Advanced Machine Learning for Surrogate Modeling in Complex Engineering Systems

Cheolhei Lee

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Industrial and Systems Engineering

Xiaowei Yue, Chair
Xi Chen
Inyoung Kim
Zhenyu (James) Kong

May 30, 2023
Blacksburg, Virginia

Keywords: Machine Learning, Active Learning, Extreme Spatial Model, Surrogate Model

Advanced Machine Learning for Surrogate Modeling in Complex Engineering Systems

Cheolhei Lee

(ABSTRACT)

Surrogate models are indispensable in the analysis of engineering systems. The quality of surrogate models is determined by the data quality and the model class but achieving a high standard of them is challenging in complex engineering systems. Heterogeneity, implicit constraints, and extreme events are typical examples of the factors that complicate systems, yet they have been underestimated or disregarded in machine learning. This dissertation is dedicated to tackling the challenges in surrogate modeling of complex engineering systems by developing the following machine learning methodologies. (i) Partitioned active learning partitions the design space according to heterogeneity in response features, thereby exploiting localized models to measure the informativeness of unlabeled data. (ii) For the systems with implicit constraints, failure-averse active learning incorporates constraint outputs to estimate the safe region and avoid undesirable failures in learning the target function. (iii) The multi-output extreme spatial learning enables modeling and simulating extreme events in composite fuselage assembly. The proposed methods were applied to real-world case studies and outperformed benchmark methods.

Advanced Machine Learning for Surrogate Modeling in Complex Engineering Systems

Cheolhei Lee

(GENERAL AUDIENCE ABSTRACT)

Data-driven decisions are ubiquitous in the engineering domain, in which data-driven models are fundamental. Active learning is a subdomain in machine learning that enables data-efficient modeling, and extreme spatial modeling is suitable for analyzing rare events. Although they are superb techniques for data-driven modeling, existing methods thereof cannot effectively address modern engineering systems complicated by heterogeneity, implicit constraints, and rare events. This dissertation is dedicated to advancing active learning and extreme spatial modeling for complex engineering systems by proposing three methodologies. The first method is partitioned active learning that efficiently learns systems, changing their behaviors, by localizing the information measurement. Second, failure-averse active learning is established to learn systems subject to implicit constraints, which cannot be analytically solved, and to minimize constraint violations. Lastly, the multi-output extreme spatial model is developed to model and simulate rare events that are associated with extremely large values in the aircraft manufacturing system. The proposed methods overcome the limitations of existing methods and outperform benchmark methods in the case studies.

Dedication

To my beloved family

Acknowledgments

Most of all, I would like to give my best gratitude to my advisor, Dr. Xiaowei Yue, for his dedicated supervision, indescribable support, and patience in my Ph.D. studies. Without his guidance, I would not be able to complete this journey. He demonstrated a great example of not only a researcher but also a leader, and the exemplars he has shown me will guide and influence my life ever after. I also would like to express my thanks to my Ph.D. committee members, Dr. Xi Chen, Dr. Inyoung Kim, and Dr. Zhenyu (James) Kong, for their insightful advice, support, and consideration. Their guidance has considerably broadened my outlook and enriched my study.

I also appreciate the Republic of Korea Navy for providing me with such a great opportunity to pursue my Ph.D. study and caring for my family in great circumstances. I thank my collaborators, Dr. Wenjun Cai, Dr. Kaiwen Wang, Dr. Wenjia Wang, Dr. Xing Wang, and Dr. Jianguo Wu for their comments, advice, and assistance that made our work publishable. I would also like to thank my fellows, Dr. Areej Albahar, Mr. Oliver Tim Lutz, and Dr. Yinan Wang.

I would like to give my gratitude to my wife for showing the deepest love, faith, and patience. Without her devotion to our family, I would not be able to keep this journey until the last. I also thank my beloved kids, Jeongyeon and Jaeyoon, who have been surprisingly encouraging and inspiring me. Lastly, I would also like to thank my parents, my parents-in-law, and my sister for their support.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Data-driven models such as surrogate models and predictive models are indispensable tools in the analysis of complex engineering systems. Especially, data-driven models pay off when the systems, on which the surrogates are established, are demanding to run. Sometimes, data-driven models can provide promising solutions that are unattainable from first principles models. Although data-driven models are equipped with high potential, they may be misleading by the poor quality of data or by choosing inappropriate model classes. There are subfields in machine learning that dedicate to improving models' quality and relevant model selection. Active learning is a machine learning framework in which learning agents intellectually query unlabelled data to learn, thereby achieving desired model quality with an acceptable data acquisition cost. The feasibility of uncertainty quantification (UQ) in machine learning is attributed to the advance of statistics, algebra, and computing power. Especially, diverse distribution families in statistics are promising apparatuses for stochastic system modeling in a data-driven manner.

However, the increasing complexity of modern engineering systems is making the achievement of any of the use of first principles models, collecting enough high-quality data, and choosing an appropriate model class challenging with machine learning techniques. The characteristics that make engineering systems difficult to analyze with their data are vari-

ous, but heterogeneity, implicit constraints, and extreme events are typical examples. Since the terminologies of the examples may have different meanings in other fields, we clarify them as follows.

- *Heterogeneity* mainly arises when systems comprise heterogeneous subsystems or when system properties vary with control and design variables. An example exhibiting heterogeneity is the phase transitions of water with different temperatures and pressures. Heterogeneity intricates data-driven modeling of the response surface with the conventional design of experiments (DOE) or active learning strategies.

- *Implicit constraints* stand for constraining (sub-) systems without analytic mathematical forms. They constrain the systems since some design or control settings violating them bring about fatal system failures or they are infeasible. The absent analytic form compels computationally expensive numerical methods to identify the desired design space, where the data-driven model should be built.

- *Extreme events* are extraordinary events associated with extreme values in stochastic systems. They are realized from the tails of heavy-tailed distributions and are of interest when they are crucial in system management. Extreme data is rare and necessitates extreme value modeling that is different from ordinary models, while extreme value models have been underutilized in the engineering domain.

Although the aforementioned characteristics are common in engineering systems, they have been frequently underestimated or ignored in machine learning. Ignoring them in data-driven modeling may lead to adopting erroneous models and catastrophic consequences in engineering system analyses (e.g., system failures and suboptimal decisions). Otherwise, it requires additional efforts to address them. Motivated by the background, this dissertation

is dedicated to developing machine learning methods for data-driven modeling in complex engineering systems.

1.2 Research Objectives

The objective of this research is to develop machine learning methodologies for data-driven modeling in engineering systems with the aforementioned characteristics. The detailed objectives with the chapter order are as follows.

1. We develop adaptive active learning that detects heterogeneity in systems so that a sampling budget can be efficiently utilized.
2. Failure-averse active learning strategy is established to consider implicit constraints in active learning or adaptive sampling for engineering systems subject to failures.
3. We develop a multi-output extreme spatial model for composite fuselage assembly that is capable of modeling correlated key characteristics.

1.3 Organization

This dissertation consists of a literature review, three methodologies, and a concluding section as illustrated in Fig. 1.1. In Chapter 2, we review seminal works in active learning for Gaussian processes (GPs), and recent literature on active learning with heterogeneity and constraints. Then, extreme value models in the engineering domain and state-of-the-art extreme spatial models are reviewed. The subsequent three chapters address the objectives presented in section 1.2. In chapter 3, we propose partitioned active learning that considers possible heterogeneity in the system. It is mainly motivated by the 'divide and conquer'

paradigm in statistical learning, so it quantifies the uncertainty of unobserved designs by partitioning the design space based on the estimated heterogeneity in the target system. Chapter 4 proposes failure-averse active learning, whose main objective is to optimize the design avoiding the occurrence of unknown failure in the system. In chapter 5, the extreme state-space modeling for multistage systems is proposed. The proposed method employs the latent variable model for the spatial extreme surface and characterizes the propagation of extreme values in the multistage system. Lastly, Chapter 6 summarizes contributions and provides future research topics.

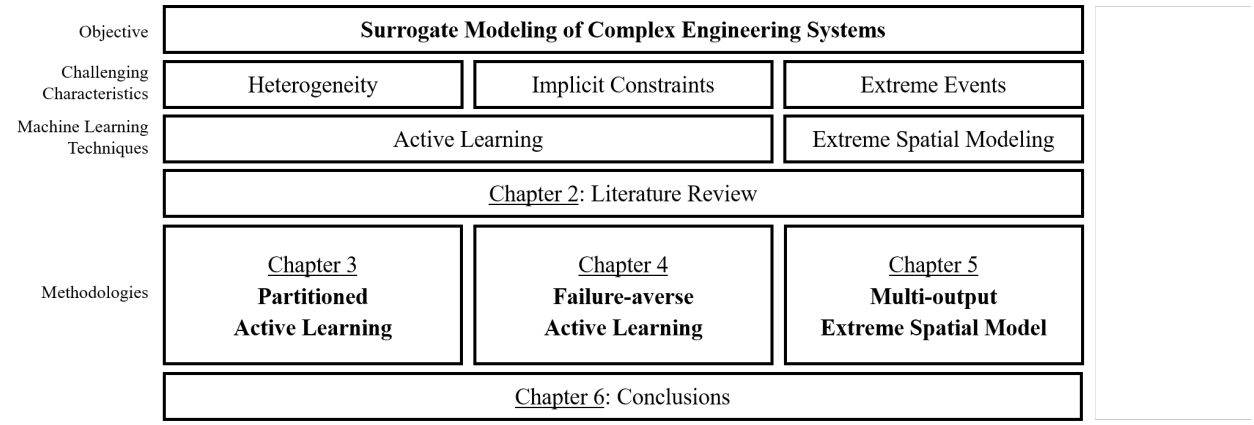

Figure 1.1: Dissertation outline.

Chapter 2

Literature Review

In this section, we review the literature on active learning and extreme spatial models. First, we review recent applications of active learning in the engineering domain, and then we investigate existing methods related to heterogeneity and implicit constraints. Some reviewed references may be focused on methodology development rather than applications. Afterward, we review extreme value models for extreme events. Since the up-to-date extreme value modeling in the engineering domain has not been developed for spatial modeling, we review state-of-the-art extreme spatial models by extending our purview to the outside of the engineering domain.

2.1 Active Learning in the Engineering Domain

Active learning [1] needs an information criterion that quantifies the informativeness of unlabeled data. The information criterion is also called the information measure or the acquisition function in different areas. A majority of existing strategies is to refer to uncertainty quantified by models in the form of variance, entropy, discrepancy, etc. In the engineering domain, active learning is mainly utilized for surrogate modeling of expensive-to-evaluate systems. In this case, active learning devotes to querying design points that minimize the generalization error of the model. Computational physics is an area where active learning is frequently used since many physics-based models are costly or have no analytic forms. Generally, predic-

tive variance or parameter variance is utilized in active learning for surrogate models of the response surface method [2], deep neural networks (DNNs) [3, 4, 5]. Especially, Lye et al. [5] assumed that the feasible region is known a priori to ensure feasibility in modeling the partial derivative equation (PDE) solver.

Unlike DNNs, utilizing uncertainty is more straightforward in GPs, so there are a number of papers associated with active learning for GPs. Yang et al. [6] proposed a physics-informed GP for the stochastic PDE simulator. The GP model is informed by replicated observations of the PDE simulator, and predictive variance is referred to in active learning. Chen et al. [7] developed the GP that incorporates linear and nonlinear PDE information. They involved the active learning strategy to determine PDE points for their GP model by employing the integrated mean squared error (IMSE) criterion. Yue et al. [8] proposed the variance-based and the Fisher information criteria for GPs considering uncertainty and applied them to the predictive modeling of composite fuselage deformation.

The recent applications of active learning in the engineering domain imply that not only active learning but also machine learning should be coupled with engineering background knowledge. This paradigm has got increasing attention recently in the computational physics domain in the name of physics-guided or physics-informed machine learning. However, most existing active learning strategies for GPs in engineering applications do not consider heterogeneity and constraints in their exploration, which are also a part of the underlying system.

2.2 Active Learning for Systems with Heterogeneity and Implicit Constraints

2.2.1 Heterogeneous Systems

Although heterogeneity is ubiquitous in contemporary engineering systems, there have been few efforts to actively address heterogeneity in active learning. Meanwhile, heterogeneity has been frequently covered in modeling. Partitioning the design space to resolve heterogeneity is a straightforward and effective technique in spatial modeling. Especially, it brings many benefits to GPs such as improving scalability and facilitating nonstationary modeling [9, 10, 11]. We refer to GPs with partitioning as partitioned GPs (PGPs) and discuss more on the partitioning methods in Section 3.2.3.

Another advantage that we can take from PGPs is the various active learning strategies in GPs. The most widely used criteria are the active learning Mackay (ALM) and the active learning Cohn (ALC) [12]. The ALM refers to the variance as the information criterion that selects the most uncertain data, while, the ALC refers to the IMSE criterion that seeks a point expected to reduce the model variance the most over the design space. Other than uncertainty-based strategies, the Kernel Hilbert space-filling design [13], the expected model output change [14], gradients [15, 16], and the discrepancy in the ensemble model [17] are considered. However, the aforementioned active learning strategies are mostly devised for single GPs (SGPs) that are inappropriate for modeling heterogeneous systems. Therefore, it is inefficient to directly apply them to PGPs, since they do not consider the partitioned structure. Pope et al. [18] proposed a PGP with active learning that takes a point in boundaries and maximizes the space-filling property of design points. However, it focuses more on detecting discontinuity in the design space rather than reducing the prediction error, although

such discontinuity is rare in engineering systems. Active learning algorithms considering the posterior structure of partitioned design space were proposed in [19, 20]. However, their approaches are highly dependent on the tree classifier, while the tree partitioning mostly induces boundaries parallel to axes that may not be realistic in practice [11, 18]. Moreover, the choice of design point candidates is dependent on the areas of partitioned regions, so it can be irrelevant when a plausible partition is not realizable with a few tree-partitioned regions.

2.2.2 Implicit Constraints

Constraints are common in optimization problems and implicit constraints make identifying the feasible set an arduous task. There are mainly two distinct manners in considering implicit constraints in sequential sampling, and the most appropriate scenario for engineering problems is that any evaluation in the infeasible design space should be avoided. We refer to this type as safe exploration since any design point violating the constraint invokes undesirable system failures. Schreiter et al. [21] proposed active learning for GPs that used the nuisance function of GP classification to discriminate the safe region from the unsafe region. They used the lower confidence interval to ensure safety during the maximum entropy-based active learning. However, their approach only focuses on the entropy of the target function, so it may not efficiently expand the explorable safe region. Consequently, it may lead to conservative learning due to the insufficiently revealed safe region and poor quality of the target model over the unexplored safe region. Furthermore, the nuisance function of GP classification needs some attention in its application. First, the nuisance function refers to binary labels, so it may distort the numerical information from the constraint observations that are usually related to closeness to the boundary of the safe region. Second, when the dataset contains only either safe or unsafe samples, it fails to provide promising discrimina-

tion. Turchetta et al. [22] suggested safe exploration for interactive machine learning that can be adopted for active learning. They conduct safe region expansion when the most informative data is not located in the current safe region. However, their safe region expansion aims to validate the safety of unobserved data, so it is inefficient to expand the explorable space to the true safe region which should be maximized to reduce the risk of the model.

Another way to address implicit constraints is that the feasibility can be disregarded in the middle of the trajectory in sequential sampling. This type is usually found in constrained Bayesian optimization, which is a sequential design strategy for the global optimization of black-box functions with constraints. First of all, it should be noted that active learning and Bayesian optimization have different aims. Active learning focuses on summarizing unobserved samples to improve the model's quality over the input space, while Bayesian optimization aims to find the optimal location over the input space by evaluating samples directly associated with improvement in the objective function. However, techniques considering constraints in Bayesian optimization are worthwhile to investigate since the techniques are also valid in active learning.

Many constrained Bayesian optimization methods [23, 24, 25, 26, 27] follow the framework of [28], which multiplies the probability of feasibility to the expected improvement (EI) function. However, multiplying the feasibility probability only aids in leading their solutions to be feasible, so it is inefficient to expand the feasible region. Gramacy and Lee [29] deemed infeasible samples also can be informative, so they proposed the integrated expected conditional improvement that is also weighted by expected feasibility. Hernandez-Lobato et al. [30] proposed the predictive entropy search with constraints that refers to the expected entropy reduction at the minimum associated with observations from the objective function and constraints. It automatically focuses on objectives and constraints by integrating them into entropy. Meanwhile, Sui et al. [31] proposed a constrained Bayesian optimization algorithm

devoted to both optimization and safe region expansion. They expanded the safe region in the first phase and then implemented typical Bayesian optimization within the disclosed safe region. However, their safe expansion may be subject to slow convergence to the ground truth unless the Lipschitz constants of constraint functions are known beforehand.

2.3 Extreme Spatial Models

In this section, we review recent literature on extreme value models in the engineering and management sciences. Extreme values in engineering systems are usually associated with undesirable abnormal events, so they are mainly considered for system monitoring, failure prediction, and outlier detection. Cho et al. [32] determined thresholds for anomaly detection with extreme value distributions. Yu et al. [33] used the extreme value model for outlier detection in open-set fault diagnosis and applied their method to the diagnosis of the rotating motor. Some other works used extreme models for the predictive modeling of intractable engineering systems. Gu et al. [34] predicted the maximum fatigue indicator parameter in metal alloys of which micro-scale defects are rare. In additive manufacturing, Boyce et al. [35] used the extreme model to predict rare failures, which may cause significant loss of quality. Yu et al. [36] utilized the extreme model for forecasting multiple failure modes in time-variant reliability problems to reduce integral errors. The proposed method is applied to the assessment of the exoskeleton under uncertainty. Liu and Meeker [37] modeled the thickness of the pipeline for inspecting possible leakage therein. Although extreme value models in the aforementioned works are adopted considering the characteristics of rare events, all of them are restricted to modeling univariate cases that are much simpler than multivariate and spatial extreme values.

Another area where the extreme values occasionally appear is reliability analysis such as

time-to-failure (TTF) prediction. Xia et al. [38] used the extreme value distribution in the regression model of TTF for multimodal manufacturing machines. Lewis-Beck et al. [39] used the Weibull distribution for the within-sample prediction of TTF in heterogeneous failure modes. King et al. [40] used the Weibull distribution for the fatigue time of polymer composites and proposed the optimal design for variance-minimizing estimation. He et al. [41] used the extreme value distribution for modeling of first-failure-time given manufactured time. However, their methodologies ignore temporal or spatial dependency so they utilize optimal design strategies to compromise the sampling cost. For cyber security, Zhan et al. [42] used the nonstationary generalized Pareto distribution with time-varying parameters for the long-term prediction of extreme cyber attacks. However, their problem did not involve spatial dependency, so theirs cannot be directly applied to the multi-output systems. Grigoriu [43] proposed the surrogate-assisted spatial extreme response model. They used a surrogate model with bounded noise for stochastic PDE and used the noisy samples for extreme response analysis. However, this method only utilizes the extreme value distribution for point-wise characterization without considering spatial extreme dependency.

There are extensive works on extreme spatial models in other applications. Davison et al. [44] concisely illustrated common spatial extreme models that are widely used and compared them in the rainfall case study. The most active research areas in extreme spatial modeling can be summarized as (i) flexibility; (ii) scalability; and (iii) simulation. The flexibility is associated with the asymptotic dependence between extreme events over a spatial or temporal domain. Since the extreme spatial models represented by max-stable processes are asymptotic dependent, spatial models with asymptotic independence have got lots of attention recently [45, 46, 47]. The scalability comes from that many extreme spatial models are subject to the prohibitive computational cost of the full likelihood estimation because of the intractable density computation. Castruccio et al. [48] showed compared performances of

full likelihoods and composite likelihoods with different orders in relatively low dimensional ($\leq$ 20) problems and provided practical suggestions. Engelke and Hitz [49] proposed the graphical models for multivariate Pareto distributions that allow efficient inference with sparse covariance matrices. However, their model is developed for peak-over-threshold, which is not considered in this dissertation. Although numerous works have been done in this field, extreme models thereof lack control variables and scalability so they do not befit the composite fuselage assembly.

Chapter 3

Partitioned Active Learning for Heterogeneous Systems

Active learning is a subfield of machine learning that focuses on improving the data collection efficiency in expensive-to-evaluate systems. Active learning applied surrogate modeling facilitates the cost-efficient analysis of demanding engineering systems, while the existence of heterogeneity in underlying systems may adversely affect the performance of active learning. In order to improve the learning efficiency under this regime, we propose partitioned active learning that seeks the most informative design points for heterogeneous systems. The proposed active learning consists of two systematic subsequent steps: the global searching scheme accelerates the exploration of active learning by investigating the most uncertain design space, and the local searching exploits the circumscribed information induced by the local GP. We also propose Cholesky update-driven numerical remedies for our active learning to address the computational complexity challenge. The proposed method is applied to numerical simulations and two real-world case studies: (i) the cost-efficient automatic fuselage shape control in aerospace manufacturing; and (ii) the optimal design of tribocorrosion-resistant alloys in materials science. The results show that our approach outperforms benchmark methods with respect to prediction accuracy and computational efficiency.

3.1 Introduction

Active learning is a subfield of machine learning and artificial intelligence that maximizes information acquisition to data efficiently train models. Contrary to passive learning such as Latin Hypercube design (LHD) and factorial design [50], active learning sequentially selects design points in the modeling phase after observing intermediate models and outputs. It is also called query learning, sequential design, adaptive sampling, or optimal design in other literature, while they pursue the same objective: finding the best subset of inputs from the design space according to information criteria that evaluate the informativeness of input referring to uncertainty, disagreement, etc. Active learning has received increasing attention from various applications in which sampling is timely and costly demanding, such as quality engineering, response surface investigation, and image recognition [2, 51].

Especially, active learning has been frequently utilized with GPs for modeling various systems spanning robotics, aerospace, and manufacturing processes [8, 52] due to the capability of UQ and the simplicity [53, 54]. However, many existing methods are confined to SGPs that impose inappropriate homogeneous information measures for systems with heterogeneity (e.g., discontinuity, and abrupt variations in gradient norms or frequencies), although heterogeneity is ubiquitous in engineering systems. For example, composite materials, one of the most versatile materials in various contemporary products, are anisotropic and highly nonlinear to external treatments [55], so they exhibit different behaviors in the design space [56]. Another example is the corrosion of alloys. Fig. 3.1 illustrates the corrosive rates of aluminum alloys emulated with the finite element method (FEM) over two pairs of control variables. We can observe that the response surface shows spatial heterogeneity, so the design space can be partitioned into three subregions according to degrees of variation. In both cases, the efficiency of active learning with a homogeneous information criterion can be significantly deteriorated by the misleadingly measured information.

A straightforward way to address heterogeneity in system analysis is to divide the design space into homogeneous subregions and use localized information such as 'divide-and-conquer'. For GPs, PGPs overcome the limitations of SGPs in heterogeneous systems by allocating multiple independent local GPs on disjoint subregions [9, 11, 18]. Subregions are defined or estimated according to distinguishable characteristics of target systems, so PGPs can efficiently accommodate heterogeneity. Moreover, partitioning improves the scalability of GPs, one of their main drawbacks, by introducing sparsity in their covariance matrices. Although it is possible to use PGPs for active learning, there are several issues that require attention. First, partitioning in PGPs is mainly driven by the hierarchical model fit, thereby inducing implausible partitions (e.g., too many subregions, generating trivial subregions). Second, many PGPs adopt Markov chain Monte Carlo (MCMC) methods to train the model, which may cause prohibitive time for partitioning. Additionally, existing active learning strategies are suboptimal for PGPs in terms of learning and computational efficiency, since they are originally devised with SGPs. Therefore, in order to consider heterogeneity in a more reasonable manner, partitioning should not be determined by local models, but by heterogeneous characteristics.

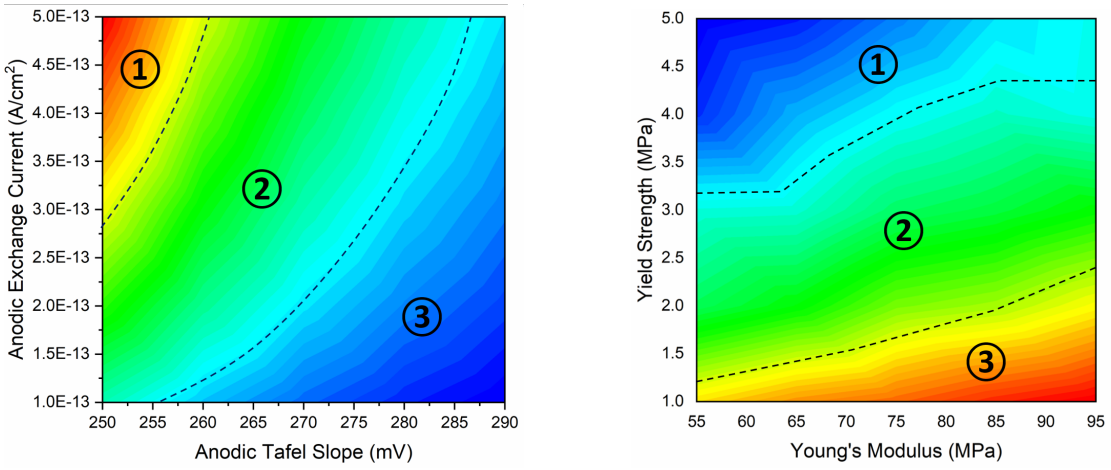

Figure 3.1: Corrosive rates of alloys with different control variables. (Left) Tafel slope and anodic exchange current. (Right) Young's modulus and yield strength.

Motivated by the aforementioned limitations, we propose an active learning method for data acquisition and modeling of heterogeneous systems. The proposed method seeks the most uncertain subregion in the design space with the global searching scheme, and the localized IMSE criterion is subsequently utilized to select the most informative design point within the subregion. Our contributions in this chapter are the following.

1. A heterogeneity-based partitioning method is developed, which tackles the limitations of existing partitioning methods in PGPs. The developed partitioning method refers to heterogeneous features and can provide the number of partitions without prior knowledge. Moreover, the partitioning performed well in a twenty-one-dimensional problem, in which Voronoi tessellation schemes could not, with a single parameter.

2. A novel active learning strategy with a two-step searching that exploits the partitioned structure is established. The strategy first searches for the region, which has the most potential to decrease the model variance, via global searching, and then the localized IMSE criterion is used to find the query location as local searching.

3. Numerical remedies are provided to accelerate the proposed algorithm. Global searching is used to reduce the number of candidates by narrowing down the area for local searching, and the Cholesky factor update method is used to reduce the cost of local searching.

The remainder of this chapter is organized as follows. Section 3.2 elucidates the new partitioned active learning algorithm and provides applicable techniques for improving the learning efficiency and numerical costs. In Section 3.3, we implement our method on function approximation problems and apply it to three real-world problems in Section 3.4 with existing learning algorithms. Finally, a brief summary of this chapter is provided in Section 3.5.

3.2 Methodology

In this section, we propose the partitioned active learning algorithm. First, we briefly review a generic framework of partitioned modeling and two widely used active learning strategies for GPs. Then, we elucidate our strategy including partitioning based on heterogeneity and the partitioned information criterion. Lastly, we provide applicable remedies to improve the computational cost of the proposed method.

As mathematical notations, we use lower letters for scalars and distinguish vectors with boldface. Upper letters indicate sets or matrices, and a set of indices are denoted as $[M] = \{1, \ldots, M\}$. In subscriptions, we parenthesize the number for the region index and use normal letters for indices of data. If an index of a subregion is apparent, we omit the region index and only use the index in the set for simplicity.

