SEQUENTIAL EXPERIMENTAL DESIGN UNDER COMPETING PRIOR KNOWLEDGE

A Thesis
Presented to
The Academic Faculty

by

Justin T. Vastola

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
May 2013

SEQUENTIAL EXPERIMENTAL DESIGN UNDER COMPETING PRIOR KNOWLEDGE

Approved by:

Dr. Jye-Chyi Lu, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Martha Grover, Co-Advisor
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Kamran Paynabar
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Jianjun Shi
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Roshan Joseph Vengazhiyil
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 5 December 2012

To my wife,

Julia.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Professor Jye-Chyi Lu. He has provided excellent guidance, insightful advice, and calming encouragement. His support goes beyond my academic life, as his kindness extended to caring for my personal well-being.

I am also very grateful of my co-advisor, Professor Martha Grover, for her supportive advice, encouragement, and kindness throughout my doctoral studies. In addition, I would like to thank Professor Jianjun Shi, Professor Roshan Vengazhiyil, and Professor Kamran Paynabar for generously offering their time to serve as committee members. Additionally, I am appreciative for the additional members of the ChBE research group: Professor Dennis Hess, Dr. Galit Levitin, and Michael Casciato.

I am very thankful for the faculty and staff members in ISyE. In particular, I am appreciative of Pam Morrison for her kindness and willingness to always help out. Additionally, I am extremely grateful for Professor Gary Parker for his honest advice and encouragement that assisted me in both academically and personally.

I am grateful for the many friends and colleagues that I have made at Georgia Tech. Among many, I am thankful for the friendships, moral support, and advice of Dr. Dong Gu Choi, Dr. Sungil Kim, Srdjan Lesaja, and Carlos Valencia.

I would like to express deep gratitude to my parents, my brother, and my sister for their support throughout my life, especially through this trying period. Last, but definitely not least, a heartfelt appreciation goes to my wife, Julia, for her never-ending support, patience, and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The overall theme of this work is the incorporation of engineering information into experimental design processes where the objective is to find the optimal operating condition of a system. In particular, novel methodology is developed for optimizing experimental systems in a high-cost, low-resource setting where the underlying response surface may be highly complex. The first part of this work is concerned with building an initial experimental design, and the second part provides a new batch sequential experimental design process. In both parts, advancements in both model building in the presence of multiple sources of information and designing experiments are provided. An introduction to this work, as well as problem motivation and objectives are provided in chapter 1.

In the second chapter, an initial experimental design criteria, applicable to computer and physical experiments, is proposed for optimization problems by combining multiple, potentially competing, sources of prior information–engineering models, expert opinion, and data from past experimentation on similar, non-identical systems. By leveraging prior information, some initial resources are dedicated to solving the optimization problem, in turn, potentially reducing the total number of resources needed in follow-up experimentation. New methodology is provided for incorporating and combining conjectured models and data into both the initial modeling and design stages. As a result, the design criteria is flexible in how it balances space filling and objective oriented properties in the presence of conjectured prior information. An application to a thin-film growth study is provided in addition to a detailed numerical study of the design properties.

In the third, and final, chapter, a batch sequential design procedure is proposed for

optimizing high-cost, low-resource experiments with complicated response surfaces. The success in the proposed approach lies in melding a flexible, sequential design algorithm with a powerful local modeling approach. Batch experiments are designed sequentially in a manner so that they adapt to balance space-filling properties and the search for the optimal operating condition. In addition, the experimental design region adapts by balancing the elimination of poor design regions with the focus on potential optimal design regions. Local model calibration and averaging techniques are introduced to easily allow incorporation of statistical models and engineering knowledge, even if such knowledge pertains to only subregions of the complete design space. The overall process iterates between adapting designs, adapting models, and updating engineering knowledge over time. Performance of the process is demonstrated on a nanowire synthesis system as well as on an optimization test function.

