasting nutrient concentrations.

B.1.3. No Spatial Correlation

In some cases, there may be no discernible spatial correlation, and observations are spatially independent. This means that the values at one location do not depend on the values at nearby locations. In practice, this is relatively rare, and some level of spatial correlation is often observed.

B.2. Spatial Correlation importance

B.2.1. Predictive model improvement

It helps in spatial data analysis, geostatistics, and spatial modeling, allowing for the development of predictive models that consider the spatial structure of the data.

B.2.2. Environmental Applications

It can provide insights into underlying spatial processes or environmental factors. For example, it can reveal patterns in climate, geology, land use, or other spatially related phenomena.

B.2.3. Kriging Algorithm

It has practical applications in fields such as spatial interpolation (e.g., Kriging), where it is used to estimate values at unsampled locations based on the values of nearby observations.

Analyzing and quantifying spatial correlation typically involves statistical techniques and tools, such as variograms, correlograms, semi-variograms, and spatial autocorrelation measures. These methods help researchers assess the strength and direction of spatial relationships, which is crucial in spatial data analysis and modeling.

REFERENCES

- Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., & Newman, P. A. (2001). *Journal of Aircraft*, 38(6), 1093-1101.
- Balabanov, V., Grossman, B., Watson, L., Mason, W., & Haftka, R. (1998, September). In 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization (p. 4804).
- Bect, J., Ginsbourger, D., Li, L., Picheny, V., & Vazquez, E. (2012). *Statistics and Computing*, 22, 773-793.
- Berger, J. O., De Oliveira, V., & Sansó, B. (2001). *Journal of the American Statistical Association*, 96(456), 1361-1374.
- Burges, C. J. (1998). *Data mining and knowledge discovery*, 2(2), 121-167.
- Cheng, K., Lu, Z., & Zhen, Y. (2019). *Computer Methods in Applied Mechanics and Engineering*, 349, 360-377.
- Chiles, J. P., & Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty* (Vol. 713). John Wiley & Sons.
- Choi, S., Alonso, J. J., Kroo, I. M., & Wintzer, M. (2008). *Journal of Aircraft*, 45(1), 106-118.
- Conti, S., & O'Hagan, A. (2010). *Journal of statistical planning and inference*, 140(3), 640-651.
- Cressie, N. (1988). *Mathematical geology*, 20, 405-421.
- Currin, C., Mitchell, T., Morris, M., & Ylvisaker, D. (1991). *Journal of the American Statistical Association*, 86(416), 953-963.
- Fang, K. T., Li, R., & Sudjianto, A. (2005). CRC press.
- Forrester, A. I., Bressloff, N. W., & Keane, A. J. (2006). *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 462(2071), 2177-2204.
- Forrester, A. I., Sobester, A., & Keane, A. J. (2007). *Proceedings of the royal society a: mathematical, physical and engineering sciences*, 463(2088), 3251-3269.
- Forrester, A.I.J., Sobester, A., and Keane, A.J., Engineering Design via Surrogate Modeling: A Practical Guide, AIAA, 2008
- Giunta, A., Narducci, R., Burgee, S., Grossman, B., Mason, W., Watson, L., & Haftka, R. (1995). In 13th Applied Aerodynamics Conference (p. 1886).
- Gneiting, T., Kleiber, W., & Schlather, M. (2010). *Journal of the American Statistical Association*, 105(491), 1167-1177.

- Goh, J., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuran, C. C., & Rutter, E. (2013). *Technometrics*, 55(4), 501-512.
- Gramacy, R. B., & Lian, H. (2012). *Technometrics*, 54(1), 30-41.
- Gutmann, H. M. (2001). *Journal of global optimization*, 19(3), 201-227.
- Huang, D., Allen, T. T., Notz, W. I., & Miller, R. A. (2006). *Structural and Multidisciplinary Optimization*, 32, 369-382.
- Kalivarapu, V., & Winer, E. (2008). *Environmental Modelling & Software*, 23(12), 1370-1383.
- Kennedy, M. C., & O'Hagan, A. (2000). *Biometrika*, 87(1), 1-13.
- Kennedy, M. C., & O'Hagan, A. (2001). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3), 425-464.
- Koehler, J. R., & Owen, A. B. (1996). 9 Computer experiments. *Handbook of statistics*, 13, 261-308.
- Krige, D. G. (1951). *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6), 119-139.
- Le Gratiet, L. (2013). *SIAM/ASA Journal on Uncertainty Quantification*, 1(1), 244-269.
- Liu, Y., Wang, S., Zhou, Q., Lv, L., Sun, W., & Song, X. (2022). *Chinese Journal of Mechanical Engineering*, 35(1), 1-15.
- Matheron, G. (1963). *Principles of geostatistics. Economic geology*, 58(8), 1246-1266.
- Matheron, G. (1969). *Le krigeage universel (Vol. 1)*. Paris: École nationale supérieure des mines de Paris.
- Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2016). John Wiley & Sons.
- Na, S., Xumin, L., & Yong, G. (2010, April). In 2010 Third International Symposium on intelligent information technology and security informatics (pp. 63-67). Ieee.
- O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91(10-11), 1290-1300.
- Qian, P. Z., & Wu, C. J. (2008). *Technometrics*, 50(2), 192-204.
- Rahman, F., Posnett, D., Herraiz, I., & Devanbu, P. (2013, August). In Proceedings of the 2013 9th joint meeting on foundations of software engineering (pp. 147-157).
- Rumpfkeil, M. P., Bryson, D. E., & Beran, P. S. (2019). In AIAA Scitech 2019 forum (p. 1998).
- Rumpfkeil, M. P., Bryson, D., & Beran, P. (2022). *Algorithms*, 15(3), 101.

- Song, Y., Cheng, Q. S., & Koziel, S. (2019). *Sensors*, 19(13), 3023.
- Spiess, A. N., & Neumeyer, N. (2010). *BMC pharmacology*, 10(1), 1-11.
- Sun, G., Li, G., Gong, Z., He, G., & Li, Q. (2011). *Engineering Optimization*, 43(12), 1351-1366.
- Wackernagel, H., & Wackernagel, H. (2003). Ordinary Kriging. *Multivariate geostatistics: an introduction with applications*, 79-88.
- Xiao, M., Zhang, G., Breitzkopf, P., Villon, P., & Zhang, W. (2018). *Applied Mathematics and Computation*, 323, 120-131.
- Yong, H. K., Wang, L., Toal, D. J., Keane, A. J., & Stanley, F. (2019). *Structural and Multidisciplinary Optimization*, 60, 1209-1226.
- Zahir, M. K., & Gao, Z. (2013). *Chinese Journal of Aeronautics*, 26(4), 841-849.
- Zhang, Y., Kim, N. H., Park, C., & Haftka, R. T. (2018). *AIAA Journal*, 56(12), 4944-4952.

BIOGRAPHY OF THE AUTHOR

Yasaman Asiaee was born in Tehran, Iran, in 1996. She earned her Bachelor's degree in Materials Science and Engineering from K. N. Toosi University of Technology of Tehran in 2020. Yasaman is a candidate for the Master of Science degree in Mechanical Engineering from the University of Maine in May 2024.