3.2.1 Partitioned Modeling for Heterogeneous Systems

A partitioned model of a heterogeneous system can be expressed as a function f defined over the design space $\Omega \subset \mathbb{R}^d$ mapping to $\mathbb{R}$. The partitioned model employs a region classifier $g : \Omega \to [M]$ that partitions Ω into M mutually disjoint subregions such that $\Omega = \bigcup_{m=1}^{M} \Omega_{(m)}$ in accordance with heterogeneity. Then, the partitioned model can be written as

$$f(\mathbf{x};\ g) = \sum_{m=1}^{M} 1_{\{g(\mathbf{x})=m\}} f_{(m)}(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{3.1}$$

where $1_{\{C\}}$ is an indicator function that has value 1 when C is true and 0 otherwise, and $f_{(m)}$ is a local GP assigned to $\Omega_{(m)}$. Although PGPs can take any valid kernel, we mainly consider stationary kernel families such as radial basis function (RBF) and Matérn. Suppose any $\mathbf{x}, \mathbf{x}' \in \Omega_{(m)}$, and $\mathbf{x} = [\, x_1 \ \cdots \ x_d \,]^\top$. The local GP defined over $\Omega_{(m)}$ with the RBF

kernel is

$$f_{(m)}(\mathbf{x}) \sim \mathcal{GP}\left(\mu_{(m)}(\mathbf{x}),\ k_{(m)}(\mathbf{x},\, \mathbf{x}')\right),$$
$$k_{(m)}(\mathbf{x},\, \mathbf{x}') = \tau^2_{(m)} \prod_{j=1}^{d} \exp\left(-\frac{(x_j - x'_j)^2}{l^2_{(m),j}}\right) + \sigma^2_{(m)} 1_{\{\mathbf{x}=\mathbf{x}'\}},$$

where $\mu_{(m)}(\mathbf{x})$ is the mean function assumed to be zero without loss of generality, and $k_{(m)}$ is the kernel of which nonnegative $\tau^2_{(m)}$, $l^2_{(m),j}$ and $\sigma^2_{(m)}$ are referred as scale, length and noise hyperparameters.

Suppose we have finite n observations on $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \Omega$, and $D = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}$, where $y_i = f(\mathbf{x}_i) + \epsilon_i$ of which $\epsilon_i \sim \mathcal{N}(0,\ \sigma^2)$. Let Φ be the hyperparameter of the region classifier, and Θ be that of local GPs. There are mainly two schemes in the hyperparameter estimation for PGPs. One is to maximize the likelihood of the hierarchical structure of Eq. (3.1), and the other way is to estimate the region classifier in advance based on some criteria and fit local models by maximizing the likelihood with the fixed region classifier. We will discuss two schemes more in Section 3.2.3. Suppose we have a region classifier from a prior or a separated algorithm so that X and D can be partitioned as $X_{(m)}$ and $D_{(m)}$, respectively. Let the covariance matrix associated with $X_{(m)}$ be $K_{(m)}$ such that $(K_{(m)})_{ij} = k_{(m)}(\mathbf{x}_{(m),i},\, \mathbf{x}_{(m),j})$ for $i,\, j \in [n_{(m)}]$, where $n_{(m)}$ is the number of samples in m-th subregion. For the fixed g, the PGP model can be trained with D by maximizing the log composite likelihood, which is

$$\ell(\Theta;\ g) \propto \sum_{m=1}^{M} \left(\mathbf{y}_{(m)}^\top K_{(m)}^{-1} \mathbf{y}_{(m)} - \log\det K_{(m)} - n_{(m)} \log 2\pi\right), \tag{3.2}$$

where $\mathbf{y}_{(m)} = [\ y_{(m),1}\ \ \ldots\ \ y_{(m),n_{(m)}}\ ]^\top$. Note that the log composite likelihood in Eq. (3.2) is the sum of local GPs' because they are independent. Consequently, possibly with some ordering process, the PGP produces a block diagonal covariance matrix which implies that

the numerical advantage of PGPs comes from the sparsity. Although constructing the entire covariance matrix is generally unnecessary in practice, it informs us that the model can be manipulated more efficiently by treating each local GP independently. Evidently, PGPs may exhibit discontinuity at partitioning boundaries, and it would be undesirable when the underlying truth is known to be continuous. Making PGPs continuity requires additional techniques, which is beyond the scope of this chapter, such as an ensemble of the posterior local models [10] or patchwork kriging [57].

3.2.2 Active Learning for Gaussian Processes

The essence of active learning is the information criterion, the function quantifying informativeness of unobserved data. By optimizing the information criterion in the design space, the learning machine determines the design point to query to the oracle. The variance and the IMSE criteria are widely considered in active learning for GPs due to their versatility and simplicity [8, 12, 19], so we briefly review two criteria following terminologies in [12] and then establish our new criterion.

Suppose we intend to determine the next sampling location $\mathbf{x}_{n+1} \in \Omega$ with the GP without partitioning. The variance criterion is

$$J_{\mathrm{V}}(\mathbf{x}) = k(\mathbf{x}) - \mathbf{k}(\mathbf{x}, X) K^{-1} \mathbf{k}(\mathbf{x}, X)^{\top}, \tag{3.3}$$

where f is the posterior GP given X. ALM maximizes the variance criterion so as to select

the location with the greatest predictive variance. Meanwhile, the IMSE criterion is

$$J_{\mathrm{IMSE}}(\mathbf{x}) = \int_{\Omega} \mathrm{Var}\left(f(\mathbf{s}|\,\mathbf{x})\right)\, p(\mathbf{s})\, \mathrm{d}\mathbf{s}, \tag{3.4}$$
$$\mathrm{Var}\left(f(\mathbf{s}|\,\mathbf{x})\right) = k(\mathbf{s}) - \mathbf{k}(\mathbf{s},\, X_{n+1}) K_{n+1}^{-1}\, \mathbf{k}(\mathbf{s},\, X_{n+1})^{\top}, \quad X_{n+1} = \left[\, X_n \;\; \mathbf{x} \,\right]^{\top},$$

where $\mathbf{s} \in \Omega$ with the density (or importance) $p(\mathbf{s})$, and K_{n+1} is the covariance matrix associated with X_{n+1}. Minimizing the IMSE criterion selects the location which is expected to reduce predictive uncertainty the most over Ω, and we refer to the active learning with the IMSE criterion as ALC.

There are more behind derivations of both criteria, but we mention them shortly herein. It turns out that ALM is equivalent to the maximum entropy design (in Gaussian cases) since the choice leads to maximizing the determinant of the covariance matrix. Meanwhile, ALC can be explained by minimizing the generalized mean squared error (MSE) in statistical learning, which can be decomposed into the bias and the variance. Although the variance criterion is straightforward and numerically inexpensive, ALC empirically has shown better performance than ALM [12, 19]. Moreover, ALC can comprehensively consider the importance of $\mathbf{s} \in \Omega$ in the information criterion. It allows us to incorporate prior knowledge and give more weight to a specific region, thereby making the algorithm more distinguishable from the space-filling design.

3.2.3 Partitioned Active Learning Strategy

Dividing the design space in heterogeneous systems allows flexible modeling for GPs, while most existing partitioning methods do not refer to heterogeneous features therein. Moreover, the most widely used active learning strategies in the previous section also do not take into

account partitions in their information criteria. This section illustrates our proposed active learning strategy that includes a separated partitioning scheme based on heterogeneity and the new information criterion that exploits the partitioned structure.

Partitioning based on Heterogeneity

The main objective of partitioning in our method is to suppress the adverse effect of the heterogeneous subsystems to improve learning efficiency. There are two ways to determine partitions in PGPs: model-driven and model-free. Model-driven partitioning establishes the region classifier by maximizing the likelihood of the hierarchical model in Eq. (3.1) [9, 10, 58]. It is promising to exploit the training data to construct a well-fitted model, while it may induce partitions that are difficult to interpret. For example, the number of subregions can be unnecessarily high, which leads to the loss of correlation information with overfitting. Another drawback is the expensive computational cost of MCMC algorithms. Especially, it may be prohibitive when the model involves a higher dimensional region classifier. Meanwhile, the model-free approach builds the region classifier separately prior to fitting local models. Although it may induce an inferior fit of the partitioned model with respect to the training data, the region classifier can be directly established based on predetermined criteria (e.g., heterogeneous factors). Moreover, the partitioning can be examined and modified with background knowledge by experts before plugging the classifier into the partitioned model. Therefore, if we have prior knowledge of partitioning (e.g., the number of subregions, and features that determine the heterogeneity), the model-free approach is more reasonable.

Agglomerative clustering in [11] is model-free, and it utilizes the finite difference in observed data as the heterogeneous feature. It is also based on Voronoi tessellation as [9, 18], while it is capable of generating nonconvex subregions and does not involve MCMC. However, the numerical complexity of Voronoi tessellation exponentially increases with the dimension, so

it can be numerically problematic in high dimensional problems such as our case study in Section 3.4.3. Furthermore, it is inapplicable when the number of samples is insufficient for the problem dimension, and the agglomerative clustering method may yield undesirable singleton clusters as shown in Fig. 3.2. Although it is possible to employ agglomerative clustering for our partitioned active learning, we provide another path of partitioning, which tackles the aforementioned limitations.

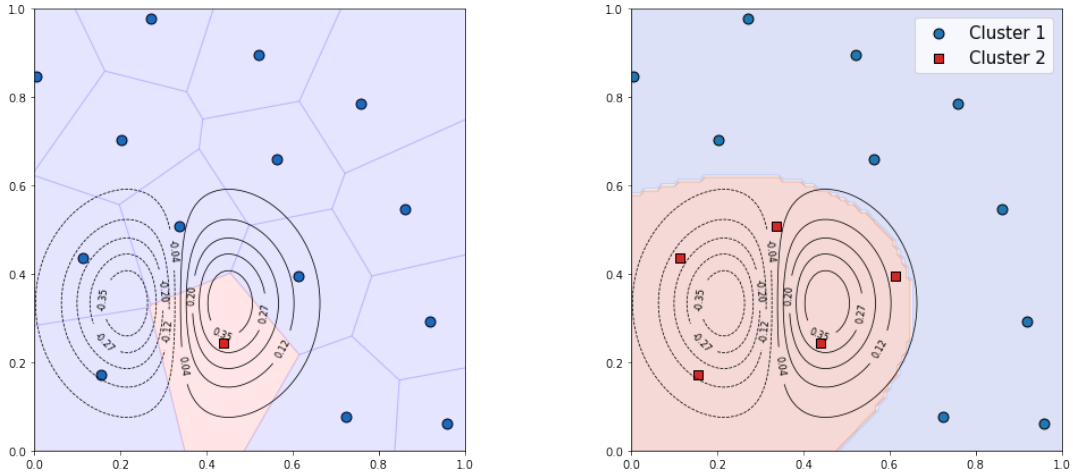

Figure 3.2: Partitioning in two-dimensional simulation. (Left) Agglomerative clustering [11]. (Right) Support vector classification with the mean shift.

To divide the design space based on heterogeneity, we first need to define heterogeneous characteristics. There could be many candidates that induce heterogeneity in systems, and the example includes variations and variances. Let h be the heterogeneous feature. The degree of variation at a point can be quantified with the gradient norm, while the best accessible reference with finite observations would be finite differences between the point and neighbors, which is

$$h_r(\mathbf{x}) = \text{avg}\left(\frac{d(y_i,\, y)}{d(\mathbf{z}_i,\, \mathbf{x})} :\ \mathbf{z}_i \in N_r(\mathbf{x})\right), \tag{3.5}$$

where $N_r(\mathbf{x}) = \{\mathbf{z} | d(\mathbf{x},\, \mathbf{z}) \leq r,\, \mathbf{z} \in \Omega\}$ for $r > 0$. The radius r should be large enough

to cover more than one adjacent point, but not too large to exclude irrelevant points. A plausible value is $1 \sim 1.5$ times the minimum distance in observations. The main difference between Eq. (3.5) and the dissimilarity in [11] is that Eq. (3.5) considers all adjacent samples for every design point, while the other calculates the dissimilarity of every pair like a graph model to conduct agglomerative clustering. For heteroscedastic stochastic systems, variance can be a good reference. The regional variance can be approximated as

$$h_r(\mathbf{x}) = \mathrm{Var}\left(y_i \colon\ \mathbf{x}_i \in N_r(\mathbf{x})\right). \tag{3.6}$$

Once heterogeneous features are evaluated for the observed data, kernel density estimation can be used to cluster them. In this chapter, we utilize mean shift [59] to detect modes of heterogeneity in the design space as follows. There are several advantages to the use of mean shift for clustering. First, it does not require a specific number of clusters to implement the procedure. There are many ways to estimate the bandwidth with data a prior [60], and it can be easily modified with a single bandwidth parameter for a specific number of clusters. Second, its complexity is mainly dependent on the number of samples rather than the input dimension, so it is numerically more advantageous than the tessellation-based methods in high-dimensional problems. Let $\tilde{\mathbf{x}}_i$ be the augmented vector of $\mathbf{x}_i$ with $h_i \equiv h(\mathbf{x}_i)$, and consider a nonnegative bandwidth parameter γ for the kernel to climb the density of the heterogeneity feature (e.g., finite difference). Then, we can find the heterogeneity mode associated with $\tilde{\mathbf{x}} \in \Omega \times \mathbb{R}$ by iteratively updating the mean shift vector ($\tilde{\mathbf{z}}$) as

$$\tilde{\mathbf{z}}_{j+1} = \tilde{\mathbf{z}}_j + \frac{\sum_{i=1}^{n} \tilde{\mathbf{x}}_i \exp\left(-\frac{1}{2}\left\|\frac{\tilde{\mathbf{z}}_j - \tilde{\mathbf{x}}_i}{\gamma}\right\|^2\right)}{\sum_{i=1}^{n} \exp\left(-\frac{1}{2}\left\|\frac{\tilde{\mathbf{z}}_j - \mathbf{x}_i}{\gamma}\right\|^2\right)}, \tag{3.7}$$

for $j = 1, 2, \ldots,$ until its convergence. The sequence of Eq. (3.7) begins from each of

$\tilde{\mathbf{z}}_1 \in \{\tilde{\mathbf{x}}_i \colon i \in [n]\}$, and the number of converging modes is the number of clusters. The number of partitions is important, while the mean shift procedure may not provide a desirable partition. Especially, without prior knowledge about the presence of heterogeneity, the decision on the number of partitions can be very challenging. In this case, the number of clusters should be set in a conservative manner (i.e., as small as possible) to keep the possible correlation between subregions, and the mean shift result should be examined and modified according to some criteria (e.g., the size of each cluster, the number of clusters).

However, the resulting mean shift model cannot be used for active learning, since it requires the heterogeneous feature of unobserved data. Thus, we need to employ a discriminative function that classifies input to the heterogeneity classes induced by the mean shift procedure. Definitely, the model class for the discriminative function should be determined by the characteristics of the underlying boundaries, while nonlinear support vector classification (SVC) is promising due to its flexibility and effectiveness in high-dimensional problems. The discriminative function will be the region classifier, and the procedure is provided in Algorithm 1.

Partitioned Information Criterion

In this section, we construct the information criterion induced by the heterogeneity-based region classifier. When the IMSE criterion is considered for a candidate location $\mathbf{x} \in \Omega_{(m)}$ with PGPs, it can be written as

$$J(\mathbf{x}) = \sum_{l \neq m} \int_{\Omega_{(l)}} \mathrm{Var}\left[f_{(l)}(\mathbf{s}_{(l)})\right] p(\mathbf{s}_{(l)}) \, \mathrm{d}\mathbf{s}_{(l)} + \int_{\Omega_{(m)}} \mathrm{Var}\left[f_{(m)}(\mathbf{s}_{(m)} | \mathbf{x})\right] p(\mathbf{s}_{(m)}) \, \mathrm{d}\mathbf{s}_{(m)}, \quad (3.8)$$

where $\mathbf{s}_{(m)} \in \Omega_{(m)}$. Interpreting each term in Eq. (3.8) is quite worthwhile. The first term is the sum of IMSEs except for $f_{(m)}$, and it is invariant to the choice of $\mathbf{x} \in \Omega_{(m)}$. The second

term is equivalent to Eq. (3.4), in which $f_{(m)}$ is only considered so that Eq. (3.8) will only take account of the local region where the candidate is located. Briefly, there are two main differences between Eq. (3.4) and Eq. (3.8): (i) considering IMSEs over other local regions; and (ii) the localized IMSE criterion. We subsequently focus on each term considering their meanings, thereby efficiently minimizing Eq. (3.8).

Heuristically, Eq. (3.8) is more likely to be minimized when the most uncertain local GP, which has more potential to be reduced with additional observations, is considered in the second term, since the local GP. Each IMSE in the first term indicates the regional uncertainty of PGP, so it can be used for investigating the most uncertain region. Let us denote the regional uncertainty of each local GP as $\mathbb{V}_{(m)}$ for $m \in [M]$. Global searching chooses the most uncertain region from the following categorical distribution,

$$m^* \sim \text{Cat}\left(\frac{\mathbb{V}_{(1)}}{\sum_{m=1}^{M}\mathbb{V}_{(m)}}, \ldots, \frac{\mathbb{V}_{(M)}}{\sum_{m=1}^{M}\mathbb{V}_{(m)}}\right). \tag{3.9}$$

A straightforward way to choose the most uncertain region is to select the maximum in the sequence of regional uncertainty while sampling from Eq. (3.9) can prevent some undesirable states such as falling into a specific subregion due to insufficient information about other subregions, or presence of multiple subregions with comparable variance.

Once the most uncertain region is determined by Eq. (3.9), we focus on the second term within $\Omega_{(m^*)}$ as the local searching with the following criterion:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\Omega_{(m^*)}} \int_{\Omega_{(m^*)}} \text{Var}\left(f_{(m^*)}\mathbf{s}_{(m^*)}|\,\mathbf{x}\right) p(\mathbf{s}_{(m^*)})\, \text{d}\mathbf{s}_{(m^*)}. \tag{3.10}$$

Since the local GP in Eq. (3.10) reflects the local behavior of underlying function excluding heterogeneity from other regions, it can lead to improvement in the exploitation of active learning by avoiding implausible predictive uncertainty. We call the sequential criteria, Eq.

(3.9) and Eq. (3.10), by Partitioned IMSE (PIMSE), and the active learning with PIMSE as Partitioned ALC (PALC). The PIMSE criterion asymptotically converges to a steady state as the number of observations increases as the following proposition.

Proposition 3.1 (Convergence of PALC)**.** *The PIMSE criterion uniformly converges on* Ω *as* $n \to \infty$.

Proof. Let us denote the local IMSE criterion in Eq. (3.10) with n observations as J_n. The sequence of $\{J_n, J_{n+1}, \ldots\}$ is monotonically decreasing for all $\mathbf{x} \in \Omega$ by Theorem 2 of [61], and clearly lower bounded by zero. Thus, it converges. □

It is noteworthy that the uniform convergence in Proposition 3.1 implies that it leads to the best estimator of f, which of uncertainty is irreducible. That is, if one employed the stochastic local kriging, the PIMSE criterion will be asymptotically dominated by the intrinsic uncertainty of the nugget effect.

The pseudocode of PALC is provided in Algorithm 1. The prerequisite values $\{N_{\text{iter}}, N_{\text{ref}}, N_{\text{cand}}\}$ stand for the numbers of attainable samples (i.e., budget), reference design points for Eq. (3.9) and candidate points for Eq. (3.10), respectively, and they are required for practical implementation of the proposed algorithm. X_{ref} is a reference set composed of N_{ref} space-filling points ($\{\mathbf{s}_i\}_{i=1}^{N_{\text{ref}}} \subset \Omega$), which is required to implement the global searching as

$$\mathbb{V}_{(m)} \approx \frac{1}{N_{\text{ref}}} \sum_{i=1}^{N_{\text{ref}}} \text{Var}\left(f_{(m)}(\mathbf{s_i}|\,\mathbf{x})\right)\, \tilde{p}(\mathbf{s}_{(m),i}), \tag{3.11}$$

where $\tilde{p}(\mathbf{s}_{(m),i})$ is the approximated probability mass function at $\mathbf{s}_{(m),i}$. Then, N_{cand} is passed to generate the candidate pool $X_{\text{cand}} \subset \Omega_{(m^*)}$, and the subset of the reference set associated with the chosen subregion will be used for solving Eq. (3.10). In this chapter, we have generated X_{ref} and X_{cand} with new LHD in every step to encourage exploration.

Algorithm 1: Partitioned ALC

Data: $D = (X, \mathbf{y}), N_{\text{iter}}, N_{\text{ref}}, N_{\text{cand}}$

Partitioning based on Heterogeneity
Calculate $\mathbf{h}$ with D
Implement mean shift, Eq. (3.7), on $(X, \mathbf{h})$ to generate M clusters
Train g on X and the generated cluster labels $\mathbf{m}$
Train f on D with g by maximizing Eq. (3.2)
while $i < N_{\text{iter}}$ **do**
    Global Searching:
    Generate $X_{\text{ref}} \subset \Omega$ with N_{ref} space-filling design
    Calculate $\mathbb{V}_{(m)}$ for $m \in [M]$ with X_{ref} with Eq. (3.11)
    $m^* \sim \text{Cat}\left(\frac{\mathbb{V}_{(1)}}{\sum_m \mathbb{V}_{(m)}}, \ldots, \frac{\mathbb{V}_{(M)}}{\sum_m \mathbb{V}_{(m)}}\right)$
    Local Searching:
    Generate $X_{\text{cand}} \subset \Omega_{(m^*)}$ with N_{cand} space-filling design
    Obtain $\mathbf{x}^*$ by solving Eq. (3.10) with $\Omega_{(m)} = X_{\text{cand}}$
    Obtain y^* at $\mathbf{x}^*$
    $D = D \cup \{(\mathbf{x}^*, y^*)\}$
    Update g and f with D
    $i = i + 1$

Aside from the main algorithm, there is no universal concrete theorem for the optimal portion of initial sampling, while some empirical suggestions can be found in [8, 62]. However, we can conjecture that the number of initial samples has a tradeoff property. If the initial sampling is weighed too much, the advantage of active learning will be diluted. Otherwise, active learning can be hindered by low-quality information coming from unreliable intermediate models. Also, when the region classifier g, the number should be enough to obtain an acceptable g. For termination of the active learning procedure, early stopping can be a reasonable choice other than the sampling budget in Algorithm 1 when a testing dataset or cross-validation scheme is available.

3.2.4 Cholesky Update based Numerical Remedies to Tackle the Computational Complexity Challenge

The IMSE criterion is numerically more demanding than the variance criterion since it involves the inversion of K_{n+1} which should be updated for every candidate. That is, the calculation of the IMSE criterion requires $\mathcal{O}(n^3)$ for each candidate. Moreover, when N candidates are provided to the active learning module, the computational cost is multiplied by the number. Although the significance of their effects may vary with situations, the effect of the candidate number can be more considerable than the inversion cost. Therefore, to improve the numerical aspect of PALC, we should provide some remedies for both matrix inversion and the number of candidates.

Global searching reduces the number of candidates by narrowing down the region of interest. Generally, candidates for active learning are given space-filling or dense-grid over the design space. Therefore, taking the subset of candidates in the most uncertain region with the global searching reduces the number of matrix inversions proportional to the ratio of the most uncertain region considering the importance of input.

The matrix inversion cost of PGP is automatically alleviated by partitioning the design space with the block diagonal covariance matrix. That is, the inversion cost reduces from $\mathcal{O}(n^3)$ to at most $\mathcal{O}({n_{(m)}}^3)$, where $n_{(m)} < n$. Another applicable remedy is updating the inverse of $K_{n_{(m)}+1}$ in Eq. (3.10) exploiting $K_{n_{(m)}}^{-1}$ iteratively. Although it is possible to apply the Sherman–Morrison formula to get the updated inverse matrix [19], the Cholesky decomposition [53] for solving the linear system $K_{n_{(m)}}^{-1}\mathbf{k}$ could be more preferable considering the numerical stability and cost. Given that the Cholesky factor (a lower triangular matrix) of $K_{n_{(m)}}$ is known, updating the Cholesky factor of $K_{n_{(m)}+1}$ only requires the forward substitution step of the size $n_{(m)}$ triangular system, thus it needs only $\mathcal{O}(n_{(m)}^2)$ instead of $\mathcal{O}({n_{(m)}}^3)$.

A more detailed procedure for the Cholesky update is provided as follows.

Suppose we have the Cholesky factor L of K_n, which is the covariance matrix of X_n, such that $K_n = LL^\top$. We aim to get the Cholesky factor $\hat{L}$ of

$$K_{n+1} = \begin{bmatrix} K_n & \mathbf{k}_\mathbf{n}^* \\ \mathbf{k}_\mathbf{n}^{*\top} & k(\mathbf{x}^*) \end{bmatrix},$$

where $\mathbf{x}^*$ is a candidate input, and $\mathbf{k}_n^* = k(X_n, \mathbf{x}^*)$. Since K_{n+1} shares the same part of K_n, it turns out that the first $n \times n$ elements of $\hat{L}$ are equivalent to L. Therefore, we can apply the Cholesky-Banachiewicz algorithm for $\hat{L}$ as

$$\hat{L}_i = \begin{cases} L_{i,i}^{-1}\left(\mathbf{k}_{\mathbf{n}i}^* - \sum_{j=1}^{i-1} \hat{L}_j L_{i,j}\right), & i \in [n] \\ \sqrt{k(\mathbf{x}^*) - \sum_{j=1}^{n} \hat{L}_j}, & i = n+1. \end{cases} \tag{3.12}$$

Rather than calculating the PIMSE directly, PALC can be faster by skipping the redundant computation and applying the Cholesky updating approach. The Cholesky factor L can be used for the predictive variance of GP in Eq. (3.3) for $\mathbf{s} \in X_{\text{Ref}}$, and the global searching criterion of Eq. (3.9) as

$$\text{Var}_n^2(\mathbf{s}) = k(\mathbf{s}) - \mathbf{v}_n{}^\top \mathbf{v}_n, \qquad \mathbf{v}_n = L^{-1}\mathbf{k}_n, \tag{3.13}$$

where $\mathbf{k}_n = k(X_n, \mathbf{s})$. In the same manner, the optimal solution of local searching in Eq. (3.10) can be obtained by minimizing

$$\text{Var}^2_{(m^*),n_{(m^*)}+1}(\mathbf{s}) = k(\mathbf{s}) - \mathbf{v}^{*\top}\mathbf{v}^*, \tag{3.14}$$

$$\mathbf{v}^* = \hat{L}^{-1} k(X_{n+1}, \mathbf{s}), \tag{3.15}$$

Since we already have the solution of Eq. (3.15) partially with $\mathbf{v}$ (i.e., $\mathbf{v}^*[1:n] \equiv \mathbf{v}_n$), we need only $v^*_{n+1} := \mathbf{v}^*[n+1]$, which can be calculated with a forward substitution as

$$v^*_{n+1} = k(\mathbf{x}^*, \mathbf{s}) - \sum_{j=1}^{n} \hat{L}_k \mathbf{v}_j. \tag{3.16}$$

Since $k(\mathbf{s})$ in Eq. (3.14) is invariant to $\mathbf{s}$ and $\mathbf{x}^*$, only the second term $\mathbf{v}^{*\top}\mathbf{v}^*$ is usually considered as the simplified PIMSE criterion, which must be maximized in PALC.

3.3 Simulation Study

In this section, we evaluate our active learning algorithm with simulation data. Two functions are considered that can be visualized straightforwardly. Both functions contain heterogeneous response surfaces, and observation noise is involved. To reduce the variability from the random initial dataset, each simulation is replicated ten times. As our benchmark methods, ALC and ALM are considered for SGPs, and the following partitioned active learning methods are considered: (i) the variance criterion for the local searching (PALM); (ii) PALC without global searching (PALC–NoG). Additionally, uniform random sampling (Rand) and LHD are also considered passive learning. For the evaluation of models, one thousand space-filling design points for each design space are used with root mean squared error (RMSE) as the metric. Total computational times spent on querying are also compared excluding the model training time.

3.3.1 One-dimensional Data

We apply our proposed active learning algorithm to a one-dimensional simulation function

$$f(x) = 2x \sin(8\pi x^3), \tag{3.17}$$

which is defined on [0, 1] as the dotted line in Fig. 3.3. Zero-mean Gaussian noise is imposed with variance $\sigma^2 = 10^{-4}$. We allocate 10 samples for initial training using Maximin LHD, and 20 samples are sequentially obtained via active learning. The function is differentiable, with heterogeneous frequency and amplitude over the domain. With the variation feature in Eq. (3.5), heterogeneity-based partitioning provided 2∼3 partitions, and the number is set to two considering the number of initial samples and the dimension by adjusting the bandwidth parameter with grid searching over $[10^{-5}, 10^3]$. Afterward, logistic regression is used for the region classifier, which resulted in the decision boundary around $x = 0.38 \sim 0.65$.