CHAPTER I

INTRODUCTION

Motivated by the emerging field of nanotechnology, the intent of this research is to provide new advances in experimental design methodology to address the challenges found in high-cost, low-resource experimental systems with complex response surfaces. The underlying theme of this work is the incorporation of (potentially competing) sources of engineering knowledge via a direct interaction with engineers to build irregular experimental designs. By leveraging this expertise, experimental designs can be made more efficient and effective, conceivably enhancing the development of advanced engineering systems.

Nanotechnology can be defined as "the construction and use of functional structures designed at the atomic or molecular scale with at least one characteristic dimension measured in nanometers ($1nm = 10^{-9}m$, which is about 1/50,000 of the width of human hair)" [18]. Nanotechnology R&D can be costly in both time and/or resources, where experiments in some cases take up to a day for completion [55] and both the preparation and characterization of materials can be expensive [1]. As nanotechnology is expected to impact many sectors of daily life in the near future [37], it is imperative to address new statistical challenges arising from nanomanufacturing for both understanding and large-scale production of nanoproducts [18]. As pointed out in Lu, Jeng, and Wang [37], Basumallick, Das, and Mukherjee [5], and Dasgupta *et al.* [18], new challenges introduced in nanoresearch include, and are not limited to, highly non-linear response surfaces with multiple optima, high experimental costs, and large, irregularly shaped failure regions. We address these issues based the methodologies and recommendations discussed in Lu, Jeng, and Wang [37],

Dasgupta *et al.* [18], and Dasgupta [17].

Concerning the choice of experimental design methods, Lu, Jeng, and Wang [37] write:

> "In some applications, simply full factorial or fractional factorial designs may be sufficient for model building and process understanding. However, as shown in several examples, nanofabrication processes can be very sensitive to small changes in controllable and noise factors. Moreover the process outcomes may not be continuous random variables. In this case, irregular designs or new experimental design methods might be considered. Finally, special experimental constraints due to physical limitations may limit the use of certain designs."

In this work, we concern ourselves with addressing the latter challenges expressed here, cases in which full factorial or fractional-factorial designs are insufficient. More specifically, based on the problem motivation, we follow the additional advice of Lu, Jeng, and Wang: "Because the response surface is very rugged, in order to improve the data collection and analysis methods used in this paper, we recommend the experimental designs (and their extensions) commonly used in computer experiments for efficiently collecting costly data." The experimental design strategies proposed in this work balance space-filling properties, as motivated from computer experiments [45], and optimization strategies, in turn, providing flexible procedures for optimizing complex response surfaces in high-cost, low-resource scenarios.

A second challenge in nanotechnology is addressed in this work by incorporating engineering knowledge for modeling and designing experiments. As noted in Lu, Jeng, and Wang [37]: "When the physical experimentation is costly and time consuming, the use of computers for process and device simulations is expected to become more prevalent in nanoresearch." Engineering knowledge doesn't just exist in the form of computer experiments. Other sources include analytical models, expert opinion,

and information coming from past experimentation on differing systems. Flexible methodology is provided to incorporate all these sources in both the modeling and design stages, even when sources pertain to subregions of the design space and/or have different levels of relevance to the current system under study. Moreover, emphasis is placed on the direct interaction of statisticians and engineers, as it is important to recognize that:

> "Nanotechnology is a multidisciplinary subject. When statisticians are more involved in understanding the issues in nanoresearch, more effective procedures can be developed to deal with challenging application problems. Statisticians should be important players in novel scientific discoveries." [37]

In the rest of this work, a comprehensive experimental design procedure is developed to address the challenges discussed above. In the second chapter, an initial experimental design criteria is proposed for optimization problems by combining multiple, potentially competing, sources of prior information–engineering models, expert opinion, and data from past experimentation on similar, non-identical systems. New methodology is provided for incorporating and combining conjectured models and data into both the initial modeling and design stages. As a result, the design criteria is flexible in how it balances space-filling and objective oriented properties in the presence of conjectured prior information. Relevant literature for modeling engineering information and designing experiments is reviewed and compared to the proposed work. To conclude the chapter, an application to a thin-film growth study is provided in addition to a detailed numerical study of the design properties.