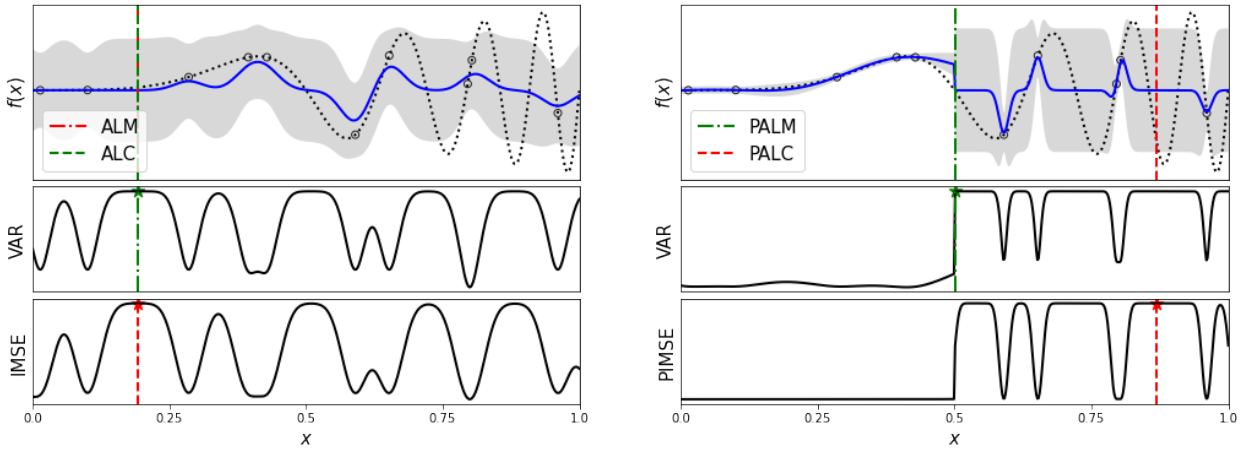

Figure 3.3: Results of different active learning algorithms in one-dimensional simulation. (Left) SGP. (Right) PGP with two local GPs.

Fig. 3.3 shows each GP model fitted with initial samples. First, the right of Fig. 3.3

illustrates that the PGP prevents misled active learning by providing appropriate predictive uncertainty. Also, we can observe from the left of Fig 3.3 that the next design point to be queried in the SGP model is chosen from the low-frequency region with both variance and IMSE criteria. Meanwhile, in the PGP model, the variance criterion takes the point from the boundary, while the IMSE does not. It implies that the IMSE criterion can be more promising when the adjacent local GPs show comparable predictive uncertainties.

Table 3.1 summarizes the results after the full data acquisition, and we grouped the considered methods into three: (i) passive learning (Rand; and LHD); (ii) variance methods (ALM; and PALM); and (iii) IMSE-based methods (ALC; PALC–NoG; and PALC). We can observe that the predictive accuracy of PALC outperforms the others. In the computational time, variance methods (ALM and PALM) are faster than the IMSE-based methods, while they are deficient in predictive accuracy; ALM is even worse than random sampling and the LHD. Among the IMSE-based methods, PALC is faster than the others. Moreover, if we focus on PALC and PALC–NoG, we can observe that global searching not only reduces computational time but also improves learning efficiency.

3.3.2 Two-dimensional Data

We extend the simulation study to a two-dimensional function (shown in the left of Fig. 3.4), which is also used in [19, 20]. The function comprises two regions: even and uneven, and zero-mean Gaussian noise with variance $\sigma^2 = 10^{-6}$ is imposed as the observation noise. Similarly, we begin with 15 samples with LHD and obtain 15 additional samples via active learning.

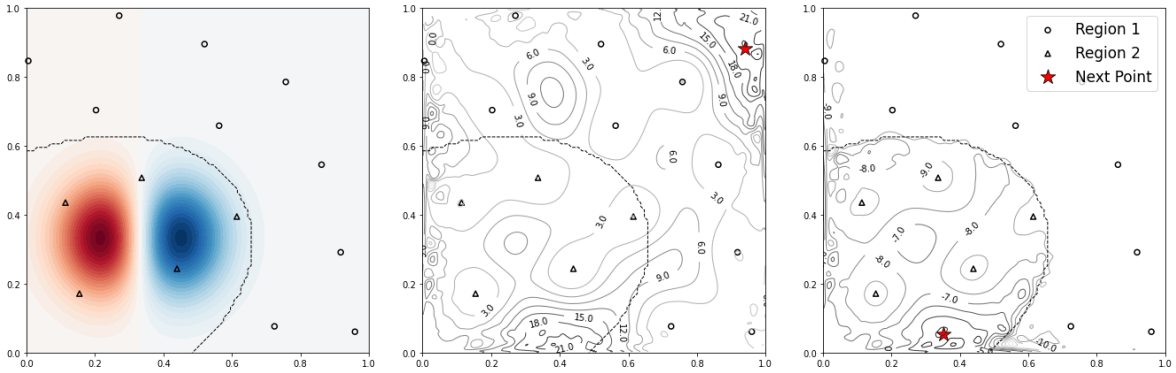

Figure 3.4: IMSE criterion contour plots in two-dimensional simulation: (Left) Ground truth and initial design points with partitioned regions. (Center) IMSE of ALC. and (Right) PIMSE of PALC.

Table 3.1: Simulation results.

Methods	1-D		2-D	
	RMSE	Time (s)	RMSE	Time (s)
Rand	0.395 (0.152)	-	0.044 (0.017)	-
LHD	0.332 (0.133)	-	0.046 (0.021)	-
ALM	0.352 (0.159)	0.341 (0.140)	0.060 (0.023)	0.344 (0.081)
PALM	0.072 (0.270)	0.144 (0.049)	0.019 (0.025)	0.151 (0.026)
ALC	0.272 (0.205)	11.468 (0.424)	0.039 (0.013)	8.368 (0.161)
PALC-NoG	0.051 (0.277)	7.049 (0.024)	0.026 (0.004)	5.570 (0.593)
PALC	0.048 (0.273)	3.553 (0.137)	0.016 (0.005)	1.942 (0.016)

For partitioning the design space, we used finite differences as the heterogeneous feature and used SVC after labeling initial samples based on the mean shift result as shown in Fig. 3.4. Some initial datasets yielded three partitions with affordable evenness, while we adhered to two partitions because of the majority of two partitions in all replications. Since region 1 is less interesting than region 2, we can observe that the PIMSE criterion induced by two

independent local GPs provides more relevant information on design points as shown in the right of Fig. 3.4. Consequently, the IMSE criterion with the SGP fails to pick from the more interesting region because of the misled information criterion. In Table 3.1, we can see that PALC surpasses the other methods again in both predictive accuracy and computational time among the IMSE-based methods.

3.4 Case Study

In this section, we apply our approach to constructing surrogate models for three different real-world cases. The purpose of the surrogate models is to embed them into automated systems and to provide UQ in posterior analysis. Since we have little prior knowledge about underlying partitioning, we referred to mean shift results and adjusted the number of partitions according to the smallest cluster size, and evenness of cluster sizes for all cases as the simulation study. Case studies include higher input dimensions than the previous simulations.

3.4.1 Residual Stress in Composite Fuselages Assembly

We apply our proposed active learning strategy to construct the predictive model of residual stress in the composite fuselage assembly process. In the aircraft manufacturing process, composite fuselages are built in several subsections independently, so they are subject to discrepancies in the junction part. The composite fuselage is reshaped using multiple fixed actuators in the automatic shape control. In order to achieve the optimal manufacturing process, shape control needs to consider not only the deformation but also the residual stress in the structure because of their fatal effects on the final product. The development of a

highly accurate predictive model for shape control is very challenging since the problem is endowed with both heterogeneity and the demanding cost of real experiments. Especially, the stress of composite fuselage is more difficult to predict than deformation [56], we apply our method and other benchmarks to predictive modeling of the stress.

In order to implement our case study cost-efficiently, we utilized the FEM model with ANSYS [63], which is well-calibrated based on the real experiment. The simulation mimics the real shape adjustment process that has ten actuators under the fuselage section as shown in the left of Fig. 3.5, and the maximum magnitude of the actuator's force is 450 lbf. The maximum residual stress on the fuselage section is our interest, which is shown in the right of Fig. 3.5 and measured in the psi scale. The ten-dimensional design space is partitioned into three regions with SVC based on finite differences since a higher number of partitions yields singleton clusters. The Matérn kernel is used for GPs. As the initial dataset, 50 LHD samples are drawn, and additional 30 samples are sequentially obtained from 1,000 LHD points with different active learning strategies. The model evaluation is conducted with a separate testing dataset composed of 100 LHD samples, and mean absolute error (MAE) is used as a metric.

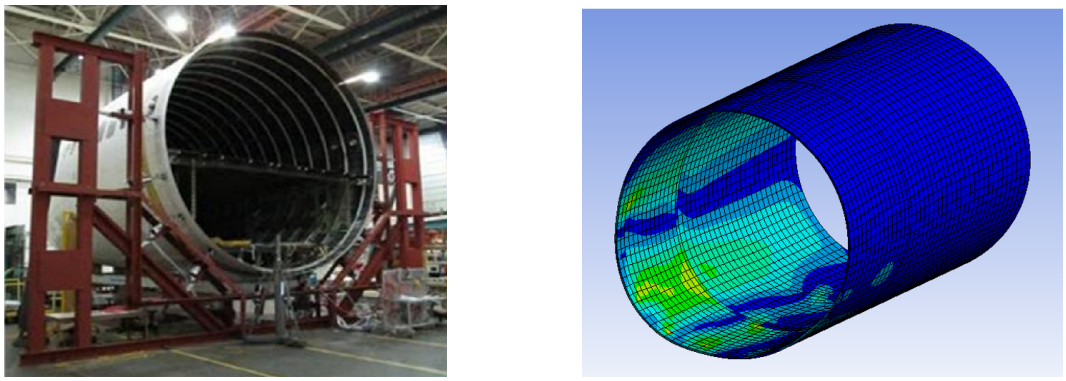

Figure 3.5: Shape adjustment of the composite fuselage. (Left) Composite fuselage installed upon the fixture with actuators [64]. (Right) Simulated residual stress of composite fuselage in ANSYS.

Table 3.2 summarizes the results of each learning strategy in the case study. PALC surpasses the other methods in both predictive error and computational time among the IMSE criterion-based methods. Interestingly, except for PALC–NoG, the other active learning methods are worse than the two passive learning strategies. Moreover, PALM performs the worst in this case, even though they chose design points to spread out the design space well. In computational times, we can observe that PALC–NoG took more time than ALC.

3.4.2 Tribocorrosion in Aluminum Alloys

In our second case study, the material loss rate during stress corrosion (i.e., tribocorrosion) in aluminum alloys with six control variables is considered. To test the tribocorrosion resistance of metals, experimental tests, and FEM simulations were carried out by scratching the surface of the samples in the corrosive environment [65, 66] as shown in Fig. 3.6. During the tribocorrosion process, the mechanical deformation and the electrochemical processes including active corrosion and passivation work synergistically to cause material degradation. The FEM model calculates the contact mechanics between the indenter and the sample and simulates the wear process as well as the wear-accelerated material dissolution of the corrosion process, and generates the volume loss results. The 6 control variables for the FEM model are material property descriptors: Young's modulus, yield strength, anodic Tafel slope, anodic exchange current density, cathodic Tafel slope, and cathodic exchange current density. The former two govern the mechanical properties while the latter four determine the corrosion behavior of the alloy. The output of the FEM model is the tribocorrosion rate of the alloy, expressed as volume loss per time.

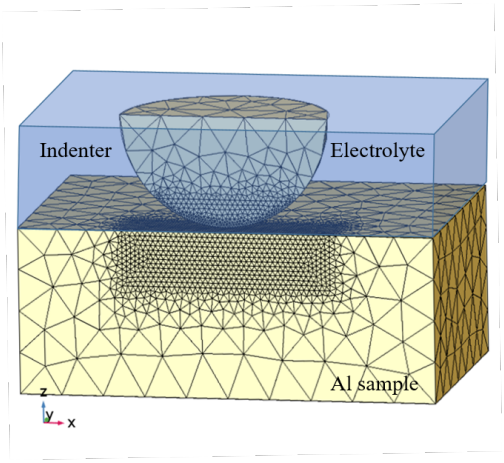

Figure 3.6: Schematic tribocorrosion simulation setup.

The surrogate model of the FEM model is constructed to assist the optimal design of alloys with UQ and to alleviate the high-computational cost of the FEM model. To establish the relationship between material property and tribocorrosion rate, a total of 106 FEM simulations were performed by systematically varying the 6 control variables. Since the scales of variables in the dataset are inconsistent, each variable is normalized to be within a unit interval. For evaluation, relative mean absolute error (RMAE) is used as a metric on account of the infinitesimal scale output. The RMAE is calculated as

$$\text{RMAE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - h(x_i)|}{|y_i|}. \tag{3.18}$$

The PGP in this case is composed of three local GPs with the RBF kernel, and the SVC model is used for partitioning based on the finite differences, which provided quite even-sized clusters. Considering the relatively small size of samples, 5-fold cross-validation is used. That is, about 84 samples are passed to each model as the candidate set, and the rest of the samples are used for the model evaluation. Compared methods are trained up to 50 samples from 20 common initial samples.

Table 3.2 shows the result of each learning strategy. We can observe that PALC achieves the minimum averaged predictive error and the computational time among the IMSE-based methods. Passive learning methods are worse than others, and the variance methods are also worse than the IMSE-based methods. In this case, overall computational times are much lower than in the previous case because of the small number of samples in the candidate pool, while the numerical remedies in PALC have significantly reduced the time of ALC.

3.4.3 Inverse Dynamics of Robot Arm

We apply partitioned active learning to the inverse dynamics problem for the seven degrees of freedom robot arm, of which original data is introduced in [53]. This problem has twenty-one-dimensional input: positions, velocities, and accelerations of seven joints, and seven joint torques as the output. The dataset contains 44,484 training samples and 4,449 testing samples. We regard the training dataset as the unlabeled data pool so we take thirty initial samples with D-optimal design [50] and obtain 30 additional samples from the pool. The testing dataset is used for the reference dataset in IMSE-based methods. For model evaluation, standardized MSE (SMSE) is used, which is

$$\text{SMSE} = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - f(x_i))^2}{\text{Var}\,(Y_{\text{test}})}, \tag{3.19}$$

considering the scale issue. Since we had no specific prior knowledge of the problem, we assigned two partitions according to the finite difference clustering result.

Results in Table 3.2 show that PALC outperforms other methods in prediction accuracy. The computational time result also shows that PALC can significantly reduce the time of naíve IMSE with our proposed method. To compare partitioning methods in a high-dimensional setting, we also implemented agglomerative clustering. For the initial dataset, agglomerative

clustering was inapplicable when the number of the initial dataset was less than the input dimension. For thirty samples, it took 3$\sim$ 4 seconds, while the proposed partitioning method took only less than 0.1 seconds. Moreover, for 60 samples, our method took only 0.3 seconds, while the agglomerative clustering took more than 1,200 seconds with memory overflow error in Voronoi tessellation.

Table 3.2: Case study results.

Methods	Case I: Fuselage Assembly		Case II: Tribocorrosion		Case III: Inverse Dynamics	
	MAE	Time (s)	RMAE	Time (s)	SMSE	Time (s)
Rand	3.858	-	0.028 (0.017)	-	1.578	0.474
LHD	3.319	-	0.028 (0.017)	-	1.441	0.551
ALM	4.545	2.6×10^{-4}	0.026 (0.017)	0.014 (0.003)	1.446	0.973
PALM	11.555	2.5×10^{-4}	0.023 (0.012)	0.104 (0.000)	-	-
ALC	4.693	11.596	0.022 (0.013)	1.283 (0.127)	1.452	425.740
PALC–NoG	3.691	12.750	0.022 (0.012)	0.764 (0.018)	-	-
PALC	3.207	9.729	0.020 (0.013)	0.757 (0.017)	1.093	237.672

3.5 Summary

Active learning is machine learning that seeks to improve sampling efficiency and lower data collection cost. Existing active learning strategies mainly focus on investigating homogeneous response surfaces, so they are insufficient for reliable and cost-efficient surrogate modeling of heterogeneous systems. This chapter is dedicated to establishing an efficient partitioned active learning strategy that adopts two-step searching schemes based on the PIMSE criterion structure. By partitioning the design space into multiple subregions according to heterogeneity in the target system, the global searching scheme refers to the integrated predictive uncertainties of local GPs to determine the most uncertain subregion. The global searching

scheme allows us to reduce the region of interest, thereby not only accelerating the searching speed but also improving the overall learning efficiency as shown in the simulation and case studies. The local searching scheme exploited the chosen local GPs in the global searching phase, so the localized IMSE criterion may provide more relevant information minimizing the interruption of heterogeneous characteristics in the other regions.

For the numerical perspective of active learning, the following applicable remedies are provided: reducing the number of candidates with the global searching scheme, and the Cholesky factor update, which can be embedded into PALC. From the simulation and the case study, PALC outperformed the benchmark methods including passive learning, the variance criterion, and the IMSE-based methods. Moreover, the proposed partitioning method is applicable to high-dimensional problems with consideration of heterogeneity. The global searching scheme dramatically improved the performance of PALC by comparing our method to the PIMSE without the global searching step. Diverse case study results imply that our method is also applicable to other domains where heterogeneity exists.

Chapter 4

Failure-averse Active Learning for Physics-constrained Systems

Industrial and engineering systems are generally subject to physics constraints that may induce fatal failures when they are violated, while such constraints are frequently underestimated in active learning. In this chapter, we develop a novel active learning method that avoids failures considering implicit physics constraints that govern the system. The proposed approach is driven by two tasks: the safe variance reduction explores the safe region to reduce the variance of the target model, and the safe region expansion aims to extend the explorable region by exploiting the probabilistic model of constraints. The global acquisition function is devised to judiciously optimize the acquisition functions of two tasks, and its theoretical properties are provided. The proposed method is applied to the composite fuselage assembly process with consideration of material failure using the Tsai–Wu criterion, and it is able to achieve zero failure without the knowledge of explicit failure regions.

4.1 Introduction

Active learning in engineering applications has been primarily utilized without considering coexisting or inherent constraints that may have different processes thereof. It is very crucial to consider such constraints in engineering systems since most of them are subject to physics

constraints that may induce fatal and irreversible failures. For example, the design of the automatic shape control system in composite aircraft manufacturing needs to consider potential material failures such as crack, buckling, and delamination caused by intolerable inputs [64]. The inverse PDE problem is another example that is used to calibrate parameters in physics models based on observations. It usually involves physics constraints composed of different PDEs, and the constraints must be satisfied in order to build first-principle models. In both cases, the application of active learning without considering physics constraints may induce fatal failures or biased models, so constraints must be considered in engineering problems.

To consider physics constraints in active learning, it is straightforward to define a safe region where design points satisfy the safe conditions (with high probability at least) and conduct active learning within the safe region. However, it is not always possible in practice, since physics constraints cannot be explicitly attained due to the complex nature of the system. Examples include the crash damage analysis of commercial vehicles composed of different materials and physicochemical interactions in the corrosion of alloys. In these cases, fundamental physics laws and equations cannot be directly applied or are insufficient to accommodate the complexities of mechanisms. Physics-based numerical methods such as finite difference methods and FEMs [67] are well-established and convincing to analyze large classes of complex structures including failures, while they are too computationally demanding to identify the entire safe region. Moreover, their deterministic solutions are vulnerable to various uncertainty sources such as material properties, geometries, and loads.

In order to circumvent the aforementioned limitations of physics-based approaches, machine learning models have been widely used in the engineering domain due to their flexibility, inexpensive prediction, and capability of UQ. Physics information can be highly advantageous for machine learning in several aspects such as generalization and physical consistency [68]. Especially, GPs have shown remarkable performance in stochastic analysis of structural

reliability that aims to evaluate the probability of system failure [69]. A common reliability analysis approach employs the GP surrogate model of performance function associated with the system failure and uses the acquisition function (e.g., [70, 71]) that leads to sampling near the boundary of safe and failure regions. The boundary is called the limit state, which is the margin of acceptable structural design. However, the reliability analysis is mainly interested in the response surface associated with failure, and it is time-consuming due to the requirement of a large number of samples to estimate the underlying distribution at the limit state. Hence, it can be data-inefficient to implement reliability analysis prior to the estimation of a safe region. So the development of flexible active learning that takes account of target and failure processes simultaneously is promising for systems with implicit physics constraints.

The principle of active learning is the exploitation of knowledge from observations and exploration by tackling the knowledge such as the design point with maximum entropy or the most disagreeable point in the set of hypotheses. However, if implicit constraints exist in the design space, active learning can be very challenging, since the most informative design point may be located in the failure region. Conversely, if active learning is too conservative to avoid failures, the resulting model will be vulnerable in the unexplored safe region. Consequently, active learning for physics-constrained systems should simultaneously take into account the following objectives:

1. maximizing the information acquisition for the target model;

2. expanding the explorable safe region by focusing on constraint functions,

and they must be achieved safely. Definitely, the two objectives are at odds since they are associated with different functions, so the active learning strategy must be judiciously controlled.

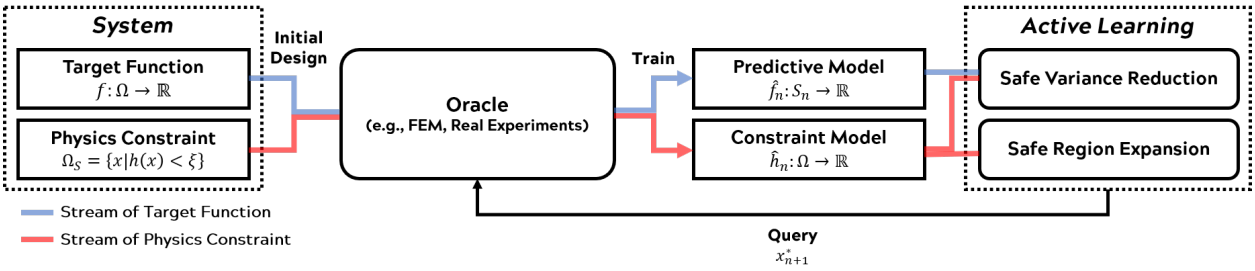

Figure 4.1: Overview of the proposed methodology.

In this chapter, we propose an active learning methodology for systems that are constrained by implicit failure processes. The overview of the proposed method is illustrated in Fig. 4.1. The target process in the system which we aim to learn is subject to failure processes with implicit physics constraints. Both processes can be evaluated via physics-based simulations or experiments, which are expensive to observe. In order to alleviate the sampling cost, we build the predictive model for the target function and constraints by imposing GP priors on them and initializing with a proper design (e.g., space-filling). The objective of our active learning is to train the predictive model of the target function data efficiently and safely by minimizing the cost from undesirable failures incurred by implicit physics constraints. The active learning strategy is built upon two sub-strategies: (i) safe variance reduction and (ii) safe region expansion. To minimize the predictive model variance with respect to the implicit safe region, safe variance reduction explores the estimated safe region induced by the constraint model. Concurrently, safe region expansion evaluates unobserved samples with respect to their closeness to the safe region boundary to improve the estimated safe region. Two sub-strategies are threaded under the multiobjective optimization (MOO) framework so that informativeness in both the target and the safety can be simultaneously considered. Our contributions in this chapter are as follows.

1. We develop the safe variance reduction strategy to improve the predictive performance of the target model under the regime of implicit physics constraints.

2. The safe region expansion strategy is devised to expand the explorable safe region for further improvement of the target model, which concurrently dedicates to avoiding failures.

3. A new acquisition function is proposed that flexibly integrates two heterogeneous strategies.

This chapter is organized as follows. In Section 4.2, we elucidate our active learning strategy considering implicit constraints to avoid failures. Section 4.3 illustrates how the proposed strategy works under the regime of implicit constraints with the simulation study. The real-world application to predictive modeling of composite fuselage deformation considering structural failures is presented in Section 4.4. Lastly, a summary of this chapter is provided in Section 4.5.

4.2 Methodology

An efficient safe exploration under implicit constraints can be accomplished when active learning devotes to both target approximation improvement and explorable region expansion. However, existing methods in Section 2.1 lack explorable region expansion and are indeliberate in avoidance of failure sampling. In this section, we propose our failure-averse active learning method for physics-constrained systems. We begin with specifying our problem whose constraints can be evaluated along with the target process. Then, we describe safe variance reduction and safe region expansion in detail, respectively. At last, they are combined in the integrated acquisition function under the MOO paradigm, and the practical implementation of the overall algorithm is discussed.

4.2.1 Problem Statement and Gaussian Process Priors

Consider a system defined over a compact and connected design space $\Omega \subseteq \mathbb{R}^D$. The system includes the target function $f : \Omega \to \mathbb{R}$, which we want to predict, and the constraint function $h : \Omega \to \mathbb{R}$, related to the system failure and assumed to be independent of f. Let $\xi \in \mathbb{R}$ be a tolerable failure threshold associated with h, which should be defined conservatively considering the intrinsic uncertainty of the process. For any design point $\mathbf{x} = [\, x_1 \cdots x_D \,]^\top \in \Omega$, the system failure occurs when

$$h(\mathbf{x}) \geq \xi, \tag{4.1}$$

and safe otherwise. For example, f can be the dimensional deformation of a solid structure given a force vector $\mathbf{x}$, and h can be von Mises stress in the structure. Both functions can be observed or evaluated with the costly real experiment and relevant physics-based models as

$$y = f(\mathbf{x}) + \epsilon_f, \qquad z = h(\mathbf{x}) + \epsilon_h, \tag{4.2}$$

where $\epsilon_f \sim \mathcal{N}(0,\, \nu_f^2)$ and $\epsilon_h \sim \mathcal{N}(0,\, \nu_h^2)$ are observation noise. We assume that both f and h are continuous, and ν_f and ν_h are known, while they can be also estimated with GP priors on f and h to be described later with nugget effects. The safe region Ω_S, the subset of design space composed of non-failure design settings, is unknown and difficult to obtain due to the prohibitive cost of the evaluation. We refer to the complementary safe region as the failure region such that $\Omega_F = \Omega \setminus \Omega_S$.

In our case, due to the high cost of evaluation of h and f, we prefer to sample the most informative set of design points that minimizes the generalization error associated with the target function. Suppose we have n samples from the system, denoted by $\mathcal{D}_n = \{(\mathbf{x}_i,\, y_i,\, z_i)\}_{i=1}^n$, and $\hat{f}_n$ and $\hat{h}_n$ are our predictors of f and h trained with $\mathcal{D}_n$, respectively. Then, the

expected risk of $\hat{f}_n$ using L_p loss is

$$\mathcal{R}_f(\hat{f}_n) = \int_{\Omega_S} L_p(f(\mathbf{x}),\ \hat{f}_n(\mathbf{x}))\ \mathrm{d}\lambda(\mathbf{x}), \tag{4.3}$$

where $L_p(y,\, y') = |y - y'|^p$ is the loss function, and $\lambda(\mathbf{x})$ is a probability measure defined over Ω. In our problem, violating Eq. (4.1) may incur a prohibitive cost of failure in the system, so any $\mathbf{x} \in \Omega_F$ will not be considered for the system.

We assume that there exists a reproducing kernel Hilbert space for each of f and h, and they are bounded therein. It allows us to model both functions with GPs with corresponding kernels such that $k_f : \Omega^2 \to \mathbb{R}$ and $k_h : \Omega^2 \to \mathbb{R}$ [53]. In this chapter, we consider the automatic relevance determination using the RBF kernel for $\mathbf{x},\, \mathbf{x}' \in \Omega$ as

$$k_f(\mathbf{x},\, \mathbf{x}') = \kappa_f^2(\mathbf{x} - \mathbf{x}')^\top M_f^2(\mathbf{x} - \mathbf{x}') + \nu_f^2\delta(\mathbf{x},\, \mathbf{x}'), \tag{4.4}$$

where κ_f is the nonnegative scale hyperparameter, M_f is the diagonal matrix of nonnegative length hyperparameters $\boldsymbol{\theta}_f = [\ \theta_{f,1},\ \ldots,\ \theta_{f,D}\ ]^\top$, and δ is the Kronecker delta function for the nugget effect. By defining k_h in the same manner, we can write f and h as

$$f(\mathbf{x}) \sim \mathcal{GP}\left(\mu_f(\mathbf{x}),\ k_f(\mathbf{x},\, \mathbf{x}')\right), \qquad h(\mathbf{x}) \sim \mathcal{GP}\left(\mu_h(\mathbf{x}),\ k_h(\mathbf{x},\, \mathbf{x}')\right), \tag{4.5}$$

where μ_f and μ_h are mean functions, assumed to be zero without loss of generality.

Instead of employing a discriminative function for estimating the safe region, the GP regressor is more suitable for physics constraints since the output of h is numerically informative. More explicitly, as $h(\mathbf{x})$ is closer to the failure threshold, we may notice that $\mathbf{x}$ is closer to the safe boundary. Moreover, discriminative functions require observations from both safe and failure regions, while regressors are not subject to such imbalance or the absence of one

class. Therefore, we fit our GP regression model directly on observed outputs from h and refer to the distance between the output and the failure threshold to infer the probability of safety.

Let us denote X_n as the $D \times n$ design matrix of $[\,\mathbf{x}_1 \cdots \mathbf{x}_n\,]$, and $\mathbf{y}_n$ and $\mathbf{z}_n$ as the vector of n observations from f and h, respectively. With GP priors on f and h, the hyperparameters $\Theta_f = \{\kappa_f, \boldsymbol{\theta}_f\}$ and $\Theta_h = \{\kappa_h, \boldsymbol{\theta}_h\}$ can be estimated by maximizing the log marginal likelihoods, which are

$$\ell(\mathbf{y}_n;\ X_n, \Theta_f) = -\frac{1}{2}\mathbf{y}^\top K_{f,n}^{-1}\mathbf{y} - \frac{1}{2}\log|K_{f,n}| - \frac{n}{2}\log 2\pi,$$
$$\ell(\mathbf{z}_n;\ X_n, \Theta_h) = -\frac{1}{2}\mathbf{z}^\top K_{h,n}^{-1}\mathbf{z} - \frac{1}{2}\log|K_{h,n}| - \frac{n}{2}\log 2\pi,$$

where $K_{f,n}$ and $K_{h,n}$ are covariance matrices consisting of every pair of $\mathbf{x},\ \mathbf{x}' \in X_n$ given Θ_f and Θ_h, respectively. Once $\hat{f}_n$ and $\hat{h}_n$ are obtained by maximizing their log marginal likelihoods, the predictive mean and variance of $\hat{f}_n$ at an unobserved design point $\mathbf{x} \in \Omega$ can be derived as

$$\mathbb{E}\left[\hat{f}_n(\mathbf{x})\right] = \mathbf{k}_f(\mathbf{x}, X_n)K_{f,n}^{-1}\mathbf{y}_n,$$
$$\mathrm{Var}\left(\hat{f}_n(\mathbf{x})\right) = k_f(\mathbf{x}) - \mathbf{k}_f(\mathbf{x}, X_n)K_{f,n}^{-1}\mathbf{k}_f(\mathbf{x}, X_n)^\top,$$

and so does $\hat{h}_n$'s.

4.2.2 Safe Variance Reduction

Let us consider L_2 loss (i.e., MSE), although it is not required in practice for our approach, and suppose f_* is an unbiased predictor of f with the minimum MSE (also called the best MSE predictor [50]) with respect to Ω_S in the family of GP. Then, Eq. (4.3) can be decom-

posed as

$$\mathcal{R}_f(\hat{f}_n) = \mathcal{R}_f(f_*) + \int_{\Omega_S} \mathrm{Var}\left(\hat{f}_n(x)\right) \mathrm{d}\lambda(\mathbf{x}), \tag{4.6}$$

which is the sum of the L_2 risk of f_* and the variance of $\hat{f}_n$. Since the L_2 risk of f_* is negligible due to its unbiasedness, Eq. (4.6) can be reduced by focusing on the variance reduction in $\hat{f}_n$. Let us denote the integrated variance of the predictor in Eq. (4.6) as

$$\mathbb{V}_{\Omega_S}(\hat{f}_n) = \int_{\Omega_S} \mathrm{Var}\left(\hat{f}_n(\mathbf{s})\right) \mathrm{d}\lambda(\mathbf{s}),$$

where $\mathbf{s} \in \Omega_S$. Then, the expected variance reduction over Ω_S in $\hat{f}_n$ for an unobserved $\mathbf{x} \in \Omega$ is

$$\Delta\mathbb{V}_{\Omega_S}(\hat{f}_n|\mathbf{x}) = \mathbb{V}_{\Omega_S}(\hat{f}_n) - \int_{\Omega_S} \mathrm{Var}\left(\hat{f}_n(\mathbf{s}|\mathbf{x})\right) \mathrm{d}\lambda(\mathbf{s}), \tag{4.7}$$

$$\mathrm{Var}\left(\hat{f}_n(\mathbf{s}|\mathbf{x})\right) = k_f(\mathbf{s}) - \mathbf{k}_f(\mathbf{s}, X_{n+1})^\top K_{f,n+1}^{-1} \mathbf{k}_f(\mathbf{s}, X_{n+1}), \tag{4.8}$$

where $X_{n+1} = [\, X_n \ \mathbf{x}\,]$, and $K_{f,n+1}$ is the covariance matrix of X_{n+1}. Eq. (4.7) is the IMSE criterion and always nonnegative (Proposition 1 and 2 in [61]) and can be simplified, by Eq. (4.8), as

$$\Delta\mathbb{V}_{\Omega_S}(\hat{f}_n|\mathbf{x}) = \int_{\Omega_S} \mathbf{k}_f(\mathbf{s}, X_{n+1})^\top K_{f,n+1}^{-1} \mathbf{k}_f(\mathbf{s}, X_{n+1}) \mathrm{d}\lambda(\mathbf{s}), \tag{4.9}$$

which is to be maximized.

Unfortunately, Eq. (4.7) cannot be used directly, since the safe region is unknown a priori. Thus, we stick to our predictor $\hat{h}_n$ to estimate the safe region as

$$S_n = \{\mathbf{x} \in \Omega \colon \ \hat{\mu}_n^h(\mathbf{x}) + \beta_n \hat{\sigma}_n^h(\mathbf{x}) < \xi\}, \tag{4.10}$$

where $\hat{\mu}_n^h(\mathbf{x})$ and $\hat{\sigma}_n^h(\mathbf{x})$ are the mean and standard deviation of $\hat{h}_n(\mathbf{x})$, and

$$\Phi(\beta_n) = \mathbb{P}\left(\hat{h}_n(\mathbf{x}) < \xi\right) = 1 - \gamma_n, \tag{4.11}$$

where $\gamma_n \in (0, 1)$. That is, β_n is related to the failure probability of $\mathbf{x} \in S_n$. By constraining our choice of next design point $\mathbf{x} \in S_n$, Eq. (4.7) can be written as

$$\Delta\mathbb{V}_{\Omega_S}(\hat{f}_n|\mathbf{x}) = \mathbb{V}_{\Omega_S \setminus S_n}(\hat{f}_n) + \Delta\mathbb{V}_{S_n}(\hat{f}_n|\mathbf{x}). \tag{4.12}$$

The first term of Eq. (4.12) indicates the irreducible variance induced by the discrepancy between Ω_S and S_n, while the second term is the reducible variance in the estimated safe region. It implies that a consequence of adopting S_n instead of Ω_S with extremely low γ is underestimating the expected variance reduction of $\mathbf{x}$ over Ω_S. In order to extend the purview of variance reduction by $\mathbf{x}$ in Eq. (4.12), we may consider another safe region, called the progressive safe region, which has a more generous safety level than S_n as

$$S_n^+ = \{\mathbf{x} \in \Omega\colon\ \hat{\mu}_n^h(\mathbf{x}) + \beta_n^+ \hat{\sigma}_n^h(\mathbf{x}) < \xi\}, \tag{4.13}$$

where $\Phi(\beta_n^+) = \mathbb{P}\left(\hat{h}_n(\mathbf{x}) < \xi\right) > 1 - \gamma_n^+$ of which $\gamma_n^+ < \gamma_n$. Straightforwardly, we have $S_n \subset S_n^+ \subseteq \Omega$. By considering S_n^+ as the reference set for the integrand, Eq. (4.8), we can reduce the irreducible variance in the first term. Consequently, we have the following acquisition function

$$J_f(\mathbf{x}) = \Delta\mathbb{V}_{S_n^+}(\hat{f}_n|\mathbf{x}), \tag{4.14}$$

where $\mathbf{x} \in S_n$.

Eq. (4.12) shows that the choice of γ_n for S_n affects the learnability of active learning and safety. More explicitly, S_n needs to be conservative to prevent failure by setting γ_n small

enough, while a conservative setting of S_n will increase the irreducible variance term in Eq. (4.12) and reduce the explorable region. Therefore, a promising choice of γ_n should consider the capacity to afford failures, and the following proposition can be prescribed.

Proposition 4.1 (Failure Probability)**.** *For N-sampling budget and any $\zeta \in (0,1)$, choosing design points $\mathbf{x}_i$'s from the safe region S_i for $i \in \{n+1, \ldots, n+N\}$ has the failure probability as*

$$\mathbb{P}\left(\bigcup_i \left(\hat{h}(\mathbf{x}_i) \geq \xi : \mathbf{x}_i \in S_{i-1}\right)\right) \leq \zeta,$$

where S_i of which $\beta = \Phi^{-1}(1-\zeta/N)$ for $\forall\ i$.

The proof is provided in Appendix A.1. For the progressive safe region, increasing γ_n^+ will reduce the unconsidered safe region $\Omega_S \setminus S_n^+$, while it can also simultaneously increase the variance in $\Omega_F \setminus S_n$, which is meaningless. Therefore, S_n^+ also need not be defined as too generous.

Even though J_f is designed carefully with appropriate safe regions, we can minimize the irreducible variance by minimizing the discrepancy between S_n and Ω_S. Generally speaking, J_f only focuses on reducing the variance of $\hat{f}_n$ and does not care about reducing the discrepancy. In order to efficiently expand the explorable region and improve the estimation accuracy of the safe region, we need to incorporate the information from h as well as f in the information criterion. In the following section, we illustrate safe region expansion that focuses on the estimation of safe region boundaries.

4.2.3 Safe Region Expansion

Safe region expansion is required to reduce the error induced by the mismatch of S_n and Ω_S and to furnish higher confidence in exploration. To expand the safe region without failure,

we need to exploit the numerically informative output of h to approach the boundary of the safe region from inside thereof, and expansion can be maximized when the design point is closest to the boundary [72]. Based on [73], we incorporate uncertainty of $\hat{h}_n$ and closeness to the boundary with the following criterion:

$$I(\mathbf{x}) = \begin{cases} \eta_n(\mathbf{x})^2 - (\hat{h}_n(\mathbf{x}) - \xi)^2 & \hat{h}_n(\mathbf{x}) \in (\xi - \eta_n(\mathbf{x}),\, \xi) \\ 0 & \text{Otherwise} \end{cases}, \tag{4.15}$$

where $\eta_n(\mathbf{x}) = \alpha\hat{\sigma}_n^h(\mathbf{x})$ of which $\alpha > 0$. $I(\mathbf{x})$ attains its maximum when $\hat{h}_n(\mathbf{x}) = \xi$, which is the case of $\mathbf{x} \in \partial\Omega_S$. Otherwise, it gets additional scores when $\hat{h}_n(\mathbf{x})$ does not exceed the threshold within an acceptable interval. The role of α is to magnify the effect of uncertainty in $I(\mathbf{x})$. Let the expected value of $I(\mathbf{x})$ with respect to $\hat{h}_n$ be $J_h(\mathbf{x})$, which is the acquisition function for safe region expansion, and expressed as

$$J_h(\mathbf{x}) = \eta_n(\mathbf{x})^2 \left(\hat{\Phi}_n^h(\xi) - \hat{\Phi}_n^h(\xi - \eta_n(\mathbf{x}))\right) - \int_{\xi-\eta_n(\mathbf{x})}^{\xi} (h - \xi)^2 \, \hat{\phi}_n^h(h) \, \mathrm{d}h, \tag{4.16}$$

where $\hat{\Phi}_n^h$ and $\hat{\phi}_n^h$ are the cumulative distribution function (CDF) and the probability density function of $\hat{h}_n(\mathbf{x})$. Eq. (4.16) is composed of two terms: the first term is related to uncertainty, and the second term is related to closeness to the boundary. Consequently, maximizing Eq. (4.16) leads to sampling near the boundary with high uncertainty if such points exist, and the most uncertain point otherwise.

Obviously, the interests of J_f and J_h are inherently different, since they are associated with different mechanisms, f and h. Also, they are formulated for different purposes. It implies that the safe approximation of the target function and the safe region expansion have tradeoffs, thus we need to compromise between both criteria to determine the most informative design point. We discuss the framework for addressing the balance between the

criteria in the next section.

4.2.4 Harmonizing Acquisition Functions

In this section, we integrate two acquisition functions to optimize (maximize) them judiciously to achieve safe active learning. Two distinct acquisition functions are proposed to accomplish different objectives, and optimization with two criteria is a MOO problem. Conceptually, we may think of a point $\mathbf{x} \in S_n$ that achieves the maximum of each criterion simultaneously, called the utopia point. However, MOO typically has no single optimal solution contrary to usual single-objective optimization. Therefore, the Pareto optimality concept is mostly referred to define the optimality in this regime. Let $\mathbf{J}(\mathbf{x}) = [\ J_f(\mathbf{x})\ J_h(\mathbf{x})\ ]^\top$, then Pareto optimality and its weaker version are defined as follows.

Definition 4.2 (Pareto Optimal)**.** A point, $\mathbf{x}_* \in \Omega$, is Pareto optimal if and only if there does not exist another point, $\mathbf{x} \in \Omega$, such that $\mathbf{J}(\mathbf{x}) \leq \mathbf{J}(\mathbf{x}_*)$, and $J_i(\mathbf{x}) < J_i(\mathbf{x}_*)$ for at least one of $i \in \{f,\ h\}$.

Definition 4.3 (Weakly Pareto Optimal)**.** A point, $\mathbf{x}_* \in \Omega$, is weakly Pareto optimal if and only if there does not exist another point, $\mathbf{x} \in \Omega$, such that $\mathbf{J}(\mathbf{x}) < \mathbf{J}(\mathbf{x}_*)$.

Note that inequalities in the definitions associated with vectors stand for element-wise inequality. Obviously, every Pareto optimal point is weakly Pareto optimal, while the reverse is not true.

It is common to scalarize the vector-valued objective functions in MOO, and the formulation of the problem is critical for the Pareto optimality of the solution. In this chapter, we use the weighted sum, which is widely used, to scalarize our criteria with the integrated acquisition function:

$$J(\mathbf{x}) = \left((1-w)J_f(\mathbf{x})^p + wJ_h(\mathbf{x})^p\right)^{1/p}, \quad w \in [0,1], \tag{4.17}$$

where $p \in \mathbb{N}$. The weight parameter w in Eq. (4.17) exactly conveys the preference of the decision maker between two objectives. For example, if one is more interested in the safe region expansion, the decision maker will weigh more on J_h, and decrease w when the estimated safe region seems acceptable. Otherwise, some may begin with small w to see if the safe region expansion is necessary. It turns out that the integrated acquisition function guarantees the Pareto optimality of its solution given w as shown in the following proposition.

Proposition 4.4 (Sufficient Pareto Optimality, [74]). *For any $w \in (0, 1)$, a solution that maximizes the integrated criterion is Pareto optimal associated with w. When $w = 0$ or $w = 1$, a solution of the integrated criterion is weakly Pareto optimal associated with w.*

However, J_f and J_h may be different in their scales, so the weight parameter w cannot be determined straightforwardly. It is common to transform component objective functions in the formulation of MOO, so we may normalize each criterion as

$$\overline{J_i}(\mathbf{x}) = \frac{J_i(\mathbf{x}) - \min J_i}{\max J_i - \min J_i},$$

where minimum and maximum of J_i for $i \in \{f, h\}$ stand for the minimum and maximum over S_n. To normalize objective functions, we need their maxima and minima, so we provide two scaling options in this chapter. The first scaling method is for global searching in a dense grid over the design space. By discretizing the design space into a dense grid, we may evaluate all criteria over the grid. It can provide the heuristically global optimal solution, and make scaling more consistent. Another scaling method is to incorporate the lower and upper bounds of criteria. We already have that both criteria are nonnegative, so their lower bounds are zero. For the upper bounds, J_f is upper bounded by

$$\mathbb{V}_{S_n^+}(\hat{f}_n) = \int_{S_n^+} \mathrm{Var}\left(\hat{f}_n(\mathbf{s})\right) \mathrm{d}\lambda(\mathbf{s}), \tag{4.18}$$

since the expected variance reduction in Eq. (4.14) is nonnegative. Meanwhile, J_h is upper bounded by

$$\sup_{\mathbf{x}\in S_n} \eta_n(\mathbf{x})^2 = \sup_{\mathbf{x}\in S_n} \alpha^2 \hat{\sigma}^h(\mathbf{x})^2 \tag{4.19}$$

from its original formulation. In this way, we can scale both criteria by their tractable bounds that can be obtained prior to the evaluation of every candidate.

The upper bounds of J_f and J_h also imply that the integrated acquisition function asymptotically converges as shown in Proposition 4.5, of which proof is given in Appendix A.2.

Proposition 4.5 (Asymptotic Convergence)**.** *Suppose a nonempty* S_n*. As* $n \to \infty$*,* $S_n \to S_* \subseteq \Omega_S$*, and also* $J(\mathbf{x}|S_n) \to 0$*, for every* $\mathbf{x} \in S_n$*, and any* $w \in [0, 1]$*.*

Proposition 4.5 shows that the integrated criterion J leads our estimators to the best estimators of f and h over a conservative estimation of Ω_S, regardless of the choice of w.

Let us refer to our active learning as PhysCAL (**Phys**ics-**C**onstrained **A**ctive **L**earning), and its pseudocode is provided in Algorithm 2. In practice, integrals in Eq. (4.14) and Eq. (4.16) require numerical methods such as averaging integrand uniformly sampled within integration limits. The computational cost of the algorithm is mostly dominated by the inverse of $K_{f,n+1}$ in Eq. (4.8), which originally takes $\mathcal{O}(n^3)$. Thus, we adopt the rank one Cholesky update in [75] to alleviate the cost to $\mathcal{O}(n^2)$ and to improve the numerical stability. If active learning has no finite candidate pool, we may need a grid or uniform space-filling designs over Ω to realize S_n and S_n^+ and solve line 5 in the algorithm. Note that PhysCAL can be terminated by not only the sampling budget but also the prediction accuracy of the target model when the sampling budget is implicit or early stopping is reasonable. In order to do so, a separate testing dataset or cross-validation is required.

Algorithm 2: Active Learning for Physics-constrained Systems

Data: N(Samplingbudget), $\mathcal{D}$, β, β^+, α, w

Train $\hat{f}$, $\hat{h}$ with $\mathcal{D}$
while $N > 0$ **do**
  Evaluate S_n, S_n^+ over Ω
  $\mathbf{x}_* = \arg\max_{\mathbf{x} \in S_n} \overline{J}(\mathbf{x})$
  Observe y_*, z_* at $\mathbf{x}_*$
  $N = N - 1$
  $\mathcal{D} = \mathcal{D} \cup \{\mathbf{x}_*, y_*, z_*\}$
  Update $\hat{f}$, $\hat{h}$ with $\mathcal{D}$
  Update β, β^+, α, w (Optional)

4.3 Simulation Study

In this section, we apply our active learning to the approximation of a constrained 2-D simulation function. The response surface of the target function is defined over $\Omega = [-0.5, 0.5]^2$, which is

$$f(\mathbf{x}) = |x_0\, x_1|, \tag{4.20}$$

where $\mathbf{x} = [\, x_0 \; x_1 \,]^\top \in \Omega$, and the constraint function is

$$h(\mathbf{x}) = (\cos(2\pi x_0) - \cos(2\pi x_1))^2 - 0.8 \exp\left(|x_0\, x_1|\right), \tag{4.21}$$

which is assumed to be implicit. Both functions and the failure region are illustrated on the left of Fig. 4.2. We set the safe region as $\Omega_S = \{\mathbf{x} \in \Omega | \, h(\mathbf{x}) < 0.7\}$, thereby the failure region ratio to the design space being approximately 0.28. Assuming we have no prior knowledge of the safe design settings, 10 initial samples are obtained over the design space using the maximin LHD, which yields 2–3 samples from the failure region in 10 replications. Observations from f and h are corrupted by Gaussian noise, and additional 20 samples are obtained with active learning. The parameters of each method were fixed in this study, so

we omit the subscription n in the parameters. For PhysCAL, we set $\gamma = 0.001/20$, which is less than the defective percentage of six sigma in statistical process control, and $\gamma^+ = 0.01$ to extend the considered region in safe variance reduction to the progressive safe region with the failure probability of 0.99. As the benchmark method, safe exploration for GP (SEGP) in [21] is considered. For the other parameter settings, $\alpha = 2$ according to [73] and $w = \{0.0, 0.1, \ldots, 0.9, 1.0\}$ are used.

During the simulation, setting the safety level in SEGP as high as ours was impossible in some replications due to the insufficient number of failure samples to make a nontrivial explorable region, which was also mentioned in their work. Furthermore, we observed that the benchmark method failed to estimate a meaningful failure region as shown at the right bottom of Fig. 4.2. These issues can be explained as follows. First, the dataset is imbalanced due to the larger safe region, so it resulted in unsatisfactory classification, which underestimates the failure region. Second, the nuisance function of SEGP is adapted for binary classification via the Laplace approximation [53], so the predictive variance induced by the nuisance function is inadequate for the safe region estimation as ours. More explicitly, encoding the failure process observation into a binary class discards numerical information from the original output (i.e., the response of the failure process). Consequently, the GP classifier determines the failure only based on the spatial input in disregarding of output's numerical information, thereby inducing improper predictive variance.

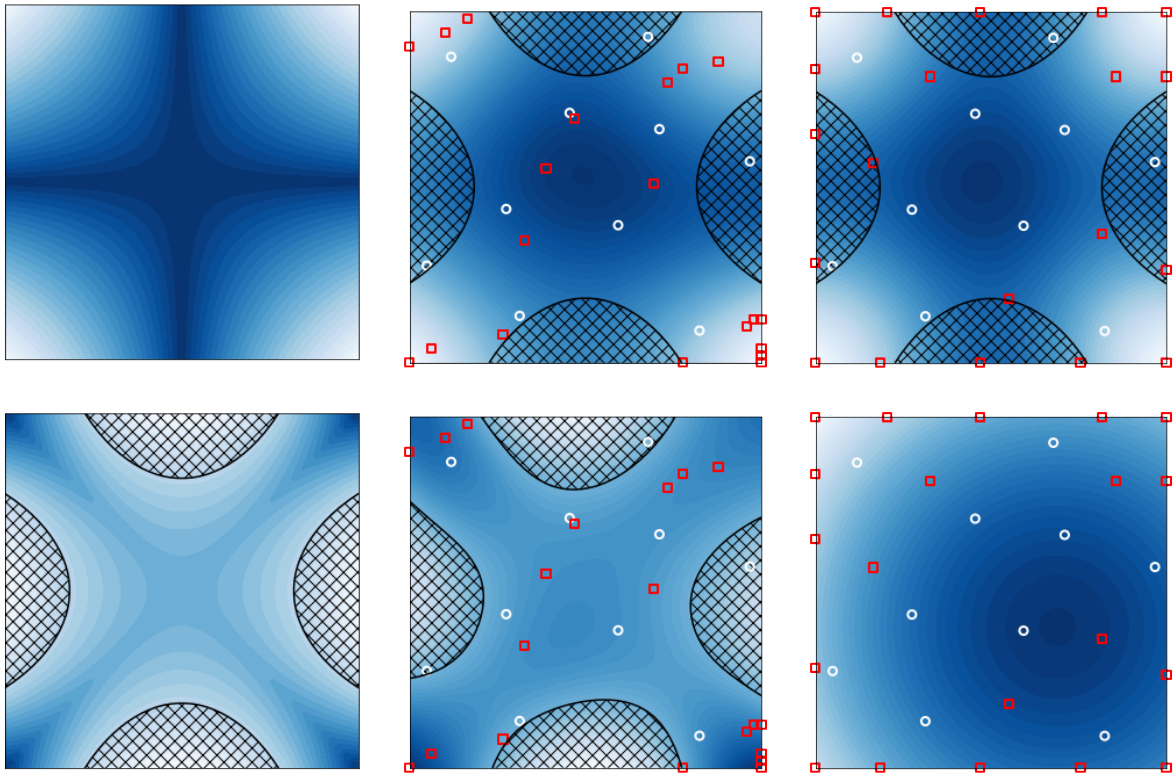

Figure 4.2: Simulation study result. (Left) Ground truths of the target (top) and constraint (bottom) functions. (Center) PhysCAL's (proposed method) estimation of the target (top) and constraint (bottom). (Right) Benchmark method's [21] estimation of the target (top) and constraint (bottom). Hatched regions in the top figures are the true failure region, and the bottom of (Center) is the estimated failure region. Note that the bottom of (Right) has no estimated failure region. White circles are initial design points, and red squares are sampled with active learning.

As a result, PhysCAL achieved the prediction error (MSE) of 0.0023 with 0.5 additional number of failures on average, and there were five zero-failure cases of ten replications. Meanwhile, the benchmark method achieved a better prediction accuracy with 0.001, while the average number of additional failures was 6.7 among 20 queries with no case of zero failure. It is not surprising that the benchmark method did better in target process prediction since exploration was unrestricted by underestimating the failure region. The right of Fig. 4.2 shows that the classifier learned that the center region is safe (with dark blue), while it

could not discriminate the failure region due to the aforementioned reasons. Consequently, the entropy-based strategy in SEGP led to evenly distributed sampling as their property [50]. Meanwhile, our method estimated the failure region much better and explored more safely as shown at the center bottom of Fig. 4.2. Hence, in the case that a single failure is very crucial, our approach will be more suitable.

Fig. 4.3 shows the performance of PhysCAL with different weight parameter settings. We can observe that a low-weight parameter may improve the prediction accuracy by focusing more on variance reduction, while it does not necessarily make our approach safe due to a lack of knowledge in the failure region. Meanwhile, setting w too high also induced an increased number of failures and low predictive accuracy due to indifference to variance reduction in the target approximation. In this simulation, $w = 0.4$ was the best choice among considered values with promising predictive accuracy and the least number of failures.

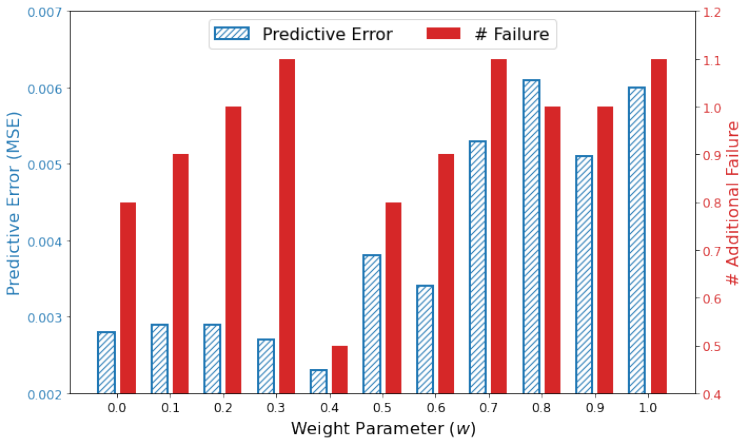

Figure 4.3: Performance of PhysCAL associated with different weight parameters in the simulation study.

4.4 Case Study

In this section, the proposed method is applied to the predictive modeling of composite fuselage deformation in the aerospace manufacturing process. Composite materials such as carbon fiber reinforced polymers are extensively applied to various domains (including aerospace, automotive, construction, and energy) due to their versatility, high strength-to-weight ratio, and corrosion resistance. However, composite materials are nonlinear and anisotropic due to their structural natures [55], so flexible models and the adaptive DOE are required to predict their deformation precisely. Moreover, since they are also subject to structural failures, manufacturers should avoid unsafe load settings in the fabrication of composite structures.

Aerospace manufacturing considered in this case study is subject to dimensional deviations at the joint rim of composite fuselage sections due to the multi-batch manufacturing system. Therefore, to assemble fuselage sections, the shape control procedure is required to reshape them homogeneously. In the shape control procedure, the fuselage section is placed on the supporting fixture, and ten equispaced hydraulic actuators are introduced to reshape the fuselage as shown in the left figure of Fig. 4.4. A promising approach for the optimal shape control procedure is to employ a highly precise predictive model of composite fuselage's deformation instead of the time-consuming physics-based model. However, the construction of such a model validated with real experiments is very challenging due to the expensive cost of sampling, and the risk of structural failure because of unsafe loads. Especially, since composite failures in the shape control process may result in the disposal of nonconforming parts or delayed delivery due to the restoration, up-to-date applications have restricted viable actuator forces in a conservative manner to prevent structural failures [56, 76]. It may lead to a suboptimal shape control, so our objective is to extend the feasible actuator forces that

may include structural failure settings, thereby providing higher degrees of freedom to the shape control.