In the third, and final, chapter, a batch sequential design procedure is proposed for optimizing high-cost, low-resource experiments with complicated response surfaces. The success in the proposed approach lies in melding a flexible, sequential design algorithm with a powerful local modeling approach. Batch experiments are designed

sequentially in a manner so that they adapt to balance space-filling properties and the search for the optimal operating condition. In addition, the experimental design region adapts by balancing the elimination of poor design regions with the focus on potential optimal design regions. Local model calibration and averaging techniques are introduced to easily allow incorporation of statistical models and engineering knowledge, even if such knowledge pertains to only subregions of the complete design space. The overall process iterates between adapting designs, adapting models, and updating engineering knowledge over time. Literature is reviewed in relation to the modeling and sequential experimental design strategies introduced here. Concluding, performance of the process is demonstrated on a nanowire synthesis system as well as on an optimization test function.

CHAPTER II

A FRAMEWORK FOR INITIAL EXPERIMENTAL DESIGN IN THE PRESENCE OF COMPETING PRIOR KNOWLEDGE

2.1 Introduction

When studying new and advanced technologies, there is often a lack of understanding of the true behavior of the process. In such cases, experimental designs are typically developed for exploratory purposes–spreading out points, in some sense uniformly, in attempt to understand and model the process as a whole. Since experimentation on new, advanced systems can be very costly, experimental designs that place points in regions that have little or no potential for being an optimal experimental input can be unnecessarily wasteful. Although a process may be new, prior sources of information can be present: researches can conjecture expert opinion data based on decades of knowledge; engineering models may exist; and data from past experiments on similar, yet non-identical, systems can provide insight into the new system.

By leveraging prior information, we propose a novel design criteria to use for solving optimization problems–problems with a goal of maximizing an objective function. In this work, we focus on maximizing the probability that a product is produced within a tolerance limit of a nominal level. The proposed methodology is applicable for both computer and physical experiments when the initial sample size is fixed before experimentation. Additionally, it can be used as a stand alone design process or in conjunction with sequential experimental design procedures such as the SMED procedure of Dasgupta [17], expected improvement designs (see Santner, Williams, and Notz [45]), or any other designs for optimization.

The success of our approach stems from the ability to build an accurate prior model to estimate an objective function before any experimentation. Prior models can be any combination of expert opinion data (see Reese *et al.* [44]), engineering models, both analytic or metamodels, and data and models from past experimentation on similar, non-identical, systems. For presentation purposes, the latter source of information is focused upon, not because of its importance, but due to its absence from literature. We do not downplay the importance of expert opinion data and engineering models. They are at least as important as past experimental data on different systems and, in many circumstances, may be the only reliable sources of prior information. By introducing a concept of *trust*, the expressed uncertainty in or quality of the prior information, all models from prior knowledge are combined into one, unified model. As a result, three measures are associated with the model at each design point: the expected value, the variance, and the trust.

A model-based experimental design procedure based off of the unified model is presented. In particular, a non-trivial modification to a flexible class of experimental designs known as *minimum energy designs* [17] is proposed to balance the trade-off between the space-filling and optimization oriented objectives in the initial experimental design. The novelty in our design is the direct incorporation of the conjectured nature of the prior model into the design criteria, in turn, allowing the trust to automatically balance this trade-off. Additionally, we provide details on extending the purely sequential nature of Dasgupta's work to select a batch of n points simultaneously.

2.2 Literature Review

The methodology we propose involves new advancements in data modeling and experimental design. We separately review the relevant literatures in relation to our work.