There are different types of failures in composite materials such as fiber, matrix, and ply failures. Likewise, a number of composite material failure criteria (e.g., Tsai–Wu, Tsai–Hill, Hoffman, Hashin) are devised for different modes of failure [77]. In this chapter, we considered the Tsai–Wu criterion, which is one of the most widely used interactive failure criteria. Note that it is possible to consider more than one criterion simultaneously by taking the most parsimonious criterion, or considering the intersection of safe regions defined by multiple constraint GP models. Briefly, the Tsai–Wu criterion considers interactions between different stress components in addition to the principal stresses (in a homogeneous element). Using the principal material coordinate system on the cubic element of composite material, consider three directions: 1 is the fiber direction; and 2 and 3 are directions perpendicular to 1, respectively. Let σ_i^T and σ_i^C be the tensile failure stress and the compressive failure stress in $i \in \{1, 2, 3\}$ direction, and τ_{12}^F be the shear failure stress in the 12 plain. The Tsai–Wu criterion is defined as

$$\left(\frac{1}{\sigma_1^T} - \frac{1}{\sigma_1^C}\right)\sigma_1 + \left(\frac{1}{\sigma_2^T} - \frac{1}{\sigma_2^C}\right)\sigma_2 + \frac{\sigma_1^2}{\sigma_1^T\sigma_1^C} + \frac{\sigma_2^2}{\sigma_2^T\sigma_2^C} + \left(\frac{\tau_{12}}{\tau_{12}^F}\right)^2 - \frac{\sigma_1\sigma_2}{\sigma_1^T\sigma_1^C\sigma_2^T\sigma_2^C} \geq 1,$$

where the left-hand side is the nonnegative criterion value, and the failure occurs when it exceeds one.

The Tsai–Wu criterion value induced by the shape adjustment solved by the FEM is shown in the right of Fig. 4.4. We can observe that failures occur at the bottom of the fuselage since fixtures that sustain the fuselage are restricting its deformation. Not only limited to the Tsai–Wu criterion, but physics constraints also have many assumptions such as homogeneity, absence of higher-order interactions, etc., although they are convincing apparatuses

to consider structural reliability. Hence, they are typically utilized with the safety-of-margin (the reciprocal of the acceptable failure criterion) or UQ to prevent unexpected failures. Likewise, the safe shape control system should consider the failure criterion not only its value but also the additional safety measures.

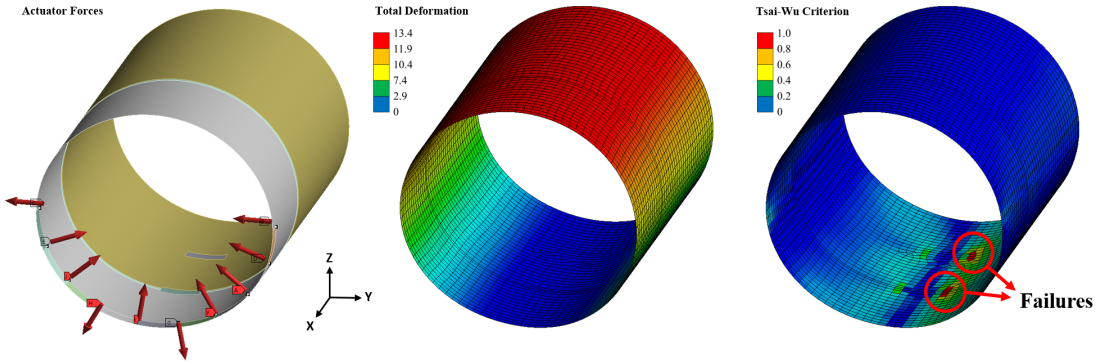

Figure 4.4: Shape control of composite fuselage in the FEM. (Left) Actuator input. (Center) Deformation. (Right) Tsai–Wu criterion.

4.4.1 Experiment Settings

A well-calibrated FEM simulator of the procedure is used as our oracle considering the risk of a real experiment. The target function's input is the vector of unidirectional forces (in lbf) of ten actuators, and the output is Y, Z-directional deformation (in microinch) of the fuselage at one of 91 critical points around the rim. The maximum magnitude of actuator force is 1,000 lbf, which may cause failures in the structure (see Fig. 4.4). As our additional safety measure, the margin of safety with the Tsai–Wu criterion is set at 1.25 (i.e., the acceptance criterion is 0.8).

For the initial design, the maximin LHD is used to generate 20 observations for both deformation and failure criterion, and additional 20 samples are queried by different methods: random, ALM, ALC, SEGP, and PhysCAL. We adopted the pool-based scenario in this case

study by providing a 400-size of candidate pool that uniformly spreads out the design space. Considering the variability in the initial design, we generated ten initial datasets independently and replicated the experiment. It is noteworthy that we do not cease active learning, even though we encounter a failure in the construction of the predictive model for this case. In practice, a composite failure is definitely an undesirable event, while it is feasible as far as the design point is within the input space (i.e., the maximum actuator forces). Therefore, failure events are also included in the training dataset, and we compare the number of additional failures in learning to evaluate how well each method avoids failures.

For PhysCAL, we also considered different weight parameters as the simulation study, and set other parameters as $\alpha = 2$, $\gamma = 0.001/20$, and $\gamma^+ = 0.1$. In SEGP, we reduced the safety level of which from the PhysCAL until that SEGP induced a nonempty explorable space. For the model evaluation, we used 100 safe samples as the testing dataset that is independently generated with the candidate pool, and the MAE is used as the metric.

4.4.2 Results

The results are summarized in Table 4.1. First, we can observe that PhysCAL outperforms other methods in the number of additional failures. It achieves nine zero failures from ten cases. Also, we can observe that ALM, ALC, and SEGP incurred more failures than the random. The reason is that the design space is almost dominated by the safe region, while the failure region may be more interesting than elsewhere. Interestingly, SEGP is inapplicable in this case when the initial dataset does not contain failure samples, since the method uses the binary classifier. It implies that employing a regressor as the constraint model is more advantageous when the prior information has no failure.

Table 4.1: Result of the case study.

Case	Random		ALM		ALC		SEGP [21]		PhysCAL	
#	MAE	# Fail	MAE	# Fail	MAE	# Fail	MAE	# Fail	MAE	# Fail
1	6.303	1	2.573	4	2.572	4	N/A		2.640	0
2	10.135	0	2.572	4	2.572	3	N/A		2.639	0
3	7.975	1	1.628	5	1.635	4	N/A		3.530	0
4	2.368	0	2.862	4	2.143	4	N/A		3.667	0
5	2.427	1	2.161	4	1.939	5	N/A		2.465	0
6	4.570	0	2.572	4	2.572	3	N/A		2.639	0
7	6.163	0	2.344	4	2.631	4	N/A		2.387	0
8	6.704	1	1.427	4	1.272	4	1.892	0	3.487	0
9	3.326	0	2.009	4	1.900	3	2.631	1	2.286	0
10	4.602	0	1.970	4	1.583	4	2.369	2	2.585	1
Mean	5.383	0.4	2.244	4.1	2.046	3.8	2.297	1.0	2.832	0.1
(Std.)	(2.345)	(0.5)	(0.470)	(0.3)	(0.441)	(0.6)	(0.305)	(0.8)	(0.518)	(0.3)

In terms of prediction accuracy, PhysCAL performs much better than random sampling and is comparable to other active learning approaches considering the scale of the metric. We can conjecture that other methods are able to observe from the failure region that may be informative, so their accuracy is the consequence of unsafe exploration. Furthermore, PhysCAL is more flexible than other methods, since we may update the weight parameter in PhysCAL during data acquisition to focus more on variance reduction as well as other methods. Therefore, PhysCAL is more promising for this case considering the risk of failure in the system.

4.4.3 Weight Parameter

Different weight parameters are considered for PhysCAL in the case study, and the performance of different weight parameters is provided in Fig. 4.5. Likewise, only focusing on one acquisition function is not optimal in this case, and the performance is better with $w = 0.7 \sim 0.9$, which is higher than the simulation study. The possible reasons are as follows. First, the failure region, in this case, is much smaller than the safe region, thus a quite aggressive exploration may be acceptable (i.e., increasing w). Second, the actuator force is positively correlated with both von Mises stress, which is linearly correlated with the Tsai–Wu criterion, and deformation [64]. Consequently, the uncertainty term of safe region expansion could be informative to model variance reduction.

4.4.4 The Margin of Safety

In order to observe the effect of the margin of safety, we increased the value from 1.25 to 1.5, which reduces the acceptable Tsai–Wu criterion to 0.66. Although we may consider the margin of safety higher than 1.5, such a high value is irrelevant in practice. Increasing the margin of safety yields the increment of failure region, so it results in more failures. In the average of 10 replications, PhysCAL got 0.5 additional failures (six zero-failures), which is the least among the considered methods, and achieved predictive accuracy of 3.74 microinches. Random sampling got 0.6 additional failures and achieved the predictive error of 5.16 microinches. Meanwhile, ALM and ALC achieved 8.3 and 7.9 failures, respectively. Likewise, SEGP was applicable only for the last three cases and induced 3.3 failures from those cases.

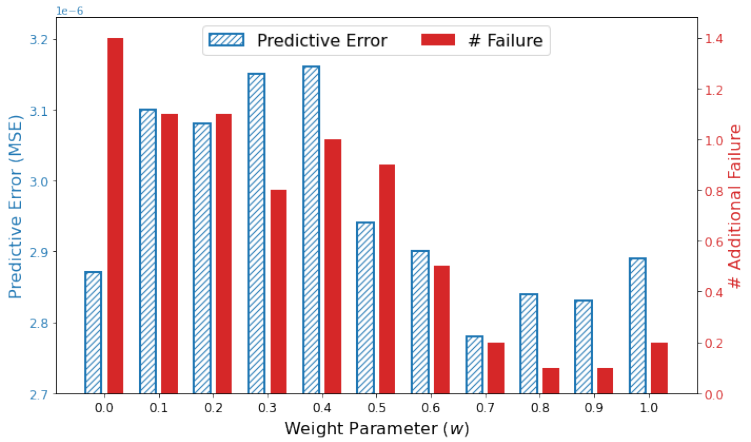

Figure 4.5: Performance of PhysCAL associated with different weight parameters in the case study.

4.5 Summary

For physics-constrained systems that are expensive to evaluate, failure-averse active learning is proposed in this chapter. In order to achieve safe active learning under the regime of implicit physics constraints, GP priors are imposed on the target function and physics constraints, and two acquisition functions are developed for safe variance reduction and safe region expansion. For safe variance reduction, two safe regions with different safety levels are employed in the IMSE criterion, thereby maximizing the safety and variance reduction over the underlying safe region. For the safe region expansion, the acquisition function is devised to sample near the safe region boundary considering uncertainty. Two acquisition functions are endowed with different objectives, so the MOO framework with Pareto optimality is applied to integrate them into the flexible global criterion. The integrated acquisition function is sufficient for the Pareto optimality of the design point to be queried and can be flexibly adjusted by the decision maker's preference considering the tradeoff between two acquisition functions. Also, it is shown that the integrated acquisition function asymptotically leads to

the best estimation of the system.

In the simulation study, the proposed approach showed promising performance with the achievement of zero failure, while the benchmark method failed to avoid failures in its learning process. Furthermore, with different parameter settings, we empirically observed that safe variance reduction and safe region expansion should be involved simultaneously for better predictive accuracy and higher safety. Our method has also shown remarkable performance in the predictive modeling of composite fuselage deformation considering its structural failure with the Tsai–Wu criterion. It achieved zero failure in most cases, while other benchmark methods induced more failures or inferior predictive accuracy. Our proposed method is adaptive since it can incorporate domain knowledge and the decision maker's preference with amenable parameters. Therefore, it is also applicable to other domains that are subject to implicit constraints.

Chapter 5

Multi-output Extreme Spatial Model for Composite Fuselage Assembly

Data-driven models in machine learning have enabled efficient management of manufacturing systems. However, a majority of machine learning models are devoted to modeling the mean response or average pattern, which is inappropriate to study abnormal extreme events that are often of the main interest. Since extreme events from heavy-tailed distributions give rise to prohibitive expenditures in system management, relevant extreme models are urgently needed to analyze extreme risks. Engineering applications of extreme models usually focus on individual extreme events, which is insufficient for complex systems with correlations. Moreover, existing extreme spatial models in other domains cannot be directly applied to controllable systems. In this chapter, we propose an extreme spatial model that facilitates efficient modeling of multi-output response control systems with a bilinear function on two spatial domains for control variables and measurement locations. Marginal parameter modeling and extremal dependence are investigated. In addition, an efficient graph-assisted composite likelihood estimation and corresponding computational algorithms are developed to cope with high dimensional outputs. Its application to composite aircraft assembly shows that the proposed model enables comprehensive analyses with superior predictive performance on extreme events to canonical methods.

5.1 Introduction

Manufacturing systems in modern industries encounter demands for unprecedented precision and adaptiveness in processing due to the increased complexity of products and diversified customers' preferences. It leads to increments in control and measurement fidelities for quality assurance. Owing to intrinsic uncertainties in manufacturing systems, statistical approaches such as statistical process control (SPC) [78] and Stream-of-variation (SoV) [79] have been extensively applied to production system management and facilitated remarkable progress in various applications (e.g., tolerance synthesis, system design, sensor allocations, quality management). However, a majority of statistical models in manufacturing systems focus on central tendencies (e.g., mean, median, average pattern) and resort to Gaussian distributions, which is inappropriate for extreme events. Moreover, many defects and anomalies of interest arise from tail distributions as extreme events, and modeling tail distributions is intractable when the underlying distributions are non-Gaussian.

Managing extreme risks is critical but exceptionally challenging in composite aircraft manufacturing. To be more specific, composite fuselage assembly in aircraft production involves shape adjustment because of inevitable dimensional deviations in fuselage sections, while the shape control process is subject to undesirable structural failures from exceptionally large stress and various uncertainties. Data-driven models are promising to automate the process [56, 76] with predictive uncertainty. However, modeling extreme values thereof is difficult because the amount of extreme data is too small owing to the small production in aircraft manufacturing.

Extreme risks characterized as tail distributions have been of great interest to system management and decision analysis [80, 81], and extreme value analysis is distinguished from typical statistical analysis in several aspects. First, statistical inference in extreme value

models is subject to data scarcity since only a small subset of observations are related to extreme values. Second, extreme value models should be able to forecast further than observed values. It implies that extreme value modeling is not interpolation but extrapolation. Third, extreme value modeling is highly intractable. Extreme value models' behaviors are highly sensitive to their shape parameters of which inference is challenging because of the aforementioned characteristics. Also, modeling multivariate extreme values is mathematically complicated and computationally demanding. All these distinctions of extreme values and their significance in the real world have led to the development of a focused theory, called extreme value theory (EVT) [82, 83].

Extreme values in composite fuselage assembly over a continuous design space can be prescribed with extreme spatial models, powered by EVT and mainly studied in environmental statistics [44]. However, they cannot be directly applied to composite fuselage assembly owing to the different characteristics of the domains. More explicitly, aircraft manufacturing includes dynamic systems in which components are discrete states and control variables, but extreme events in environmental problems are continuous and uncontrollable. Furthermore, because of the high sensitivity of the shape parameter in extreme models, conventional extreme spatial models in spatial statistics assume the shape parameter is constant over spatial domains. However, this assumption may be inappropriate in fuselage assembly since system characteristics can be changed by control inputs [75]. Lastly, monitoring the assembly process involves high-dimensional data to maintain high quality, but modeling and simulating high-dimensional extreme values are technically challenging.

To cope with the limitations of existing extreme spatial models for composite fuselage assembly, we propose the multi-output extreme spatial model that enables extreme modeling of dynamic systems and the multi-output response thereof. The proposed model builds marginal parameter models for marginal distributions on the input space and a max-stable

process to model dependency between output variables. We define a proper metric space with geometric locations of critical points, where key characteristics of products are measured. The metric space enables the use of not only the max-stable process but also a graph structure of critical points. The undirected graph facilitates incorporating background knowledge and improving computational efficiency and estimation performance by utilizing truncated composite likelihoods. The contributions in this chapter can be summarized as follows.

1. A multi-output extreme spatial model for controllable manufacturing systems is proposed. The proposed model enables modeling multivariate extreme values conditional on control variables in the continuous design space.

2. Graph-assisted composite likelihood estimation is developed for high-dimensional problems. The graph structure facilitates the incorporation of background knowledge and geometric locations of critical points.

3. End-to-end cutting-edge extreme modeling of composite fuselage assembly is demonstrated, which is innovative. The modeling process and application extensions for the usage in general manufacturing systems are provided.

Extreme modeling is more advantageous than other risk assessment-related methods such as value at risk, importance sampling, and quantile estimation when the underlying distribution is not a nominal distribution. These methods rely on underlying tail distributions that are difficult to estimate without extreme value principles. Meanwhile, extreme models may reduce the risk of model misspecification in tail distribution modeling owing to the max-stable property. More details are provided in Section 5.2.1.

The remainder of this chapter is organized as follows. In Section 5.2, fundamental EVT and extreme spatial models are provided. Thereafter, we propose the multi-output extreme spatial model, including marginal parameter modeling, extremal dependence investigation,

graph-assisted composite likelihood estimation, and computational algorithms. Section 5.3 illustrates the application of the proposed method to extreme residual stress modeling in composite aircraft production. At last, the summary of this chapter and further applications of the method are discussed in Section 5.4.

5.2 Methodology

To illustrate the multi-output extreme spatial model, we briefly introduce EVT and extreme spatial models. A mathematical formulation of manufacturing systems and the multi-output extreme spatial model will be discussed afterward. For notations, capital letters denote functions, random variables, either univariate or multivariate, and the dimension. The dimension will be specified, e.g., $X \in \mathbb{R}^D$, and square brackets indicate the set of indices, e.g., $[D] = \{1, \ldots, D\}$. Parenthesized superscripts indicate indices of i.i.d. samples, and a component index in a vector and a sequence will be denoted by corresponding lower letters.

5.2.1 Extreme Value Models

Consider a random variable X with a distribution F_X, and let $\{X^{(1)}, \ldots, X^{(n)}\}$ be i.i.d. samples from X. Extreme value theory concerns the limit value (i.e., maximum or minimum) of the random variable. For clarity in exposition, we confine our interest to the maximum without the loss of generality hereafter. The limiting maximum value of X can be shown as

$$\lim_{n\to\infty} \frac{\max\{X^{(1)}, \ldots, X^{(n)}\} - b_n}{a_n} \xrightarrow{D} Y,$$

for some $b_n \in \mathbb{R}$ and $a_n > 0$ for $n \in \mathbb{N}$. Given that Y has a nondegenerate distribution, the distribution is a member of

$$F_Y(y) = \exp\left[-\left\{1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right\}^{-1/\xi}\right], \tag{5.1}$$

which is defined on $\{y\colon 1 + \xi(y-\mu)/\sigma > 0\}$, where $\xi, \mu \in \mathbb{R}$ and $\sigma > 0$. The distribution family of Eq. (5.1) has three parameters and is called the generalized extreme value (GEV) family. The GEV family subsumes three subfamilies with different tail decaying behaviors according to the shape parameter ξ: (i) Weibull ($\xi < 0$); (ii) Gumbel ($\xi = 0$); and (iii) Fréchet ($\xi > 0$). An important property of the GEV family is that only they are max-stable, which is defined to have some constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$F_Y^n(a_n y + b_n) = F_Y(y),$$

for every $n \in \mathbb{N}$. Generally, any distribution in the GEV family is transformed to unit Fréchet with

$$Z = \left\{1 + \xi\left(\frac{Y-\mu}{\sigma}\right)\right\}^{1/\xi} \sim \mathrm{GEV}(1,1,1)$$

for convenience.

Multivariate extreme values where each component is transformed to unit Fréchet can be modeled as

$$F_Z(Z_1 \le z_1, \ldots, Z_Q \le z_Q) = \exp\left\{-V(z_1, \ldots, z_Q)\right\}, \tag{5.2}$$

where V is called the exponent measure, which determines the joint distribution. Multivariate extreme modeling can be achieved by estimating the exponent measure, which may be

either parametric or nonparametric, satisfying

$$V(\infty, \ldots, z_q, \ldots, \infty) = 1/z_q, \qquad V(tz_1, \ldots, tz_Q) = t^{-1}V(z_1, \ldots, z_q),$$

for any $t > 0$ [84]. Modeling extreme value dependency is one of the most challenging parts of extreme value modeling, and not all models can accommodate either dependence or independence. Therefore, modeling multivariate extreme values is subject to model misspecification, and this is one of the active research areas in extreme modeling. Furthermore, Eq. (5.2) requires every partial derivative of the exponent measure for the likelihood estimation, which is computationally prohibitive with respect to the dimension Q. Hence, composite likelihood inference is usually used for high-dimensional problems.

Extreme spatial models are random processes defined over spatial domains in which marginals are extreme value distributions. Generally, an extreme spatial model is built upon a smoothly varying model of marginal GEV parameters. However, it cannot address concurrent events at multiple locations in a spatial domain, so several copula models are proposed to introduce dependency [85]. Max-stable processes are valid extreme spatial models since they represent random processes satisfying the max-stable property, and the widely used models include the Brown–Resnick process [86] and the extremal-t process [87]. Consider a spatial domain $\mathcal{S} \in \mathbb{R}^D$, where s is an arbitrary location in $\mathcal{S}$. A max-stable process defined over $\mathcal{S}$ has the spectral representation [88], which is

$$Z(s) = \sup_{i \in \mathbb{N}} \{W^{(i)}(s)/r^{(i)}\}, \tag{5.3}$$

where $r^{(i)}$'s are samples from the unit rate Poisson process on $\mathbb{R}_+$, and $W^{(i)}(s)$'s are from a stationary random process, satisfying $\mathbb{E}\left[\max\{0, W(s)\}\right] = 1$. The random process characterizes $Z(s)$ with the corresponding exponent measure.

5.2.2 Multi-output Extreme Spatial Model

In this section, we illustrate the proposed multi-output extreme spatial model. Consider the composite fuselage assembly process, $\Phi\colon \mathcal{S} \to \mathcal{X}$, where $\mathcal{S} \subset \mathbb{R}^D$, and $\mathcal{X} \subset \mathbb{R}^Q$. The input s can be incoming dimensional deviations, actuator force, or both, and the output X is the resulting state or key characteristics of the fuselage measured at Q critical points. The process can be modeled as

$$X(s) = \Phi(s) + \varepsilon(s), \tag{5.4}$$

where $s \in \mathcal{S}$, and $\varepsilon \in \mathbb{R}^Q$ is observation noise. The asembly process is maintained by controlling the input to satisfy

$$X_q(s) < \kappa_q, \quad q \in [Q] \tag{5.5}$$

where κ_q is a criterion for acceptable product quality. Since the process is stochastic because of not only observation noise but also intrinsic uncertainty in Φ attributed to unmeasurable or uncontrollable features, a general approach to quantify the probability of Eq. (5.5) is modeling Eq. (5.4) with a stochastic spatial model such as GPs. However, the approach may underestimate the risk when X_q is heavy-tailed or non-Gaussian. Therefore, a stringent analysis on Eq. (5.5) can be achieved when we focus on extreme values of $X_q(s)$.

The max-stable process in Eq. (5.3) is unsuitable to model extreme values of Eq. (5.4), since X is always observed conditional on s. That is, modeling dependency of extreme values of X over $\mathcal{S}$ is unnecessary, and we only need the marginal distribution for each $s \in \mathcal{S}$. Meanwhile, for $q, q' \in [Q]$, where $q \neq q'$, $X_q(s)$ and $X_{q'}(s)$ may be correlated as concurrent events, which should be considered in the model. Furthermore, aircraft manufacturing systems involve a large number of critical points, and it intensifies inferential computations in the max-stable process. The multi-output extreme spatial model addresses these drawbacks, and

it is outlined in Fig. 5.1. The proposed model is in two folds: modeling the marginal parameters and constructing the max-stable process for the multi-output response; and it will be explained in detail in the following sections.

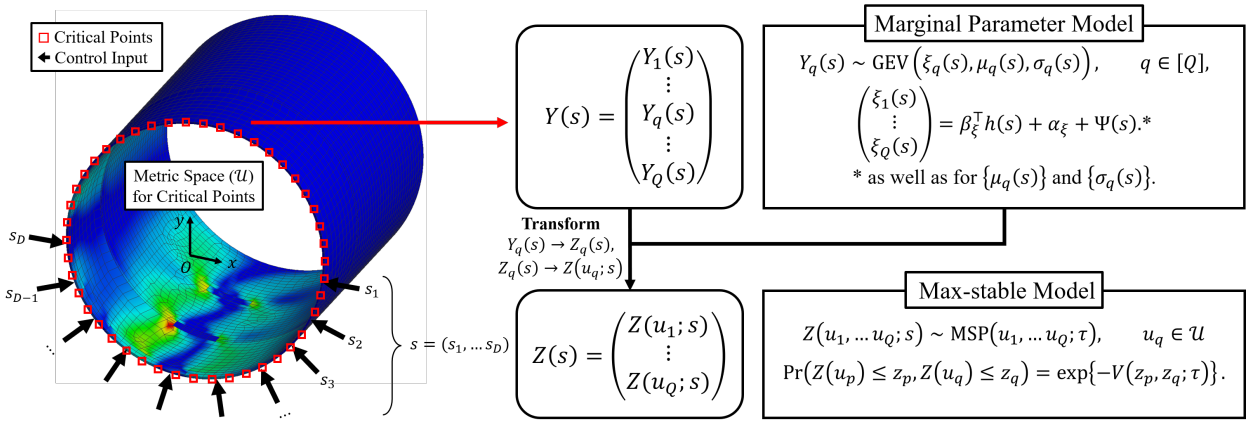

Figure 5.1: Overview of the multi-output extreme spatial model.

Marginal Parameter Modeling

Let $Y(s)$ be the component-wise maxima of $X(s)$, and denote q-th component of $Y(s)$ by $Y_q(s)$. We assume that Y_q is max-stable uniformly over $\mathcal{S}$ for $q \in [Q]$ so that we can use the GEV family to model the marginal distributions of $\{Y(s) : s \in \mathcal{S}\}$. Given that we have L replicated observations from N design points $S = \{s_n : n \in [N]\}$, we may write observations as $\mathbf{X}(s_n) = \{\mathbf{x}(s_n)^{(l)} : l \in [L]\}$ for each $n \in [N]$. Let T be the block size for block maxima, which should be chosen based on backgrounds and inference stability. For example, one year is used as the block size for precipitation data, while, the batch size, gross order, or delivery capacity can be considered in manufacturing systems. Otherwise, for inference stability, the block size can be chosen by looking at the return-level plot [82] with the available data amount. Using block maxima with the block size T, extreme values of size $B = \lceil L/T \rceil$ can be extracted from observations as $\mathbf{Y}(s_n) = \{\mathbf{y}(s_n)^{(b)} : b \in [B]\}$. Note that $\mathbf{Y} \subset \mathbf{X}$. It leads to the estimation of GEV parameters using the maximum likelihood for each pair of

$(n, q) \in [N] \times [Q]$.

Let us denote the vector of each GEV parameter as $\Xi(s_n) = [\xi_1(s_n), \ldots, \xi_Q(s_n)]^\top$, and $\mathrm{M}(s_n), \Sigma(s_n)$, respectively. We may assume that $\xi_q(s_n)$, $\mu_q(s_n)$, and $\sigma_q(s_n)$ are independent for $q \in [Q]$. We interpolate the marginal parameters of $\{\mathbf{Y}(s_n) \colon n \in [N]\}$ with

$$\Xi(s_n; \beta_\xi, \alpha_\xi, \theta_\xi) = \beta_\xi^\top h(s_n) + \alpha_\xi + \Psi(s_n; \theta_\xi), \tag{5.6}$$

where $h : \mathcal{S} \to \mathbb{R}^P$ is a deterministic function of s (e.g., polynomial), $\beta_\xi \in \mathbb{R}^{P \times Q}$, $\alpha_\xi \in \mathbb{R}^Q$ are parameters for linear regression, and Ψ is a Q-dimensional zero-mean GP with hyperparameter θ_ξ. In practice, logarithm should be taken on $\Sigma(s_n)$ to ensure positiveness. Although components of Ξ can be modeled independently via Eq. (5.6), using a multi-output GP [89] can be taken into account since parameters at geometrically close critical points may be correlated. However, the computational complexity of the multi-output GP is $\mathcal{O}(Q^3N^3)$ so it can be prohibitive when Q is large. Therefore, the output dimension and the degree of parametric correlation should be considered in determining Eq. (5.6).