2.2.1 Modeling with Combined Data Sources

Incorporating data from many sources is not a new topic in statistics, extending at least as far back to meta-analytic literature pioneered by Zeckhauser [57]. The goal of meta-analytic approaches is to combine information across multiple studies, typically when only summary statistics are available, to increase the accuracy of the statistical analysis. For a review of meta-analysis, see Hedges and Olkin [24]. Relying on meta-analytic literature isn't sufficient for our purposes for two main reasons. First, meta-analytic methods typically assume a certain level of homogeneity among the data sources, whereas the data sources we consider come from very different, competing sources. Second, before initial experimentation, no observed data is available on the current system. Instead, information is drawn from expert opinions, engineering models, and past experiments on similar, non-identical systems.

Combining engineering models with physical data is not a new field in statistics. Model calibration, in which computer [30] or analytical models [29] are adjusted based on the collected physical data, is one such approach. Engineering knowledge has also been used for model building. Zhao *et al.* [58] built a model by decomposing the expected response into an analytic engineering model and a Gaussian process model. Reese *et al.* [44] proposed a modeling approach to incorporate physical and computer data with expert opinion data in a hierarchical Bayesian framework. Additionally, model prediction has been addressed by Craig *et al.* [16] by using a computer simulator and past data from the actual, true system.

Our problem formulation differs from these works in one key aspect: no physical data on the new system is present to include in the modeling or for updating the computer or engineering models. Hence, new methodology is proposed in the absence of any "true" data. In particular, we combine all prior sources of information via a direct interaction with engineers. Additionally, new methodology for incorporating data from past experiments on similar, yet non-identical, engineering systems is

proposed.

2.2.2 Model-based and Space Filling Experimental Designs

Space filling designs, designs in which design points are chosen to spread out "evenly" over the design region, are common in, but not limited to, computer experiments. Typical examples of these designs include Latin hypercube, maximin, minimax, and uniform designs. See Santner, Williams, and Notz [45] for more on these types of designs. These designs are useful for obtaining a model over the complete design space, especially when the true response surfaces are quite complex, however, when designing experiment for process optimization, a global model may not be necessary. Instead, a design that balances space-filling properties with the optimization objective can be just as effective and more efficient than designs that simply spread out points evenly. By using prior sources of information, our design criteria is able to balance the space-filling and objective criteria so that resources are not unnecessarily wasted on regions with low potential of satisfying the optimization objective. For a recent approach in leveraging engineering models to design a computer experiment, see Ba [3], who's motivation–efficiently enhancing model prediction–is fundamentally different than our optimization objective.

Model-based experimental designs exploit a model's characteristics in order to more efficiently design an experiment to satisfy some criteria. Traditionally these designs have been addressed through the *alphabetical* experimental designs, e.g., A-, D-, or T-optimal designs. Typically, the goals of these designs are to minimize some form of variance. For more information, see Federov and Hackl [20]. Recent research in computer experiments provides additional advances in model-based experimental design. Approaches include maximizing the entropy, minimizing prediction error, or maximizing an improvement criteria [45]. Each of these methods relies on a model-free initial experimental design to collect data, which in turn is used to build a model

for follow-up, model-based designs. The initial design plan typically is given little attention, as experimenters resort to spreading points evenly over the design space. The novelty of our approach is that initial experiments can be model-based, thus saving resources from the beginning of the experimental process.

Our initial design plan is based on the concept of the *minimum energy design* (MED) proposed by Dasgupta [17]. The basic idea behind MED's begins with the assignment of a "charge" to each point in the design space. Following, n points in the design space are chosen to minimize some form of potential energy. A similar design, the *minimum potential design* (MPD), was used by Kessels *et al.* [31] and is provided in the statistical software JMP. In MPD's, potential design points are essentially assigned an identical charge, resulting in a completely space-filling design. Alternatively, MED's allow the charges to vary as a function of the response surface so that designs can be objective oriented instead of purely space-filling. Dasgupta introduced this concept to use in a purely sequential experimental design procedure. We propose a new batch minimum energy design that incorporates the trust model in addition to the response surface.

2.3 Building a Unified Model

A prior model is built by unifying multiple, competing sources of information into one model. Information can come in the form of expert opinion, engineering models, and past data or models from previous experiments on similar, non-identical systems. The latter data source is focused on here due to its lacking presence in literature, yet we reiterate that expert opinion and engineering models are just as important and may be the only existing or reliable sources of information before experimentation. Before discussing the incorporation of past data on non-identical systems, information on modeling expert opinion and engineering models is provided.

Eliciting expert opinion from researchers is a complex topic unto itself, combining

ideas from both statistics and psychology. A good discussion of such methods and concerns can be found in Garthwaite, Kadane and O'Hagan [22]. Our methodology is quite general in that it can include any techniques for eliciting information to build a unified model. We do not focus on these modeling strategies here, but ensure that, with a specification of a trust model, models from expert elicitations can be directly incorporated as one of the competing models. For its ease of use and flexibility, we base our expert opinion modeling methodology on those expressed in Reese *et al.* [44]. Details are withheld for the appendix. There is a great wealth on building models from computer experiments. Interested readers may refer to Santner, Williams, and Notz [45]. We do not focus on this task but merely mention that these modeling procedures can be directly used in our approach for incorporating computer data. When dealing with analytical engineering models, only a mean model, with no specification of variance, is given. Examples of works incorporating engineering models can be found in Joseph and Melkote [29] and Zhao *et al.* [58], who provide modeling methods when data from the real system are available. If no data is available, as is the case for designing an initial experiment, other techniques are to be used. We provide a brief discussion in the appendix on one alternative approach to incorporating engineering models.

Using data from past experiments on different experimental systems is a topic that has received little attention, if any at all. Although the past data is coming from a different system, information can still be gleaned from the past data since many engineering processes behave in similar, albeit non-identical, manners. In reference to physical experiments, a great motivating example is provided in Zong and Watkins [59]. They consider a supercritical fluid deposition (SFD) process: a process when a gas is placed under a high pressure resulting in fluid like properties. In particular, solid material can be dissolved in a supercritical fluid and deposited on a surface. The kinetics behind supercritical deposition are not well understood, but a

similar, yet quite different process, chemical vapor deposition (CVD), is better understood. If a connection between the CVD and SFD processes can be made, then when describing properties such as the growth rate of a compound, experimenters can indirectly rely on growth rates as understood from the CVD point of view.

For a SFD process involving three compounds, Zong and Watkins depict how for two compounds, the rough, overall trend of the growth rate is captured as described by the CVD kinetics. For a third compound, they explain how the trend, although still accurately shaped, is shifted in the domain space. This shift, however, can be understood based on the differing physics between the role of the third substance in the CVD process versus the SFD process. In this circumstance, data from past CVD experiments can be translated to the domain of the new SFD experiment, giving "data" or conjectured models that can be used as a reference for designing a new experiment on the new system. We formalize these ideas into a statistical modeling framework below.

2.3.1 Past Experiments on Non-identical Systems

At the most basic level, to build a model based on previous experiments, a new design point and a response need to be specified for each design point and response coming from a previous experiment. Denote the past and current design spaces by D_p and D_c, respectively, and let $\xi_p \in D_p$ and $\xi_c \in D_c$ denote the experimental designs of the previous and current experiments, respectively. Here and throughout the paper, *current* refers to aspects of the experiment that we are seeking to build a design for. With the existence of a relationship between past and current systems, *pseudo-data* can to be generated as follows:

1. each design point of the previous experiment is mapped to the design space of the current experiments:

$$\tilde{\boldsymbol{x}}_i \in \xi_p \mapsto \boldsymbol{x}_i \in \xi_c;$$

2. the data responses, $\tilde{y}_i$, from the previous experiments will be augmented by adding a value λ_i to best reflect the expected result of the current system:

$$\tilde{y}_i \mapsto \tilde{y}_i + \lambda_i \equiv y_i.$$

Two different methods are considered to implement the above steps: updating each individual design point and response explicitly and updating model parameters, in addition to the design space by implicitly updating all the design points and responses. The first scenario arises when there is no statistical or analytical model to begin with or when model parameters may not have a directly interpretable relationship to relate the past and current systems. This scenario often arises since, typically, the importance of literature on new, advanced technologies can be the illustration of a phenomena's existence rather than trying to understand the process on a range of settings. As a result, past experimental data is provided for only a few design points. For example, Ye *et al.* [56]present an environmentally friendly method for fabricating metal nanoparticle-multiwall carbon nanotube composites using supercritical fluid synthesis. In terms of nanoparticle size, for each of three types of metals, they present the size measurement for one experimental run, disallowing any formal statistical treatment for model building.