It is also possible to utilize Eq. (5.6) as a latent variable model (LVM) [44] in a hierarchical model as

$$Y_q(s_n) \sim \mathrm{GEV}(\xi_q(s_n), \mu_q(s_n), \sigma_q(s_n)), \quad q \in [Q],$$

where proper priors are imposed on parameters. Although the LVM is straightforward, inference and sampling from the posterior distribution are computationally expensive when the control variable dimension is large as the fuselage assembly. Moreover, adjusting an MCMC method for a good convergence is intractable due to the high dimension of the parameter space. On the other hand, interpolating MLE parameters for S is not hindered by the dimension as in the Bayesian framework. To sum up, the LVM is preferable when the control variable dimension is small, otherwise, interpolating the MLE parameters would

compensate for less computational burdens.

Multi-output Modeling

Although the marginal parameter model is sufficient to model the marginal extreme distribution for any $s \in \mathcal{S}$, we need to consider the extremal dependency of Q components in $Y(s)$ to simulate the extreme spatial model. We use the max-stable process in Eq. (5.3) to model the dependency between key characteristics. Since we have marginal distributions, we can component-wisely transform observations into unit Fréchet vectors, and we denote the transformed output following the max-stable process parameterized with τ by $Z(s) \sim \text{MSP}(s;\ \tau)$.

Extremal Dependence Investigation An important step in extreme spatial modeling is investigating extremal dependence prior to choosing the max-stable process. The extremal dependence in a bivariate case can be empirically measured by

$$\chi = \lim_{u \to 1} \mathbb{P}\left(F_{Z_i}(z_i) > u | F_{Z_j}(z_j) > u\right) \tag{5.7}$$

$$= 2 - V(1,1), \tag{5.8}$$

where Z_i and Z_j have unit Fréchet margins. Eq. (5.7) measures asymptotic dependency between two extreme variables: $\chi = 0$ indicates two variables are asymptotically independent, and $\chi = 1$ means they are totally dependent. To alleviate the model misspecification, the model class should be chosen based on the estimated χ. Conversely, the model can be validated with Eq. (5.8) if the exponent measure can be readily obtained.

To model the intercorrelated $\{Z_q \colon q \in [Q]\}$ with the max-stable process, the critical points should be represented in a spatial domain. Generally, measurement points in a production

system are located on a rigid body, so it is reasonable to consider the geometric adjacency of critical points by defining a proper metric space, denoted by $(\mathcal{U}, d_{\mathcal{U}})$. A straightforward way is to assign the Q critical points to $\{u_q \colon q \in [Q]\} \subset \mathcal{U} \subset \mathbb{R}^P$ with the Euclidean distance $d_{\mathcal{U}}$, where P is typically two or three. Consequently, engaging the metric space associated with critical points transforms the max-stable process to a bilinear function on $\mathcal{S} \times \mathcal{U}$ as

$$\begin{bmatrix} Z_1(s) \\ \vdots \\ Z_Q(s) \end{bmatrix} \coloneqq \begin{bmatrix} Z(u_1, s) \\ \vdots \\ Z(u_Q, s) \end{bmatrix} \sim \mathrm{MSP}(u_1, \ldots, u_Q, s;\ \tau), \tag{5.9}$$

where the exponent measure is

$$\mathbb{P}\left(Z(u_1, s) \leq z_1, \ldots, Z(u_Q, s) \leq z_Q\right) = V(z_1, \ldots, sz_Q, s;\ \tau). \tag{5.10}$$

Eq. (5.10) can be more simplified by assuming that dependency is an intrinsic characteristic of the system, so the extremal dependence of Z is independent of s as

$$V(z_1, \ldots, z_Q, s;\ \tau) \equiv V(z_1, \ldots, z_Q, s';\ \tau), \tag{5.11}$$

for any $s, s' \in \mathcal{S}$. To check the assumption is satisfied, χ must be irrelevant to distance in $\mathcal{S}$. To estimate χ, the extremal coefficient is a useful measure, which is defined as $\theta = V(1, 1)$, and can be estimated with F-madogram [90],

$$\theta = \mathbb{E}\left[\left|F_Z(Z_q(s)) - F_Z(Z_{q'}(s'))\right|\right]/2,$$

where F_Z is the CDF of the unit Fréchet distribution. Although Eq. (5.11) is not necessary in our model by defining the max-stable process on $\mathcal{S} \times \mathcal{U}$, it does not only simplifies ex-

tremal dependence modeling but also alleviates data scarcity by transforming observations at different input locations in $\mathcal{S}$ as additional replications. It also allows using the additional replications for cross-validation in the parameter estimation by partitioning S into K-fold.

Graph-assisted Composite Likelihood Estimation For large Q (e.g., $Q > 20$), computing the full likelihood is prohibitive since the joint density function of Eq. (5.2) involves all partial derivatives, of which the number of terms is the Bell number of Q. Moreover, some max-stable processes even have no closed form. Therefore, a composite likelihood with a low-order (e.g., bivariate or trivariate) density is standard in the inference of max-stable processes. Consider the set of transformed observations $\mathbf{z}^{(b)} = \{z_q^{(b)} : q \in [Q], \}$ for $b \in [NB]$. For the order $H \in \{2, \ldots, Q\}$, $\mathcal{G}_H$ denotes the collection of all subvectors of $\{\mathbf{z}^{(b)} : b \in [NB]\}$ with the size H, and $\mathbf{z}_{[H]}$ is a member of $\mathcal{G}_H$. Let the parameter to be estimated in the max-stable process be τ and $f(\,\cdot\,;\tau)$ is the H-th order density from Eq. (5.2) characterized by the max-stable process. Then, the conventional composite likelihood is

$$\mathrm{CL}_H(\tau|\boldsymbol{Z}) = \prod_{b=1}^{B} \prod_{\mathbf{z}_{[H]} \in \mathcal{G}_H} f(\mathbf{z}_{[H]}^{(b)}; \tau), \tag{5.12}$$

where logarithm can be taken on the RHS as the composite log-likelihood.

However, even though a low-order composite likelihood is used, the computation also can be demanding due to the size of G_H which is $C(Q, H)$. Truncated composite likelihoods use a nonempty subset $G_H \subset \mathcal{G}_H$ in Eq. (5.12) and can improve both the computation and the inference performance [48]. A truncated composite likelihood can be written as

$$\mathrm{TCL}_H(\tau|\boldsymbol{Z}, G_H) = \prod_{b=1}^{B} \prod_{\mathbf{z}_{[H]} \in G_H} f(\mathbf{z}_{[H]}^{(b)}; \tau), \tag{5.13}$$

and regarded as an exponentially weighted composite likelihood in which weights are binary.

The performance of the truncated likelihood estimation can be improved when G_H consists of correlated components. Since key characteristics from geometrically neighboring critical points are likely to be correlated, the adjacency between points in each group can be a reasonable measure. In this chapter, we refer to the average distance of the subvector, which is defined as

$$\delta(g) = \frac{\sum_{u,u' \in g} d_{\mathcal{U}}(u, u')}{\mathrm{C}_2^H}, \tag{5.14}$$

where $g \in \mathcal{G}_H$ and $u \neq u'$. Ordering $\mathcal{G}_H$ with respect to Eq. (5.14) and taking a quantile (or specifying an exact threshold based on background knowledge) will generate G_H in which components are closely located. Algorithm 3 is a pseudo algorithm for generating G_H. The subcollection G_H can be regarded as an undirected graph in which cliques have the size H, so we refer to Eq. (5.13) with G_H generated in this way as a graph-assisted composite likelihood.

Algorithm 3: Generating graph

```
Δ = ∅
for g ∈ 𝒢_H do
    δ(g) = Σ_{u,u'∈g} d_𝒰(u, u') / C_2^H  for u ≠ u'
    Δ = Δ ∪ {δ(g)}
Estimate q-th quantile of Δ as Δ_q
G_H = {g : g ∈ 𝒢_H, δ(g) ≤ Δ_q}
```

Simulation Once the model is fitted by maximizing the graph-assisted composite likelihood, it is possible to simulate with the extreme model for unobserved $s \in \mathcal{S}$. The simulation with the extreme spatial model is not simple as other statistical models since Eq. (5.3) requires infinite samples from the spectral representation. Although it is possible to take maximum on large samples, it is beneficial to use the exact simulation method proposed by [91]. It facilitates more efficient sampling using conditional spectral random processes to

which we resort in the case study. Samples from the max-stable process need to be inverse-transformed to the original GEV distribution with the parameters obtained from Eq. (5.6). The pseudo algorithm of the proposed model with the sampling procedure is provided in Algorithm 4.

Algorithm 4: Multi-output extreme spatial modeling

Modeling Fitting

- Extract maxima from $\{\mathbf{x}(s_n)^{(l)} \colon l \in [L], n \in [N]\}$ with T block size.
- **for** $n \in [N]$ **do**
  - Get MLEs of the marginal parameters: $\mathbf{y}(s_n) \sim \mathrm{GEV}(\Xi(s_n), \mathrm{M}(s_n), \Sigma(s_n))$.
- Fit Eq. (5.6) for each of $\{\Xi(s_n), \mathrm{M}(s_n), \Sigma(s_n) \colon n \in [N]\}$.
- **for** $(b, n) \in [B] \times [N]$ **do**
  - Transform $\mathbf{y}(s_n)^{(b)}$ into unit Fréchet, $\mathbf{z}(s_n)^{(b)}$.
- Define the metric space $(\mathcal{U}, d_{\mathcal{U}})$ for Q critical points.
- Construct a graph G_H with Algorithm 3.
- Fit the max-stable process $Z(\,\cdot\,;\tau)$ on $\{\mathbf{z}(u_q) \colon q \in [Q]\}$ maximizing Eq. (5.13).

Sampling

- For $s \in \mathcal{S}$, sample $Z(s) \sim \{Z(u_q, \tau) \colon q \in [Q]\}$
- Inverse transform $Z(s)$ to $Y(s)$ with $\{\Xi(s), \mathrm{M}(s), \Sigma(s)\}$.

5.3 Case Study: Composite Fuselage Assembly

In this section, we apply the proposed method to model maximum residual stress in composite aircraft production, specifically fuselage assembly. Briefly, the assembly process initiates fuselage shape control using hydraulic actuators on fuselage sections installed on fixtures due to inevitable dimensional deviations. Once the fuselage sections are reshaped into desired

shapes, they are assembled using rivet joints, and the series of procedures causes residual stress in the structure as shown in Fig. 5.1. More details on the fuselage assembly process can be found in [92]. The characteristics of the assembly process can be summarized as follows which are common in modern manufacturing systems.

- The process is subject to intrinsic uncertainty due to actuator load noise, fixture locations, material inhomogeneity, etc. These factors complicate modeling because they are unpredictable.
- Deformation and residual stress in composite parts are complicated because composite materials are composed of different materials and multiple layers with different orientations. Generally, their behaviors to external effects are nonlinear and anisotropic.
- The process yields inevitable residual stress in composite parts after releasing the actuators from the fuselage. Extremely large stress induced by improper actuator loads may lead to structural failures (e.g., fracture, brittle).
- The quality of assembled fuselage is related to safety issues, so ultra-high precision control is required which entails an excessively rigorous inspection with a large number of repetitive measurements and sensors. It compels high dimensional data analytics.

The traditional shape control process requires proficient experts and computationally prohibitive simulations, so surrogate modeling, Bayesian optimization, and active learning methods are proposed to address the aforementioned challenges[27, 76]. For structural failures, [93] considered residual stress to prevent failures in surrogate modeling, while they referred to deterministic simulation with the GP, which is unrealistic in practice due to intrinsic uncertainty. Since the underlying process is stochastic and non-Gaussian [94], the structural failure risk associated with maximum residual stress must be modeled with extreme spatial models.

5.3.1 Experiment Setting

In this case study, we used the surrogate model from [56] as the ground truth to simulate composite fuselage assembly with 20 actuators (10 actuators for each section) and residual stress measured at 128 equispaced critical points around the joint rim. The actuators are subject to intrinsic Gaussian noise where the mean is zero and the variance is 5% of the force magnitude. For training samples, 30 design points are selected with the maximin LHD [50] over $[-200, 200]^{20}$ lbf and replicated the simulation 500 times for each design point. Consequently, we have the design matrix with the size 30×20 and observations with $30 \times 500 \times 128$. For testing, we select 20 design points in the same way as the training data. To extract maxima, different block sizes are considered to see the effect based on the number of samples and the reported yearly orders of Boeing 787 [95], which roughly varies between 25 and 350. For the metric space for critical points, we defined $\mathcal{U} \subset \mathbb{R}^2$ where the origin is the center of the nominal joint planar fuselage in the Z direction (see Fig. 5.1). For $d_{\mathcal{U}}$, the shortest path along the surface between two points is considered.

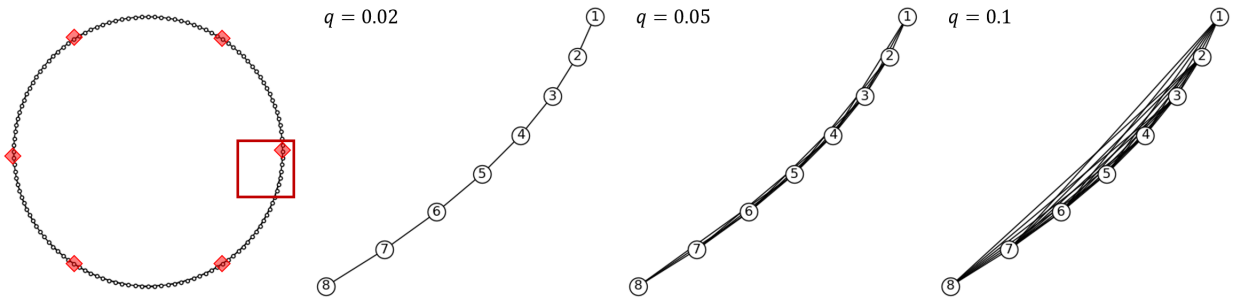

Figure 5.2: Graphs of critical points with different quantiles. The critical points highlighted with diamonds are the representative critical points that are used in evaluating extremal dependence (i.e., 6 out of 128). The three right figures show subgraphs with 1-8 critical points in the red box. The left graph with $q = 0.02$ is used in the case study.

To investigate extremal dependence in maximum residual stress, we refer to empirical estimates of Eq. (5.7) between points as shown in the left of Fig. 5.3. To reduce the number of

pairs for evaluating Eq. (5.7), six representative critical points, with the largest variance, are chosen as shown in Fig. 5.2. The empirical estimate is nonzero (0.0025), which implies that maximum residual stress between critical points is asymptotically dependent. The two right figures in Fig. 5.3 show empirical F-madograms with respect to Euclidean distance in the actuator design space and the metric space of critical points, respectively. The F-madogram with respect to actuator force shows that the extremal dependence barely changes in the design space which validates Eq. (5.11) on $\mathcal{S}$. Meanwhile, the F-madogram for the critical point distance presents a usual dependence behavior, which increases with the distance. In conclusion, the exponent measure in Eq. (5.11) is valid in this case, and the defined metric space $\mathcal{U}$ is proper for the extreme spatial model.

To compare the proposed method with canonical predictive models, we consider quantile stochastic kriging (QSK) [44], stochastic kriging (SK) [96] and quantile linear regression (QLR) with different quantiles: $q = \{0.9, 0.95, 0.99\}$. The QSK is trained with data over q-th quantile, while the SK and the QLR are trained with all observations (i.e., including non-extreme data). Also, to see the effect of the block size in our model, we considered different sizes: $T \in \{10, 20, 25, 50\}$. For the proposed model, we used Eq. (5.6) as the marginal parameter model, and the Brown–Resnick process as the max-stable process for multi-output modeling, respectively.

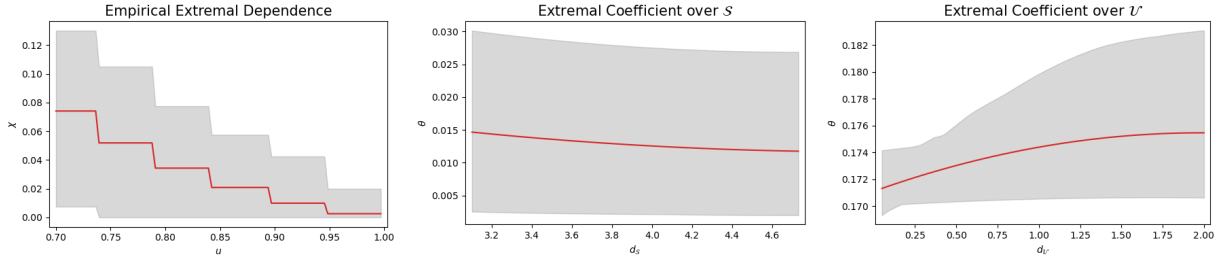

Figure 5.3: Extremal dependence investigation on residual stress with 95% confidence intervals from bootstrapping.

For SK, QSK, and the GP in the marginal parameter model, the RBF kernel with the scale, length, and nugget parameters is considered, and the parameters are estimated with maximum likelihoods using the quasi-Newton method. The Brown–Resnick process also used the RBF kernel but without scale and nugget parameters (i.e., τ is a 2-D length parameter). The parameters of the QLR are estimated using the linear optimizer in `scipy` [97] on the pinball loss associated with the quantile. To fit the max-stable process, the bivariate density ($H = 2$) is used for the truncated likelihood with the graph with quantile $q = 0.02$, which consists of all adjacent critical points pairs as shown in the left of Fig. 5.2. The number of terms in the truncated composite likelihood is 162, while the non-truncated likelihood consists of 8,128 terms. It is noteworthy that it is possible to introduce further prior knowledge to the graph such as fixture or actuator locations that may influence the adjacent dependency. However, we did not consider such information in this case study.

For model evaluation, we consider two metrics because the QLR's distributional prediction is not straightforward as well as other methods, and the techniques are beyond the scope of this dissertation. As the major metric, we use the empirical Wasserstein distance (WD) between the testing dataset and samples from the marginal distributions of each model to comprehensively compare the distributional distance. The empirical distance between two random variables is defined as

$$\mathrm{WD}(Y_1, Y_2) = \sum_{n=1}^{N} \frac{|Y_1^{(n)} - Y_2^{(n)}|}{N}, \tag{5.15}$$

where $Y_1^{(n)}$ and $Y_2^{(n)}$ are ordered samples from Y_1 and Y_2, respectively. Since the metric is empirical, we use bootstrapping in the comparison. Due to its own right, the WD metric is appropriate to evaluate the discrepancy between two distributions, which is frequently used in distributionally robust optimization problems. The second metric is the percentage mean

distance (PMD), which refers to the point estimation of the mean as

$$\mathrm{PMD}(Y_1, Y_2) = \frac{|\bar{Y_1} - \bar{Y_2}|}{|\bar{Y_1}|}, \tag{5.16}$$

where $\bar{Y}$ is the arithmetic mean of Y. Although it is as not comprehensive as the WD, it facilitates comparing the QLR to other methods in terms of extremal trends in residual stress.

5.3.2 Results

Predictive Performance

Table 5.1 summarizes the results of considered methods, where the MESM stands for the multi-output extreme spatial model, the proposed method. The scores for all critical points are averaged and the testing dataset and the standard deviations are calculated using bootstrapping. The result shows that the MESM outperforms the other models in WD regardless of the block size. The worst performance of the SK implies that the conventional Gaussian models are inappropriate when the extreme values are of interest in the non-Gaussian production system. With the PMD, the QSK performed the best, while the QSK is inferior to the MESM in fully describing the extreme stress distributions. Moreover, we can observe that the QLR is even worse than the MESM in the PMD.

Table 5.1: Averaged distances on the testing dataset and training times of models.

Models	MESM				QSK			QLR			SK
Parameter	$T = 10$	$T = 20$	$T = 25$	$T = 50$	$q = 0.90$	$q = 0.95$	$q = 0.99$	$q = 0.90$	$q = 0.95$	$q = 0.99$	-
WD	3.555 (2.396)	3.572 (2.376)	**3.553 (2.399)**	3.684 (2.562)	5.965 (3.989)	5.948 (3.956)	5.911 (3.933)	-	-	-	8.901 (10.187)
PMD	0.150 (0.046)	0.161 (0.096)	0.152 (0.047)	0.158 (0.069)	**0.112 (0.024)**	0.114 0.026)	0.121 (0.029)	0.169 (0.057)	0.182 (0.064)	0.203 (0.070)	0.121 (0.018)
Training Time (min.)	17.105	17.961	18.518	17.621	0.509	0.484	0.494	50.679	50.193	32.588	0.402

Fig. 5.4 shows the marginal predictive distributions of the MESM and the QSK for an instance of extremal residual stress from the testing dataset, obtained using the block maxima with $T = 25$. It clearly shows that the MESM describes the behavior of the extremal residual stress distribution well with similar confidence intervals, while the QSK has too large intervals, which is implausible. Also, it shows that residual stress tends to be high at the bottom side of the fuselage due to the fixture and direct contact with the actuators.

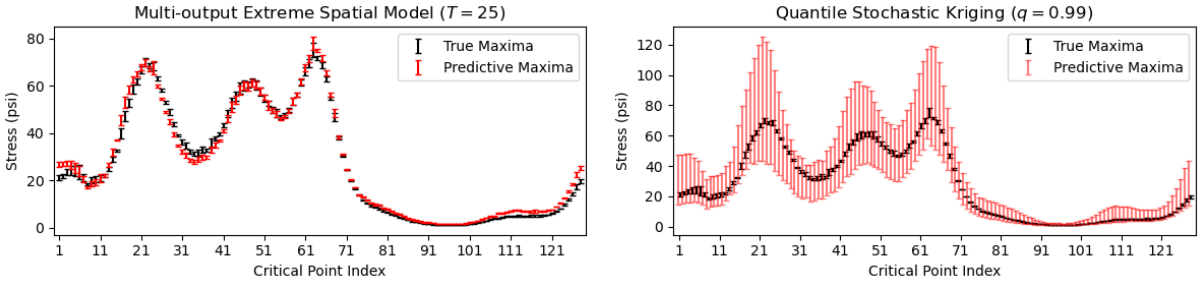

Figure 5.4: Confidence intervals (95%) of marginal distributions of extremal residual stress in fuselage assembly.

Training Time

The most of time in training the MESM was taken by fitting the max-stable process for multi-output modeling, which took about 15 minutes with the graph with $q = 0.02$ and

resulted in $\tau = (10.569, 11.126)$. Meanwhile, training the max-stable process with the full composite likelihood took about 180 minutes and resulted in $\tau = (11.911, 11.795)$. Behaviors of the sample paths from the two models were negligible as shown in Fig. 5.5. The QSK and the SK took less than a minute, while it is due to their lack of dependence modeling. The QLR also does not include the dependence model, but it took the most time among the compared models.

5.3.3 Application Extensions

The extreme spatial model enables more useful analyses of extreme values that cannot be done with the other models. The first thing that we can do is simulate extreme events in the assembly process. Samples from the extreme model represent the maximum stress trend in the process, and it may be used to efficiently monitor the assembly process by allocating sensors to a few vulnerable critical points. For example, an observed actuator input is expected to yield extremal residual stress in Fig. 5.5, which produces the largest stress at the critical point # 62, to which the manufacturer should refer with more cautions.

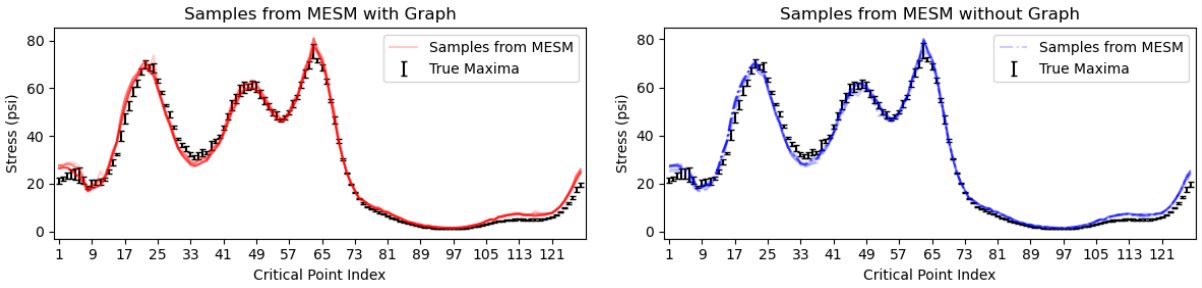

Figure 5.5: Simulations with the MESM. Ten samples are drawn from each MESM.

A further promising application is to use the R-return level of a GEV distributed value Y in the design and control of the production system, robust to extreme risks. The R-return

level for $s \in \mathcal{S}$ at q-th critical point is defined as

$$\mathrm{RL}_q(s; R) = \mu_q + \frac{\sigma_q}{\xi_q}\{(-\log p)^{-\xi_q} - 1\}, \quad p \in (0, 1), \tag{5.17}$$

where $p = 1 - 1/R$ and $R \in \mathbb{N}$. Eq. (5.17) can be interpreted as the value Y exceeding RL_q once on average every R cycle of the block size (e.g., lot size, batch size, yearly deliveries). Fig. 5.6 illustrates $R = 100$ return levels for all $s \in \mathcal{S}$, and the red line is the maximum. The maximum is at most 100 psi, which is far less than the yield stress of the composite material, so the assembly process may utilize higher actuator loads for efficiency. If the system is designed and controlled based on return levels, the manufacturer also may save the inspection cost by concentrating the monitoring assets at the most vulnerable critical points or locations.

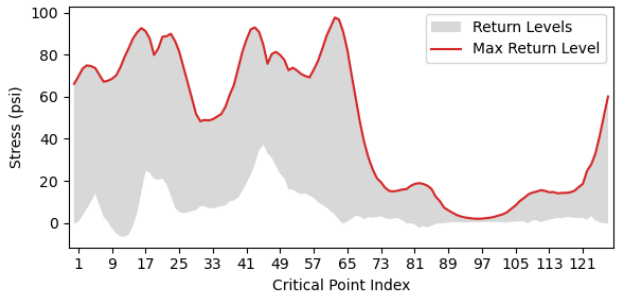

Figure 5.6: Return levels ($R = 100$) on the control design space.

5.4 Summary

Extreme events in composite aircraft manufacturing are extremely rare but significant. They have been frequently ignored or underestimated with misspecified models in the modeling of engineering systems, we proposed the multi-output extreme spatial model for composite fuselage assembly that is capable of modeling correlated key characteristics. The proposed model

establishes the marginal parameter model over the control input space to estimate marginal distributions, and the max-stable process is used to link key characteristics based on the geometric locations of critical points. We also developed graph-assisted composite likelihood estimation to incorporate background knowledge and improve estimation efficiency.

The case study result shows that the conventional GPs are insufficient to model extreme residual stress in composite fuselage assembly. Furthermore, simulation with the proposed model is demonstrated, and the application of return level for system design is discussed. Apart from simulation and system design, extreme risk-robust control also can be a promising application with the extreme spatial model by embedding the model into constraints in control policy updates. As a further discussion, the underlying process of composite fuselage assembly is nonlinear [56] to which the proposed model is appropriate to apply. For extreme events in other multi-output manufacturing systems, the proposed method is expected to perform well once the system is properly investigated a priori to avoid lacking power.

Chapter 6

Summary and Future Research

6.1 Summary of Contributions

In this dissertation, we proposed two active learning strategies and the multi-output extreme spatial model for complex engineering systems. The contributions of this dissertation are as follows. First, PAL is proposed that improves informativeness quantification relevance in active learning under the existence of heterogeneity. The novel partitioning scheme coupling the mean shift and the SVC is developed, which provided a more relevant partition in the simulation study than [11]. The twofold active learning strategy accelerates exploration with global searching and focuses on the most desirable subregion with local searching. Moreover, the Cholesky factor update algorithm is proposed to alleviate the computational cost of PAL. PAL outperformed the benchmark methods in three real-world cases with the same sampling budget. Second, failure-averse active learning is developed for systems with implicit constraints. The method is devised to avoid failures from violating the implicit constraints in active learning. The information criterion in failure-averse active learning consists of safe variance reduction and safe region expansion, and we incorporated MOO to achieve optimality and reflect the decision maker's preference in active learning. For theoretical analysis, we showed that the information criterion asymptotically converges with a desirable safe region estimation. The simulation and case study results showed that failure-averse active learning outperforms the existing safe exploration method. Lastly, we proposed the multi-output

extreme spatial model to analyze extreme events in composite fuselage assembly. The proposed extreme spatial model is efficiently designed to model controllable systems with the marginal parameter model. The max-stable process is utilized for multi-output response by establishing the background knowledge-assisted framework. Aided by the graph-assisted composite likelihood estimation, the multi-output extreme spatial model outperformed the quantile-based models in modeling the extreme residual stress from composite fuselage assembly.