When converting the previous design points individually, the modeling process is similar to modeling expert opinion data but with two differences. First, the true response is not conjectured; instead, the expected difference between the past and future responses is conjectured. The resulting *pseudo-data* is not a conjectured mean response but the past response shifted by the conjectured mean difference. This distinction leads to a slight difference in modeling compared to existing approaches to modeling expert opinion data since the lack of knowledge is modeled through the vector, $\boldsymbol{\lambda}$, of the expected differences in the past and conjectured responses instead of $\boldsymbol{Y}$, the conjectured expected response. The other distinction arises from the conversion

of the design points in step 1. In converting the past experimental data, for each $\tilde{\boldsymbol{x}}_i \in \xi_p$, an expert specifies what he or she believes to be the corresponding design point $\boldsymbol{x}_i \in \xi_c$. This conversion is only a postulate, so $\boldsymbol{x}_i$ cannot be viewed as a fixed, known value. An additional source of variability needs to account for this uncertainty in the modeling process. In light of these two differences, traditional methods for modeling elicited data need to be modified. We provide one such modification procedure in the appendix.

The second scenario for modeling past experimental data arises when model parameters are interpretable to the extent that the differences between the past and current experiments can be understood in relation to differences in the parameter values. For example, Casciato *et al.* [11] provide a method for generating silver nanoparticles under high temperature and high pressure. The Arrhenius equation is used to model the dependence between particle size and temperature. In particular the mean particle size, y, relates to temperature T in the following manner:

$$\ln(y) = \beta_0 + E_a \frac{1}{R \cdot T},$$

where R is the Universal gas constant and E_a is the activation energy–an unknown parameter that needs to be estimated in addition to β_0. The activation energy of a process is a quantity that depends on the compounds involved in addition to other system settings. Based on the past experiments to produce silver nanoparticles, suppose new experiments are to be conducted to produce copper nanoparticles instead. Here, D_p and D_c refer to the temperature (input) ranges for the experiments to fabricate silver and copper nanoparticles, respectively. To convert the past data into psuedo-data for the current system, each temperature $\tilde{x}_i \in \xi_p$ can be implicitly translated to its "equivalent" temperature $x_i \in \xi_c$ by the translation required to convert D_p to D_c. Then, since the difference between the reactivity of copper and silver precursors is understood to some degree based on known differing physical principles, the activation energy for the new system can be reasonably conjectured based on

the estimated activation energy of the past experiments. The conjectured change in activation energy can be used to implicitly define the vector $\boldsymbol{\lambda}$ to convert each $\tilde{y}_i$ in the past experiment to conjectured responses y_i for the current experiment. Details are provided later.

In a different scenario, the domain of a model may need to be converted as well as the model parameters. Zong and Watkins [59] show that engineering models for a chemical vapor deposition (CVD) process can be useful for describing the response function of a certain supercritical fluid deposition (SFD) process, thus, providing a link between the CVD and SFD processes. In particular, by solving a system of kinetic equations, they provide the theoretical growth rate of a thin film for the CVD process, in turn, providing an approximate growth rate model for the SFD process. When studying certain aspects of the SFD process, they show how the model can be used almost directly, yet in another instance, they show how this model needs to be (and can be) changed to reflect a difference between the CVD and SFD processes. In particular, when studying the effect of the concentration of hydrogen, $H_{2(g)}$, on the thin film growth rate in the SFD system, Zong and Watkins noted a mismatch between the observed data and the CVD based model. The appropriate "shape" of the model was correct, but the location of the model in the experimental domain was wrong. This mismatch is understood, however, by noting that hydrogen plays an additional role (two roles total) in their SFD process: it serves as a reducing agent for metalization as well as to reduce copper oxides. As a result, the hydrogen in their SFD process in being used more quickly than predicted by the CVD system.