The proposed models can be further discussed for more realistic settings and larger problems. The partitioning scheme of PAL is subject to data imbalance owing to obsessional local searching. Although the stochastic global searching in Eq. (3.9) is proposed to prevent the localization, it may be insufficient in extreme cases. We believe that SVC for imbalanced data techniques [98] and lower bounding the categorical distribution in global searching may resolve the issue. For failure-averse active learning, the choice of weight parameter in the integrated acquisition function can be aided by other machine learning techniques such as reinforcement learning. Unfortunately, considering the Pareto frontier was not effective in the considered simulation and the case study, which resulted in $w = 0$ or $w = 1$. Meanwhile, including an additional objective function for the weight parameter (e.g., keep balancing between two scores) performed promisingly. In the multi-output extreme spatial model, we only used the Brown–Resnick process, which is asymptotic dependent, based on the preliminary dependence analysis. However, it should be investigated that asymptotic independent systems are also common in the engineering domain as the environmental science.

6.2 Future Research Directions

The research on this dissertation topic aroused several new topics in my future research. Although the development of machine learning and computing power realized unprecedented artificial intelligence, it is still reluctant to apply data-driven models to critical problems, where human lives are involved or a suboptimal decision results in catastrophic consequences. Currently, involving domain knowledge in data analytics is inevitable to enhance credibility. Domain knowledge can be embedded into mathematical modeling in the form of additional objective functions, constraints, and boundary conditions that must be sophisticatedly integrated with a multi-objective optimization algorithm. One of my future research is not only to develop a way of including domain knowledge in machine learning but also to integrate the system in a data-driven way instead of arbitration by humans. Another topic is methodology development in extreme value modeling. Extreme value models are equipped with lots of tempting properties for data analytics, while they are unexploited by the machine learning community. I believe there are more applicable machine learning techniques that exploit advanced computing power to extreme value models. For example, ensemble modeling may improve the stability of extreme value models and data augmentation can resolve the data scarcity in extreme value analysis. Developing machine learning-driven extreme value models with inspiring case studies will encourage its generality.

6.3 Codes for Implementation

To improve the repeatability and reproducibility of the proposed methods, the codes used in this dissertation are open to the public on the website: https://github.com/cheolheil.

Bibliography

[1] B. Settles. *Active Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning Series. Morgan & Claypool, 2012. ISBN 9781608457250. doi: 10.1007/978-3-031-01560-1.

[2] Adel Alaeddini, Edward Craft, Rajitha Meka, and Stanford Martinez. Sequential laplacian regularized v-optimal design of experiments for response surface modeling of expensive tests: An application in wind tunnel testing. *IISE Transactions*, 51(5): 559–576, February 2019. doi: 10.1080/24725854.2018.1508928.

[3] Raphaël Pestourie, Youssef Mroueh, Thanh V. Nguyen, Payel Das, and Steven G. Johnson. Active learning of deep surrogates for PDEs: Application to metasurface design. *npj Computational Materials*, 6(1):1–7, October 2020. doi: 10.1038/s41524-020-00431-2.

[4] Francisco Sahli Costabal, Yibo Yang, Paris Perdikaris, Daniel E. Hurtado, and Ellen Kuhl. Physics-informed neural networks for cardiac activation mapping. *Frontiers in Physics*, 8:42, 2020. doi: 10.3389/fphy.2020.00042.

[5] Kjetil O. Lye, Siddhartha Mishra, Deep Ray, and Praveen Chandrashekar. Iterative surrogate model optimization (ISMO): An active learning algorithm for PDE constrained optimization with deep neural networks. *Computer Methods in Applied Mechanics and Engineering*, 374:113575, feb 2021. doi: 10.1016/j.cma.2020.113575.

[6] Xiu Yang, David Barajas-Solano, Guzel Tartakovsky, and Alexandre M. Tartakovsky. Physics-informed CoKriging: A gaussian-process-regression-based multi-

fidelity method for data-model convergence. *Journal of Computational Physics*, 395: 410–431, oct 2019. doi: 10.1016/j.jcp.2019.06.041.

[7] Jialei Chen, Zhehui Chen, Chuck Zhang, and C. F. Jeff Wu. APIK: Active physics-informed kriging model with partial differential equations. *SIAM/ASA Journal on Uncertainty Quantification*, 10(1):481–506, March 2022. doi: 10.1137/20m1389285.

[8] Xiaowei Yue, Yuchen Wen, Jeffrey H. Hunt, and Jianjun Shi. Active learning for gaussian process considering uncertainties with application to shape control of composite fuselage. *IEEE Transactions on Automation Science and Engineering*, 18(1):36–46, jan 2021. doi: 10.1109/tase.2020.2990401.

[9] Hyoung-Moon Kim, Bani K. Mallick, and C. C. Holmes. Analyzing nonstationary spatial data using piecewise gaussian processes. *Journal of the American Statistical Association*, 100(470):653–668, 2005. doi: 10.1198/016214504000002014.

[10] Robert B. Gramacy and Herbert K. H. Lee. Bayesian treed gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483):1119–1130, 2008. doi: 10.1198/016214508000000689.

[11] Matthew J. Heaton, William F. Christensen, and Maria A. Terres. Nonstationary gaussian process models using spatial hierarchical clustering from finite differences. *Technometrics*, 59(1):93–101, 2017. doi: 10.1080/00401706.2015.1102763.

[12] Sambu Seo, Marko Wallat, Thore Graepel, and Klaus Obermayer. Gaussian process regression: Active data selection and test point rejection. In *Mustererkennung 2000*, pages 27–34. Springer, 2000. doi: 10.1007/978-3-642-59802-9_4.

[13] Edoardo Pasolli and Farid Melgani. Gaussian process regression within an active learn-

ing scheme. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 3574–3577. IEEE, IEEE, jul 2011. doi: 10.1109/igarss.2011.6049994.

[14] Christoph Käding, Erik Rodner, Alexander Freytag, Oliver Mothes, Björn Barz, Joachim Denzler, and A. G. Carl Zeiss. Active learning for regression tasks with expected model output changes. In *BMVC*, page 103, 2018. URL http://www.bmva.org/bmvc/2018/contents/papers/0362.pdf.

[15] Collin B. Erickson, Bruce E. Ankenman, Matthew Plumlee, and Susan M. Sanchez. Gradient based criteria for sequential design. In *2018 Winter Simulation Conference (WSC)*, pages 467–478. IEEE, IEEE, dec 2018. doi: 10.1109/wsc.2018.8632546.

[16] Sébastien Marmin, David Ginsbourger, Jean Baccou, and Jacques Liandrat. Warped gaussian processes and derivative-based sequential designs for functions with heterogeneous variations. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):991–1018, 2018. doi: 10.1137/17m1129179.

[17] Byeongsoo Kim, Yongbin Lee, and Dong-Hoon Choi. Construction of the radial basis function based on a sequential sampling approach using cross-validation. *Journal of Mechanical Science and Technology*, 23(12):3357–3365, 2009. doi: 10.1007/s12206-009-1014-z.

[18] Christopher A. Pope, John Paul Gosling, Stuart Barber, Jill S. Johnson, Takanobu Yamaguchi, Graham Feingold, and Paul G. Blackwell. Gaussian process modeling of heterogeneity and discontinuities using voronoi tessellations. *Technometrics*, 63(1):53–63, 2021.

[19] Robert B. Gramacy and Herbert K. H. Lee. Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51(2):130–145, 2009.

[20] Bledar Konomi, Georgios Karagiannis, Avik Sarkar, Xin Sun, and Guang Lin. Bayesian treed multivariate gaussian process with adaptive design: Application to a carbon capture unit. *Technometrics*, 56(2):145–158, 2014. doi: 10.1080/00401706.2013.879078.

[21] Jens Schreiter, Duy Nguyen-Tuong, Mona Eberts, Bastian Bischoff, Heiner Markert, and Marc Toussaint. Safe exploration for active learning with gaussian processes. In *Machine Learning and Knowledge Discovery in Databases*, pages 133–149. Springer International Publishing, 2015. doi: 10.1007/978-3-319-23461-8_9.

[22] Matteo Turchetta, Felix Berkenkamp, and Andreas Krause. Safe exploration for interactive machine learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/4f398cb9d6bc79ae567298335b51ba8a-Paper.pdf.

[23] Jacob Gardner, Matt Kusner, Zhixiang, Kilian Weinberger, and John Cunningham. Bayesian optimization with inequality constraints. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 937–945, Bejing, China, 22–24 Jun 2014. PMLR. URL https://proceedings.mlr.press/v32/gardner14.html. Cycle 2.

[24] Remi Lam and Karen Willcox. Lookahead bayesian optimization with inequality constraints. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/83f97f4825290be4cb794ec6a234595f-Paper.pdf.

[25] Benjamin Letham, Brian Karrer, Guilherme Ottoni, and Eytan Bakshy. Constrained bayesian optimization with noisy experiments. *Bayesian Analysis*, 14(2):495–519, 2019.

[26] David Eriksson and Matthias Poloczek. Scalable constrained bayesian optimization. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 730–738. PMLR, 13–15 Apr 2021. URL https://proceedings.mlr.press/v130/eriksson21a.html.

[27] Areej AlBahar, Inyoung Kim, Xing Wang, and Xiaowei Yue. Physics-constrained bayesian optimization for optimal actuators placement in composite structures assembly. *IEEE Transactions on Automation Science and Engineering*, pages 1–12, 2022. doi: 10.1109/TASE.2022.3200376.

[28] Matthias Schonlau, William J. Welch, and Donald R. Jones. Global versus local search in constrained optimization of computer models. In *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, pages 11–25. Institute of Mathematical Statistics, 1998. doi: 10.1214/lnms/1215456182.

[29] Robert B. Gramacy and Herbert K. H. Lee. Optimization under unknown constraints. In *Bayesian Statistics 9*, pages 229–256. Oxford University Press, oct 2011. ISBN 9780199694587. doi: 10.1093/acprof:oso/9780199694587.003.0008.

[30] Jose Miguel Hernandez-Lobato, Michael Gelbart, Matthew Hoffman, Ryan Adams, and Zoubin Ghahramani. Predictive entropy search for bayesian optimization with unknown constraints. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1699–1707, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/hernandez-lobatob15.html.

[31] Yanan Sui, Vincent Zhuang, Joel Burdick, and Yisong Yue. Stagewise safe Bayesian optimization with Gaussian processes. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4781–4789. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/sui18a.html.

[32] Woojin Cho, Youngrae Kim, and Jinkyoo Park. Hierarchical anomaly detection using a multioutput gaussian process. *IEEE Transactions on Automation Science and Engineering*, 17(1):261–272, jan 2020. doi: 10.1109/tase.2019.2917887.

[33] Xiaolei Yu, Zhibin Zhao, Xingwu Zhang, Qiyang Zhang, Yilong Liu, Chuang Sun, and Xuefeng Chen. Deep-learning-based open set fault diagnosis by extreme value theory. *IEEE Transactions on Industrial Informatics*, 18(1):185–196, jan 2022. doi: 10.1109/tii.2021.3070324.

[34] Tang Gu, Krzysztof S. Stopka, Chuan Xu, and David L. McDowell. Prediction of maximum fatigue indicator parameters for duplex ti–6al–4v using extreme value theory. *Acta Materialia*, 188:504–516, apr 2020. doi: 10.1016/j.actamat.2020.02.009.

[35] Brad L. Boyce, Bradley C. Salzbrenner, Jeffrey M. Rodelas, Laura P. Swiler, Jonathan D. Madison, Bradley H. Jared, and Yu-Lin Shen. Extreme-value statistics reveal rare failure-critical defects in additive manufacturing. *Advanced Engineering Materials*, 19(8):1700102, 2017. doi: 10.1002/adem.201700102.

[36] Shui Yu, Zhonglai Wang, and Debiao Meng. Time-variant reliability assessment for multiple failure modes and temporal parameters. *Structural and Multidisciplinary Optimization*, 58(4):1705–1717, 2018. doi: 10.1007/s00158-018-1993-4.

[37] Shiyao Liu and William Q. Meeker. Statistical methods for estimating the mini-

mum thickness along a pipeline. *Technometrics*, 57(2):164–179, 2015. doi: 10.1080/00401706.2014.915236.

[38] Tangbin Xia, Xiaolei Fang, Nagi Gebraeel, Lifeng Xi, and Ershun Pan. Online analytics framework of sensor-driven prognosis and opportunistic maintenance for mass customization. *Journal of Manufacturing Science and Engineering*, 141(5):051011, April 2019. doi: 10.1115/1.4043255.

[39] Colin Lewis-Beck, Qinglong Tian, and William Q. Meeker. Prediction of future failures for heterogeneous reliability field data. *Technometrics*, 64(1):125–138, June 2022. doi: 10.1080/00401706.2021.1921036.

[40] Caleb B. King, Yili Hong, Stephanie P. Dehart, Patrick A. Defeo, and Rong Pan. Planning fatigue tests for polymer composites. *Journal of Quality Technology*, 48(3):227–245, 2016. doi: 10.1080/00224065.2016.11918163.

[41] Shuguang He, Zhaomin Zhang, Wei Jiang, and Dejun Bian. Predicting field reliability based on two-dimensional warranty data with learning effects. *Journal of Quality Technology*, 50(2):198–206, 2018. doi: 10.1080/00224065.2018.1436831.

[42] Zhenxin Zhan, Maochao Xu, and Shouhuai Xu. Predicting cyber attack rates with extreme values. *IEEE Transactions on Information Forensics and Security*, 10(8):1666–1677, aug 2015. doi: 10.1109/tifs.2015.2422261.

[43] Mircea Grigoriu. Estimates of system response maxima by extreme value theory and surrogate models. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):922–955, January 2017. doi: 10.1137/17m1114223.

[44] Anthony C. Davison, Simone A. Padoan, and Mathieu Ribatet. Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186, 2012. doi: 10.1214/11-sts376.

[45] Raphaël Huser and Jennifer L. Wadsworth. Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, 114(525): 434–444, 2019. doi: 10.1080/01621459.2017.1411813.

[46] Jean-Noël Bacro, Carlo Gaetan, Thomas Opitz, and Gwladys Toulemonde. Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data. *Journal of the American Statistical Association*, 115(530): 555–569, June 2020. doi: 10.1080/01621459.2019.1617152.

[47] Gregory P. Bopp, Benjamin A. Shaby, and Raphaël Huser. A hierarchical max-infinitely divisible spatial model for extreme precipitation. *Journal of the American Statistical Association*, 116(533):93–106, May 2021. doi: 10.1080/01621459.2020.1750414.

[48] Stefano Castruccio, Raphaël Huser, and Marc G. Genton. High-order composite likelihood inference for max-stable distributions and processes. *Journal of Computational and Graphical Statistics*, 25(4):1212–1229, 2016. doi: 10.1080/10618600.2015.1086656.

[49] Sebastian Engelke and Adrien S. Hitz. Graphical models for extremes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):871–932, June 2020. doi: 10.1111/rssb.12355.

[50] Thomas J. Santner, Brian J. Williams, and William I. Notz. *The Design and Analysis of Computer Experiments*, volume 2. Springer, 2018. doi: 10.1007/978-1-4939-8847-1.

[51] Xiangyong Cao, Jing Yao, Zongben Xu, and Deyu Meng. Hyperspectral image classification with convolutional neural network and active learning. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4604–4616, jul 2020. doi: 10.1109/tgrs.2020.2964627.

[52] Marc Peter Deisenroth, Dieter Fox, and Carl Edward Rasmussen. Gaussian processes for data-efficient learning in robotics and control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):408–423, feb 2015. doi: 10.1109/tpami.2013.218.

[53] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, USA, 2005. ISBN 9780262256834. doi: 10.7551/mitpress/3206.001.0001.

[54] Seyede Fatemeh Ghoreishi and Mahdi Imani. Bayesian surrogate learning for uncertainty analysis of coupled multidisciplinary systems. *Journal of Computing and Information Science in Engineering*, 21(4), 2021. doi: 10.1115/1.4049994.

[55] Michael W. Hyer and Scott R. White. *Stress Analysis of Fiber-Reinforced Composite Materials*. DEStech Publications, Inc, 2009. ISBN 978-1-932078-86-2.

[56] Cheolhei Lee, Jianguo Wu, Wenjia Wang, and Xiaowei Yue. Neural network gaussian process considering input uncertainty for composite structure assembly. *IEEE/ASME Transactions on Mechatronics*, 27(3):1267–1277, June 2022. doi: 10.1109/tmech.2020.3040755.

[57] Chiwoo Park and Daniel Apley. Patchwork kriging for large-scale gaussian process regression. *Journal of Machine Learning Research*, 19(7):1–43, 2018. URL http://jmlr.org/papers/v19/17-042.html.

[58] Carl Rasmussen and Zoubin Ghahramani. Infinite mixtures of gaussian process experts. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2001. URL https://proceedings.neurips.cc/paper_files/paper/2001/file/9afefc52942cb83c7c1f14b2139b09ba-Paper.pdf.

[59] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (5):603–619, may 2002. doi: 10.1109/34.1000236.

[60] Dorin Comaniciu. An algorithm for data-driven bandwidth selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):281–288, feb 2003. doi: 10.1109/tpami.2003.1177159.

[61] Xi Chen and Qiang Zhou. Sequential design strategies for mean response surface metamodeling via stochastic kriging with adaptive exploration and exploitation. *European Journal of Operational Research*, 262(2):575–585, 2017. doi: 10.1016/j.ejor.2017.03.042.

[62] Haitao Liu, Yew-Soon Ong, and Jianfei Cai. A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. *Structural and Multidisciplinary Optimization*, 57(1):393–416, June 2017. doi: 10.1007/s00158-017-1739-8.

[63] Yan Wang, Xiaowei Yue, Rui Tuo, Jeffrey H. Hunt, Jianjun Shi, et al. Effective model calibration via sensible variable identification and adjustment with application to composite fuselage simulation. *Annals of Applied Statistics*, 14(4):1759–1776, 2020. doi: 10.1214/20-aoas1353.

[64] Yuchen Wen, Xiaowei Yue, Jeffrey H. Hunt, and Jianjun Shi. Feasibility analysis of composite fuselage shape control via finite element analysis. *Journal of Manufacturing Systems*, 46:272–281, 2018. doi: 10.1016/j.jmsy.2018.01.008.

[65] Kaiwen Wang, Yinan Wang, Xiaowei Yue, and Wenjun Cai. Multiphysics modeling and uncertainty quantification of tribocorrosion in aluminum alloys. *Corrosion Science*, 178:109095, 2021. doi: 10.1016/j.corsci.2020.109095.

[66] Kaiwen Wang and Wenjun Cai. Modeling the effects of individual layer thickness and orientation on the tribocorrosion behavior of al/cu nanostructured metallic multilayers. *Wear*, 477:203849, 2021. doi: 10.1016/j.wear.2021.203849.

[67] Saeed Moaveni. *Finite Element Analysis Theory and Application with Ansys, 3/e*. Pearson Education India, 2011. ISBN 9780136681632.

[68] Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating scientific knowledge with machine learning for engineering and environmental systems, 2020.

[69] Seung-Kyum Choi, Ramana Grandhi, and Robert A. Canfield. *Reliability-Based Structural Design*. Springer Science & Business Media, 2006.

[70] Barron J. Bichon, Michael S. Eldred, Laura Painton Swiler, Sandaran Mahadevan, and John M. McFarland. Efficient global reliability analysis for nonlinear implicit performance functions. *Aiaa Journal*, 46(10):2459–2468, 2008. doi: 10.2514/1.34321.

[71] B. Echard, N. Gayton, and M. Lemaire. AK-MCS: An active learning reliability method combining kriging and monte carlo simulation. *Structural Safety*, 33(2):145–154, mar 2011. doi: 10.1016/j.strusafe.2011.01.002.

[72] Rebecca Willett, Robert Nowak, and Rui Castro. Faster rates in regression via active learning. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *Advances in Neural Information Processing Systems*, volume 18. MIT Press, 2005. URL https://proceedings.neurips.cc/paper_files/paper/2005/file/4ea6a546c19499318091a9df40a13181-Paper.pdf.

[73] Pritam Ranjan, Derek Bingham, and George Michailidis. Sequential experiment design

for contour estimation from complex computer codes. *Technometrics*, 50(4):527–541, 2008. doi: 10.1198/004017008000000541.

[74] R. Timothy Marler and Jasbir S. Arora. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization*, 26(6):369–395, 2004. doi: 10.1007/s00158-003-0368-6.

[75] Cheolhei Lee, Kaiwen Wang, Jianguo Wu, Wenjun Cai, and Xiaowei Yue. Partitioned active learning for heterogeneous systems. *Journal of Computing and Information Science in Engineering*, 23(4):041009, January 2023. ISSN 1530-9827. doi: 10.1115/1.4056567.

[76] Xiaowei Yue, Yuchen Wen, Jeffrey H. Hunt, and Jianjun Shi. Surrogate model-based control considering uncertainties for composite fuselage assembly. *Journal of Manufacturing Science and Engineering*, 140(4), February 2018. ISSN 1087-1357. doi: 10.1115/1.4038510. 041017.

[77] Adrian C. Orifici, Izrael Herszberg, and Rodney S. Thomson. Review of methodologies for composite material modelling incorporating failure. *Composite Structures*, 86(1-3): 194–210, 2008. doi: 10.1016/j.compstruct.2008.03.007.

[78] Douglas C. Montgomery. *Introduction to Statistical Quality Control*. John Wiley & Sons, 2020. ISBN 978-1-119-39930-8.

[79] Jianjun Shi. *Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*. CRC press, 2006. ISBN 9780849321511. doi: 10.1201/9781420003901.

[80] Ni Ma and Ward Whitt. A rare-event simulation algorithm for periodic single-server queues. *INFORMS Journal on Computing*, 30(1):71–89, January 2018. doi: 10.1287/ijoc.2017.0766.

[81] Zdravko I. Botev and Pierre L'Ecuyer. Sampling conditionally on a rare event via generalized splitting. *INFORMS Journal on Computing*, 32(4):986–995, April 2020. doi: 10.1287/ijoc.2019.0936.

[82] Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. *An Introduction to Statistical Modeling of Extreme Values*, volume 208. Springer, 2001. doi: 10.1007/978-1-4471-3675-0.

[83] Laurens de Haan and Ana Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2007. ISBN 978-0-387-34471-3.

[84] Sidney I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer London, Limited, 2013. ISBN 9780387759531.

[85] Huiyan Sang and Alan E. Gelfand. Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15 (1):49–65, 2010. doi: 10.1007/s13253-009-0010-1.

[86] Bruce M. Brown and Sidney I. Resnick. Extreme values of independent stochastic processes. *Journal of Applied Probability*, 14(4):732–739, 1977. doi: 10.1017/s0021900200105261.

[87] T. Opitz. Extremal t processes: Elliptical domain of attraction and a spectral representation. *Journal of Multivariate Analysis*, 122:409–413, November 2013. doi: 10.1016/j.jmva.2013.08.008.

[88] Laurens De Haan. A spectral representation for max-stable processes. *The Annals of Probability*, 12(4):1194–1204, 1984. doi: 10.1214/aop/1176993148.

[89] Edwin V. Bonilla, Kian Chai, and Christopher Williams. Multi-task gaussian process prediction. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Ad-*

vances in Neural Information Processing Systems, volume 20. Curran Associates, Inc., 2007. URL https://proceedings.neurips.cc/paper_files/paper/2007/file/66368270ffd51418ec58bd793f2d9b1b-Paper.pdf.

[90] Dan Cooley, Philippe Naveau, and Paul Poncet. Variograms for spatial max-stable random fields. In *Lecture Notes in Statistics*, pages 373–390. Springer New York, 2006. doi: 10.1007/0-387-36062-x_17.

[91] Clément Dombry, Sebastian Engelke, and Marco Oesting. Exact simulation of max-stable processes. *Biometrika*, 103(2):303–317, 2016. doi: 10.1093/biomet/asw008.

[92] Yuchen Wen, Xiaowei Yue, Jeffrey H. Hunt, and Jianjun Shi. Virtual assembly and residual stress analysis for the composite fuselage assembly process. *Journal of Manufacturing Systems*, 52:55–62, 2019. doi: 10.1016/j.jmsy.2019.04.001.

[93] Cheolhei Lee, Xing Wang, Jianguo Wu, and Xiaowei Yue. Failure-averse active learning for physics-constrained systems. *IEEE Transactions on Automation Science and Engineering*, pages 1–12, 2022. doi: 10.1109/tase.2022.3213827.

[94] Tim Lutz, Xiaowei Yue, and Jaime Camelio. Towards a Digital Twin: Simulation and Residual Stress Analysis in Aerospace Composite Structures Assembly. In *Proceedings of the ASME 2022 17th International Manufacturing Science and Engineering Conference*, volume Manufacturing Processes; Manufacturing Systems, June 2022. ISBN 978-0-7918-8581-9. doi: 10.1115/MSEC2022-85439. URL https://doi.org/10.1115/MSEC2022-85439. V002T05A036.

[95] The Boeing Company. Orders & Deliveries. https://www.boeing.com/commercial/#/orders-deliveries, 2023. Accessed April 17, 2023.

[96] Bruce E. Ankenman, Barry L. Nelson, and Jeremy Staum. Stochastic kriging for

simulation metamodeling. *Operations Research*, 58(2):371–382, 2010. doi: 10.1287/opre.1090.0754.

[97] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C. J. Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

[98] Rehan Akbani, Stephen Kwek, and Nathalie Japkowicz. Applying support vector machines to imbalanced datasets. In *Machine Learning: ECML 2004*, pages 39–50. Springer Berlin Heidelberg, 2004. doi: 10.1007/978-3-540-30115-8_7.

[99] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006. ISBN 978-0-387-31073-2.

[100] Roger Frigola, Yutian Chen, and Carl Edward Rasmussen. Variational gaussian process state-space models. In *Advances in Neural Information Processing Systems*, pages 3680–3688, 2014.

[101] Peter Bühlmann, Torsten Hothorn, et al. Boosting algorithms: Regularization, prediction and model fitting. *Statistical Science*, 22(4):477–505, 2007. doi: 10.1214/07-sts242.

[102] Peter Bühlmann, Bin Yu, et al. Analyzing bagging. *The Annals of Statistics*, 30(4): 927–961, 2002. doi: 10.1214/aos/1031689014.

[103] Andreas Buja and Werner Stuetzle. Observations on bagging. *Statistica Sinica*, 16(2): 323–351, 2006. URL https://www.jstor.org/stable/24307547.

[104] Xiaowei Jia, Jared Willard, Anuj Karpatne, Jordan Read, Jacob Zwart, Michael Steinbach, and Vipin Kumar. Physics guided rnns for modeling dynamical systems: A case study in simulating lake temperature profiles. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pages 558–566. SIAM, 2019. doi: 10.1137/1.9781611975673.63.

[105] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378: 686–707, 2019. doi: 10.1016/j.jcp.2018.10.045.

[106] Dongkun Zhang, Lu Lu, Ling Guo, and George Em Karniadakis. Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. *Journal of Computational Physics*, 397:108850, 2019. doi: 10.1016/j.jcp.2019.07.048.

[107] Yibo Yang and Paris Perdikaris. Adversarial uncertainty quantification in physics-informed neural networks. *Journal of Computational Physics*, 394:136–152, 2019. doi: 10.1016/j.jcp.2019.05.027.

[108] Hieu T. Nguyen and Arnold Smeulders. Active learning using pre-clustering. In *Proceedings of the 21st International Conference on Machine Learning*, page 79, 2004. doi: 10.1145/1015330.1015349.

[109] Yuhang Zhang, Hsiuhan Lexie Yang, Saurabh Prasad, Edoardo Pasolli, Jinha Jung, and Melba Crawford. Ensemble multiple kernel active learning for classification of

multisource remote sensing data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(2):845–858, feb 2015. doi: 10.1109/jstars.2014.2359136.

[110] Prem Melville and Raymond J. Mooney. Diverse ensembles for active learning. In *Proceedings of the 21st International Conference on Machine Learning*, page 74, 2004. doi: 10.1145/1015330.1015385.

[111] Martina Hasenjäger and Helge Ritter. Active learning with local models. *Neural Processing Letters*, 7(2):107–117, 1998. doi: 10.1023/A:1009688513124.

[112] Robert Burbidge, Jem J. Rowland, and Ross D. King. Active learning for regression based on query by committee. In *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, pages 209–218. Springer Berlin Heidelberg, 2007. doi: 10.1007/978-3-540-77226-2_22.

[113] Edoardo Vignotto and Sebastian Engelke. Extreme value theory for anomaly detection – the GPD classifier. *Extremes*, 23(4):501–520, sep 2020. doi: 10.1007/s10687-020-00393-0.

[114] Jouchi Nakajima, Tsuyoshi Kunihama, Yasuhiro Omori, and Sylvia Frühwirth-Schnatter. Generalized extreme value distribution with time-dependence using the ar and ma models in state space form. *Computational Statistics & Data Analysis*, 56 (11):3241–3259, 2012. doi: 10.1016/j.csda.2011.04.017.

[115] Charles E. McCulloch and John M. Neuhaus. Improving predictions when interest focuses on extreme random effects. *Journal of the American Statistical Association*, 0 (0):1–10, 2021. doi: 10.1080/01621459.2021.1938583.

[116] R. Huser and A. C. Davison. Space–time modelling of extreme events. *Journal of the*

Royal Statistical Society Series B: Statistical Methodology, 76(2):439–461, September 2013. doi: 10.1111/rssb.12035.

[117] Andrew Kusiak. Extreme science and engineering. *Journal of Intelligent Manufacturing*, 31(7):1607–1610, 2020. doi: 10.1007/s10845-020-01643-5.

[118] Tsui-Wei Weng, Huan Zhang, Pin-Yu Chen, Jinfeng Yi, Dong Su, Yupeng Gao, Cho-Jui Hsieh, and Luca Daniel. Evaluating the robustness of neural networks: An extreme value theory approach, January 2018.

[119] Ethan M. Rudd, Lalit P. Jain, Walter J. Scheirer, and Terrance E. Boult. The extreme value machine. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40 (3):762–768, mar 2018. doi: 10.1109/tpami.2017.2707495.

[120] Sreelekha Guggilam, S. M. Arshad Zaidi, Varun Chandola, and Abani Patra. Bayesian anomaly detection using extreme value theory, May 2019.

[121] Walter J. Scheirer. Extreme value theory-based methods for visual recognition. *Synthesis Lectures on Computer Vision*, 7(1):1–131, 2017. doi: 10.1007/978-3-031-01817-6.

[122] H. Sebastian Seung, Manfred Opper, and Haim Sompolinsky. Query by committee. In *Proceedings of the 5th Annual Workshop on Computational Learning Theory*, pages 287–294, 1992. doi: 10.1145/130385.130417.

[123] Michael D. McKay, Richard J. Beckman, and William J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1):55–61, 2000. doi: 10.1080/00401706.2000.10485979.

[124] Jaeho Kang, Kwang Ryel Ryu, and Hyuk-Chul Kwon. Using cluster-based sampling to select initial training set for active learning in text classification. In *Pacific-Asia*

Conference on Knowledge Discovery and Data Mining, pages 384–388. Springer, 2004. doi: 10.1007/978-3-540-24775-3_46.

[125] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL http://jmlr.org/papers/v12/pedregosa11a.html.

[126] Jian-Xun Wang, Jin-Long Wu, and Heng Xiao. Physics-informed machine learning approach for reconstructing reynolds stress modeling discrepancies based on dns data. *Physical Review Fluids*, 2(3):034603, 2017. doi: 10.1103/physrevfluids.2.034603.

[127] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations, 2017.

[128] Joshua L. Lansford and Dionisios G. Vlachos. Infrared spectroscopy data-and physics-driven machine learning for characterizing surface microstructure of complex materials. *Nature Communications*, 11(1):1–12, 2020. doi: 10.1038/s41467-020-15340-7.

[129] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1050–1059, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/gal16.html.

[130] Maria-Florina Balcan, Alina Beygelzimer, and John Langford. Agnostic active learn-

ing. *Journal of Computer and System Sciences*, 75(1):78–89, 2009. doi: 10.1016/j.jcss.2008.07.003.

[131] Peter L. Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. In *Lecture Notes in Computer Science*, pages 224–240. Springer Berlin Heidelberg, 2001. doi: 10.1007/3-540-44581-1_15.

[132] V. Vapnik. Principles of risk minimization for learning theory. In J. Moody, S. Hanson, and R.P. Lippmann, editors, *Advances in Neural Information Processing Systems*, volume 4. Morgan-Kaufmann, 1991. URL https://proceedings.neurips.cc/paper_files/paper/1991/file/ff4d5fbbafdf976cfdc032e3bde78de5-Paper.pdf.

[133] Yiming Zhang, Nam H. Kim, and Raphael T. Haftka. General-surrogate adaptive sampling using interquartile range for design space exploration. *Journal of Mechanical Design*, 142(5), November 2019. doi: 10.1115/1.4044432.

[134] Binbin Zhang, Nagavenkat Adurthi, Rahul Rai, and Puneet Singla. A novel sampling technique for probabilistic static coverage problems. *Journal of Mechanical Design*, 138(3), January 2016. doi: 10.1115/1.4032395.

[135] Soumalya Sarkar, Sudeepta Mondal, Michael Joly, Matthew E. Lynch, Shaunak D. Bopardikar, Ranadip Acharya, and Paris Perdikaris. Multifidelity and multiscale bayesian framework for high-dimensional engineering design and calibration. *Journal of Mechanical Design*, 141(12), September 2019. doi: 10.1115/1.4044598.

[136] Xufeng Yang, Yongshou Liu, Caiying Mi, and Xiangjin Wang. Active learning kriging model combining with kernel-density-estimation-based importance sampling method for the estimation of low failure probability. *Journal of Mechanical Design*, 140(5), 2018. doi: 10.1115/1.4039339.

[137] Conner Sharpe, Tyler Wiest, Pingfeng Wang, and Carolyn Conner Seepersad. A comparative evaluation of supervised machine learning classification techniques for engineering design applications. *Journal of Mechanical Design*, 141(12), 2019. doi: 10.1115/1.4044524.

[138] Sangjune Bae, Chanyoung Park, and Nam H. Kim. Estimating effect of additional sample on uncertainty reduction in reliability analysis using gaussian process. *Journal of Mechanical Design*, pages 1–27, 2020. doi: 10.1115/1.4047002.

[139] Liye Lv, Maolin Shi, Xueguan Song, Wei Sun, and Jie Zhang. A fast-converging ensemble infilling approach balancing global exploration and local exploitation: The go-inspired hybrid infilling strategy. *Journal of Mechanical Design*, 142(2), October 2019. doi: 10.1115/1.4044112.

[140] Xiaowei Yue and Jianjun Shi. Surrogate model–based optimal feed-forward control for dimensional-variation reduction in composite parts' assembly processes. *Journal of Quality Technology*, 50(3):279–289, 2018. doi: 10.1080/00224065.2018.1474688.

[141] Mostafa Reisi Gahrooei, Kamaran Paynabar, Massimo Pacella, and Bianca Maria Colosimo. An adaptive fused sampling approach of high-accuracy data in the presence of low-accuracy data. *IISE Transactions*, 51(11):1251–1264, April 2019. doi: 10.1080/24725854.2018.1540901.

[142] Apoorv Agnihotri and Nipun Batra. Exploring bayesian optimization. *Distill*, 5 (5), May 2020. doi: 10.23915/distill.00026. URL https://distill.pub/2020/bayesian-optimization.

[143] David A. Cohn, Zoubin Ghahramani, and Michael I. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996.

[144] Masashi Sugiyama and Neil Rubens. Active learning with model selection in linear regression. In *Proceedings of the 2008 SIAM International Conference on Data Mining*, pages 518–529. SIAM, 2008. doi: 10.1137/1.9781611972788.47.

[145] Masashi Sugiyama. Active learning in approximately linear regression based on conditional expectation of generalization error. *Journal of Machine Learning Research*, 7 (6):141–166, 2006. URL http://jmlr.org/papers/v7/sugiyama06a.html.

[146] Wenbin Cai, Ya Zhang, and Jun Zhou. Maximizing expected model change for active learning in regression. In *2013 IEEE 13th International Conference on Data Mining*, pages 51–60. IEEE, IEEE, dec 2013. doi: 10.1109/icdm.2013.104.

[147] Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. Bayesian cart model search. *Journal of the American Statistical Association*, 93(443):935–948, 1998. doi: 10.1080/01621459.1998.10473750.

[148] Alan Tan Wei Min, Abhishek Gupta, and Yew-Soon Ong. Generalizing transfer bayesian optimization to source-target heterogeneity. *IEEE Transactions on Automation Science and Engineering*, 18(4):1754–1765, oct 2021. doi: 10.1109/tase.2020.3017644.

[149] Alexander Freytag, Erik Rodner, Paul Bodesheim, and Joachim Denzler. Labeling examples that matter: Relevance-based active learning with gaussian processes. In *German Conference on Pattern Recognition*, pages 282–291. Springer, 2013. doi: 10.1007/978-3-642-40602-7_31.

[150] Fabien Lauer. On the complexity of piecewise affine system identification. *Automatica*, 62:148–153, 2015.

[151] Jack Sherman and Winifred J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, March 1950. doi: 10.1214/aoms/1177729893.

[152] William H. Press, H. William, Saul A. Teukolsky, A. Saul, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge university press, 2007. ISBN 9780521880688.

[153] David J. C. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, 1992. doi: 10.1162/neco.1992.4.4.590.

[154] Elena Uteva, Richard S. Graham, Richard D. Wilkinson, and Richard J. Wheatley. Active learning in gaussian process interpolation of potential energy surfaces. *The Journal of Chemical Physics*, 149(17):174114, 2018. doi: 10.1063/1.5051772.

[155] David Higdon. A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, 5(2):173–190, 1998. doi: 10.1023/a:1009666805688.

[156] Markus Heinonen, Henrik Mannerström, Juho Rousu, Samuel Kaski, and Harri Lähdesmäki. Non-stationary gaussian process regression with hamiltonian monte carlo. In Arthur Gretton and Christian C. Robert, editors, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 732–740, Cadiz, Spain, 09–11 May 2016. PMLR. URL https://proceedings.mlr.press/v51/heinonen16.html.

[157] Christopher Paciorek and Mark Schervish. Nonstationary covariance functions for gaussian process regression. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press,

2003. URL https://proceedings.neurips.cc/paper_files/paper/2003/file/326a8c055c0d04f5b06544665d8bb3ea-Paper.pdf.

[158] Jasper Snoek, Hugo Larochelle, and Ryan P. Adams. Practical bayesian optimization of machine learning algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf.

[159] Robert B. Gramacy. *Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences*. CRC Press, 2020. doi: 10.1201/9780367815493.

[160] Foivos Psarommatis, Gökan May, Paul-Arthur Dreyfus, and Dimitris Kiritsis. Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research. *International Journal of Production Research*, 58(1):1–17, April 2019. doi: 10.1080/00207543.2019.1605228.

[161] Louise Wright and Stuart Davidson. How to tell the difference between a model and a digital twin. *Advanced Modeling and Simulation in Engineering Sciences*, 7(1):1–13, 2020. doi: 10.1186/s40323-020-00147-4.

[162] Corinna Cortes, Giulia DeSalvo, Claudio Gentile, Mehryar Mohri, and Ningshan Zhang. Region-based active learning. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 2801–2809. PMLR, 16–18 Apr 2019. URL https://proceedings.mlr.press/v89/cortes19a.html.

[163] Chun-Teh Chen and Grace X. Gu. Generative deep neural networks for inverse materi-

als design using backpropagation and active learning. *Advanced Science*, 7(5):1902607, 2020. doi: 10.1002/advs.201902607.

[164] Xufang Zhang, Lei Wang, and John Dalsgaard Sørensen. REIF: A novel active-learning function toward adaptive kriging surrogate models for structural reliability analysis. *Reliability Engineering & System Safety*, 185:440–454, may 2019. doi: 10.1016/j.ress.2019.01.014.

[165] Stephen W. Tsai and Edward M. Wu. A general theory of strength for anisotropic materials. *Journal of Composite Materials*, 5(1):58–80, 1971. doi: 10.21236/ada306350.

[166] Xiu Yang, Guzel Tartakovsky, and Alexandre Tartakovsky. Physics-information-aided kriging: Constructing covariance functions using stochastic simulation models, 2018.

[167] Michael A. Gelbart, Jasper Snoek, and Ryan P. Adams. Bayesian optimization with unknown constraints, 2014.

[168] Victor Picheny. A Stepwise uncertainty reduction approach to constrained global optimization. In Samuel Kaski and Jukka Corander, editors, *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 787–795, Reykjavik, Iceland, 22–25 Apr 2014. PMLR. URL https://proceedings.mlr.press/v33/picheny14.html.

[169] Robert B. Gramacy, Genetha A. Gray, Sébastien Le Digabel, Herbert K. H. Lee, Pritam Ranjan, Garth Wells, and Stefan M. Wild. Modeling an augmented lagrangian for blackbox constrained optimization. *Technometrics*, 58(1):1–11, 2016. doi: 10.1080/00401706.2015.1014065.

[170] Meng Li, Mohammadkazem Sadoughi, Chao Hu, Zhen Hu, Amin Toghi Eshghi, and Soobum Lee. High-dimensional reliability-based design optimization involving highly

nonlinear constraints and computationally expensive simulations. *Journal of Mechanical Design*, 141(5):051402, 2019. doi: 10.1115/1.4041917.

[171] Anirban Basudhar, Christoph Dribusch, Sylvain Lacaze, and Samy Missoum. Constrained efficient global optimization with support vector machines. *Structural and Multidisciplinary Optimization*, 46(2):201–221, 2012. doi: 10.1007/s00158-011-0745-5.

[172] Clément Chevalier, Julien Bect, David Ginsbourger, Emmanuel Vazquez, Victor Picheny, and Yann Richet. Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4): 455–465, 2014. doi: 10.1080/00401706.2013.860918.

[173] Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, 2012.

[174] Zequn Wang and Pingfeng Wang. A maximum confidence enhancement based sequential sampling scheme for simulation-based design. *Journal of Mechanical Design*, 136 (2), December 2014. doi: 10.1115/1.4026033. 021006.

[175] M. K. Sadoughi, Chao Hu, Cameron A. MacKenzie, Amin Toghi Eshghi, and Soobum Lee. Sequential exploration-exploitation with dynamic trade-off for efficient reliability analysis of complex engineered systems. *Structural and Multidisciplinary Optimization*, 57(1):235–250, July 2017. doi: 10.1007/s00158-017-1748-7.

[176] Peter Koepernik and Florian Pfaff. Consistency of gaussian process regression in metric spaces. *Journal of Machine Learning Research*, 22(244):1–27, 2021. URL http://jmlr.org/papers/v22/21-0853.html.

[177] Jun Wang and Jida Huang. Functionally graded non-periodic cellular structure design

and optimization. *Journal of Computing and Information Science in Engineering*, 22 (3), November 2022. doi: 10.1115/1.4053039.

[178] Payam Ghassemi and Souma Chowdhury. An extended bayesian optimization approach to decentralized swarm robotic search. *Journal of Computing and Information Science in Engineering*, 20(5), 2020. doi: 10.1115/1.4046587.

[179] John C. Platt. *Probabilities for SV Machines*, pages 61–73. The MIT Press, Cambridge, MA, USA, 2000. ISBN 9780262283977.

[180] Daniela Castro-Camilo and Raphaël Huser. Local likelihood estimation of complex tail dependence structures, applied to u.s. precipitation extremes. *Journal of the American Statistical Association*, 115(531):1037–1054, September 2020. doi: 10.1080/01621459.2019.1647842.

[181] Brian J. Reich and Benjamin A. Shaby. A hierarchical max-stable spatial model for extreme precipitation. *The Annals of Applied Statistics*, 6(4):1430–1451, 2012. doi: 10.1214/12-AOAS591.

[182] Corina Birghila, Maximilian Aigner, and Sebastian Engelke. Distributionally robust tail bounds based on wasserstein distance and f-divergence, June 2021.

[183] Martin Schlather. Models for stationary max-stable random fields. *Extremes*, 5(1): 33–44, 2002. doi: 10.1023/a:1020977924878.

[184] Jayde Eustace. *An Extreme Value State Space Model with Gumbel Marginals*. Master's thesis, Memorial University of Newfoundland, Newfoundland and Labrador, Canada, April 2015. URL http://research.library.mun.ca/id/eprint/8435.

[185] Jinjun Tong and Jayde Eustace. *Time Series and State Space Model with Generalized Extreme Value Distributed Marginals and α-Stable Distributed Errors*. Phd's thesis,

Memorial University of Newfoundland, Newfoundland and Labrador, Canada, May 2018. URL http://research.library.mun.ca/id/eprint/13314.

[186] Jonathan Jalbert, Anne-Catherine Favre, Claude Bélisle, and Jean-François Angers. A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over north america. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 66(5):941–962, January 2017. doi: 10.1111/rssc.12212.

[187] Sebastian Engelke and Jevgenijs Ivanovs. Sparse structures for multivariate extremes. *Annual Review of Statistics and Its Application*, 8(1):241–270, March 2021. doi: 10.1146/annurev-statistics-040620-041554.

[188] Axel Bücher and Chen Zhou. A horse race between the block maxima method and the peak–over–threshold approach. *Statistical Science*, 36(3):360–378, 2021. doi: 10.1214/20-sts795.

[189] Raphaël Huser and Jennifer L. Wadsworth. Advances in statistical modeling of spatial extremes. *WIREs Computational Statistics*, 14(1), November 2020. doi: 10.1002/wics.1537.

[190] Paul G. Constantine. *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*. SIAM, 2015. ISBN 9781611973853. doi: 10.1137/1.9781611973860.

[191] Paul G. Constantine, Eric Dow, and Qiqi Wang. Active subspace methods in theory and practice: Applications to kriging surfaces. *SIAM Journal on Scientific Computing*, 36(4):A1500–A1524, 2014.

[192] Ruda Zhang, Simon Mak, and David Dunson. Gaussian process subspace regression for model reduction, 2021.

[193] Nathan Wycoff, Mickael Binois, and Stefan M. Wild. Sequential learning of active subspaces. *Journal of Computational and Graphical Statistics*, 30(4):1224–1237, 2021.

[194] Shancong Mou, Michael Biehler, Xiaowei Yue, Jeffrey H. Hunt, and Jianjun Shi. SPAC: Sparse sensor placement-based adaptive control for high precision fuselage assembly. *IISE Transactions*, pages 1–11, October 2022. doi: 10.1080/24725854.2022.2116133.

[195] Yinan Wang, Kaiwen Wang, Wenjun Cai, and Xiaowei Yue. NP-ODE: Neural process aided ordinary differential equations for uncertainty quantification of finite element analysis. *IISE Transactions*, 54(3):1–16, April 2021. doi: 10.1080/24725854.2021.1891485.

[196] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009. doi: 10.1137/07070111x.

[197] Marco Signoretto, Lieven De Lathauwer, and Johan A. K. Suykens. A kernel-based framework to tensorial data analysis. *Neural Networks*, 24(8):861–874, 2011. doi: 10.1016/j.neunet.2011.05.011.

[198] Xiaowei Yue, Jin Gyu Park, Zhiyong Liang, and Jianjun Shi. Tensor mixed effects model with application to nanomanufacturing inspection. *Technometrics*, 62(1):116–129, 2020.

[199] Yinan Wang, Weihong "Grace" Guo, and Xiaowei Yue. Tensor decomposition to compress convolutional layers in deep learning. *IISE Transactions*, 54(5):481–495, April 2022. doi: 10.1080/24725854.2021.1894514.

[200] Rose Yu, Guangyu Li, and Yan Liu. Tensor regression meets gaussian processes. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceed-*

ings of Machine Learning Research, pages 482–490. PMLR, 09–11 Apr 2018. URL https://proceedings.mlr.press/v84/yu18a.html.

[201] Adrià Garriga-Alonso, Carl Edward Rasmussen, and Laurence Aitchison. Deep convolutional networks as shallow gaussian processes, 2018.

[202] Roman Novak, Lechao Xiao, Jaehoon Lee, Yasaman Bahri, Greg Yang, Jiri Hron, Daniel A. Abolafia, Jeffrey Pennington, and Jascha Sohl-Dickstein. Bayesian deep convolutional networks with many channels are gaussian processes, 2018.

[203] Lifang He, Chun-Ta Lu, Guixiang Ma, Shen Wang, Linlin Shen, Philip S. Yu, and Ann B. Ragin. Kernelized support tensor machines. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1442–1451. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/he17a.html.

[204] Lynn Houthuys and Johan A. K. Suykens. Tensor learning in multi-view kernel PCA. In *Artificial Neural Networks and Machine Learning – ICANN 2018*, pages 205–215. Springer International Publishing, 2018. doi: 10.1007/978-3-030-01421-6_21.

[205] Alex Campbell and Pietro Liò. tvgp-vae: Tensor-variate gaussian process prior variational autoencoder, 2020.

[206] Andrew V. Knyazev and Merico E. Argentati. Principal angles between subspaces in an a-based scalar product: Algorithms and perturbation estimates. *SIAM Journal on Scientific Computing*, 23(6):2008–2040, January 2002. doi: 10.1137/s1064827500377332.

[207] Qibin Zhao, Guoxu Zhou, Liqing Zhang, and Andrzej Cichocki. Tensor-variate gaussian processes regression and its application to video surveillance. In *2014 IEEE In-*

ternational Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1265–1269. IEEE, IEEE, may 2014. doi: 10.1109/icassp.2014.6853800.

[208] Hua Zhou, Lexin Li, and Hongtu Zhu. Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502): 540–552, 2013.

[209] Carlos Llosa-Vite and Ranjan Maitra. Reduced-rank tensor-on-tensor regression and tensor-variate analysis of variance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2):2282–2296, feb 2023. doi: 10.1109/tpami.2022.3164836.

[210] Zenglin Xu, Feng Yan, Yuan, and Qi. Infinite tucker decomposition: Nonparametric bayesian models for multiway data analysis, 2011.

[211] Lingbing Tang, Pin Peng, Qibin Zhao, Lihua Gui, and Luochang Qing. Tensor-based gaussian processes regression using a probabilistic kernel with information divergence. *International Journal of Simulation–Systems Science & Technology*, 16(5), January 2015. doi: 10.5013/ijssst.a.16.05.13.

[212] Yipeng Liu, Jiani Liu, and Ce Zhu. Low-rank tensor train coefficient array estimation for tensor-on-tensor regression. *IEEE Transactions on Neural Networks and Learning Systems*, 31(12):5402–5411, dec 2020. doi: 10.1109/tnnls.2020.2967022.

[213] Eric F. Lock. Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics*, 27(3):638–647, 2018.

[214] Changqing Xu and Kaijie Xu. Random tensors and their normal distributions, 2019.

[215] Michael Brazell, Na Li, Carmeliza Navasca, and Christino Tamon. Solving multilinear systems via tensor inversion. *SIAM Journal on Matrix Analysis and Applications*, 34 (2):542–570, 2013. doi: 10.1137/100804577.

[216] Mohammad Arashi. Some theoretical results on the tensor elliptical distribution. *Journal of Data Science and Modeling*, 1(2):27–38, 2021. doi: 10.22054/jcsm.2021.47404.1019.

[217] Yu Wang, Zeyu Sun, Dogyoon Song, and Alfred Hero. Kronecker-structured covariance models for multiway data. *Statistics Surveys*, 16:238–270, January 2022. doi: 10.1214/22-ss139.

[218] Martin Ohlson, M. Rauf Ahmad, and Dietrich Von Rosen. The multilinear normal distribution: Introduction and some basic properties. *Journal of Multivariate Analysis*, 113:37–47, 2013. doi: 10.1016/j.jmva.2011.05.015.

[219] Jamie Fairbrother, Amanda Turner, and Stein W. Wallace. Scenario generation for single-period portfolio selection problems with tail risk measures: Coping with high dimensions and integer variables. *INFORMS Journal on Computing*, 30(3):472–491, August 2018. doi: 10.1287/ijoc.2017.0790.

[220] Xi Chen and Kyoung-Kuk Kim. Efficient VaR and CVaR measurement via stochastic kriging. *INFORMS Journal on Computing*, 28(4):629–644, November 2016. doi: 10.1287/ijoc.2016.0705.

[221] Pradip K. Saha. *Aerospace Manufacturing Processes*. CRC Press, September 2016. doi: 10.1201/9781315367965.

[222] Richard L. Smith. Max-stable processes and spatial extremes. Unpublished Preprint, 1990. URL https://rls.sites.oasis.unc.edu/postscript/rs/spatex.pdf.

Appendices

Appendix A

Appendix of Chapter 4

A.1 Proof of Proposition 4.1

For any $\mathbf{x}_i \in S_i$, the probability of failure is

$$\begin{aligned}
&\mathbb{P}\left(\hat{h}_i(\mathbf{x}_i) \geq \xi \,\middle|\, \mathbf{x}_i \in S_{i-1}\right) \\
&\leq \mathbb{P}\left(\hat{h}_i(\mathbf{x}_i) \geq \hat{\mu}^h_{i-1}(\mathbf{x}_i) + \beta_{i-1}\hat{\sigma}^h_{i-1}(\mathbf{x}_i) \,\middle|\, \mathbf{x}_i \in S_{i-1}\right) \\
&= \gamma_{i-1},
\end{aligned}$$

and it holds for $\forall\, i$. Then, by Boole's inequality (union bound), we have

$$\begin{aligned}
&\mathbb{P}\left(\bigcup_i \left(\hat{h}(\mathbf{x}_i) \geq \xi \,\middle|\, \mathbf{x}_i \in S_{i-1}\right)\right) \\
&\leq \sum_i \mathbb{P}\left(\hat{h}_i(\mathbf{x}_i) \geq \xi \,\middle|\, \mathbf{x}_i \in S_{i-1}\right) \\
&\leq \sum_i \gamma_{i-1}.
\end{aligned}$$

By setting $\gamma_i = \gamma$ for $\forall\, i$, the failure probability can be simplified as

$$\sum_i \gamma_{i-1} = N\gamma \triangleq \zeta \implies \gamma = \frac{\zeta}{N}.$$

A.2 Proof of Proposition 4.5

First, we show that S_n converges to a nonempty subset of Ω_S as $n \to \infty$ by the consistency of GP regression over Ω.

Lemma A.1. *Let S_0 be the initial estimated safe region that contains a safe design point at least. With any dense sequence of design points in S_n including PhysCAL, S_n converges to $S_* \subseteq \Omega_S$.*

Proof. Let us denote a union of estimated safe regions as

$$T_n = \bigcup_{i=0}^{n} S_i.$$

Since T_n is monotone (i.e., $T_n \subset T_{n+1} \subset \cdots$), and bounded by Ω, it converges almost surely to a subset of Ω as $n \to \infty$. Let $\hat{T} = \lim_{n\to\infty} T_n$, and suppose any dense sequence of design points $T_* = \{\mathbf{x}_1, \mathbf{x}_2, \ldots\} \subset \hat{T}$. Then, the consistency of GP regression holds over T_* [176]. That is, $\hat{f}_n$ and $\hat{h}_n$ converge almost surely to the best MSE predictor $\hat{f}_*$ and $\hat{h}_*$ in S_*. Since $\hat{h}_*$ is unbiased and has zero-variance over T_*, for a large enough n, we have that

$$\begin{aligned}
\lim_{n\to\infty} S_n &= \{\mathbf{x} \in \Omega |\ \hat{\mu}_*^h(\mathbf{x}) + \gamma \hat{\sigma}_*^h(\mathbf{x}) < \xi\} \\
&= \{\mathbf{x} \in \Omega |\ h(\mathbf{x}) < \xi\} \\
&\triangleq S_* \subseteq \Omega_S,
\end{aligned}$$

where $\hat{h}_*(\mathbf{x}) \sim \mathcal{N}(\hat{\mu}_*^h,\ \hat{\sigma}_*^h = 0)$. □

Lemma A.1 allows us to utilize the consistency of GP models to show the degeneracy of $J(\mathbf{x})$

in S_*. Recall that both J_f and J_h are nonnegative, and they are upper bounded by

$$J_f(\mathbf{x}) \leq \mathbb{V}_{S_n^+}(\hat{f}_n) \leq \sup_{\mathbf{x}\in S_*} \mathrm{Var}\left(\hat{f}_n(\mathbf{x})\right),$$
$$J_h(\mathbf{x}) \leq \eta^2(\mathbf{x}) \leq \sup_{\mathbf{x}\in S_*} \alpha^2 \mathrm{Var}\left(\hat{h}_n(\mathbf{x})\right),$$

for any $\mathbf{x} \in S_*$. Then, by the consistency of GP, we have

$$\lim_{n\to\infty} \sup_{\mathbf{x}\in S_*} \mathrm{Var}\left(\hat{f}_n(\mathbf{x})\right) = 0,$$
$$\lim_{n\to\infty} \sup_{\mathbf{x}\in S_*} \mathrm{Var}\left(\hat{h}_n(\mathbf{x})\right) = 0.$$

Thus, by the squeeze theorem, $J(\mathbf{x})$ uniformly converges to zero over S_*.