Suppose in the above scenario new experiments are to be conducted on the SFD system to optimize the thin film growth rate based on the amount of hydrogen concentration. Let $\tilde{g}(\tilde{x})$ denote the engineering model to describe the growth rate versus the hydrogen concentration based on the kinetics of the CVD system, where $\tilde{x} \in D_p$, the experimental domain of the past (CVD) system. In order for $\tilde{g}(\tilde{x})$ to be useful

to design an optimization experiment for the SFD process, the additional role that hydrogen plays in the SFD process needs to be accounted for before designing the new experiment. The $H_{2(g)}$ compounds are being used in greater abundance in the SFD system, so each $\tilde{x}_i \in D_p$ needs to be increased to a value $x_i \in D_c$ by relying on engineering knowledge, resulting in the experimental domain of the SFD system. Following, step 2 of the conversion process is addressed by rescaling $\tilde{g}$ so its "shape" is maintained on D_c. The resulting conjectured model is then exploited to design an initial experiment to optimize the SFD system. These ideas are formalized below.

2.3.1.1 Converting Past Experiments by Updating Model Parameters

Consider a model, either statistical or analytical, and data obtained from a past experiment on a system non-identical to the current system. Denote this model by $f(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}})$, where the tilde indicates that the domain and parameter values are from the past experiment. Based on a physical understanding of the differences between the past and current experiments, new parameter values for the current system, $\boldsymbol{\beta}$, and a new domain, $\boldsymbol{X}$, for which the model is relevant are conjectured. The whole past domain space, $\tilde{\boldsymbol{X}}$, is mapped to a new space, $\boldsymbol{X}$, by translation, adding a vector $\boldsymbol{\lambda_X}$, and/or dilation, multiplying by a vector $\boldsymbol{\omega_X}$. In the new domain pertaining to the new process, parameters for the new process, $\boldsymbol{\beta}$, are conjectured in reference to $\tilde{\boldsymbol{\beta}}$, the past parameter. Denote the mean of the conjectured model on the new experimental region by $g(\boldsymbol{X})$. Then $g(\boldsymbol{X}, \boldsymbol{\beta}) = f(\boldsymbol{\omega_X}(\boldsymbol{X} - \boldsymbol{\lambda_X}), \boldsymbol{\beta})$ by definition. With the conjecture for the expected value of the model in hand, the past data can be converted based on the differences between the past and conjectured model.

We do not want to convert the past data by just simply taking the direct difference between the past and conjectured model, namely $g(\boldsymbol{X}, \boldsymbol{\beta}) - f(\tilde{\boldsymbol{X}}, \tilde{\boldsymbol{\beta}})$, when setting $\boldsymbol{X} = \tilde{\boldsymbol{X}}$, since this difference is comparing different "parts" of the curve. Instead, the difference in the mean models needs to be taken on a *comparable domain*. To compare

two models, we consider the difference $g(\boldsymbol{X}, \boldsymbol{\beta}) - g(\boldsymbol{X}, \tilde{\boldsymbol{\beta}})$; that is, we compare both models on the new domain. At each previously observed design point, this difference quantifies the amount that the past data needs to be augmented to represent an "observation" from the new experiment.

Conjectured data for the new system is obtained as follows. Let $\tilde{\mathcal{X}} = [\tilde{\boldsymbol{X}}_1^T; \ldots; \tilde{\boldsymbol{X}}_n^T]$ be the $n \times k$ design matrix of the past experiment, $\Lambda_{\boldsymbol{X}} = [\boldsymbol{\lambda}_{\boldsymbol{X}}{}_1^T; \ldots; \boldsymbol{\lambda}_{\boldsymbol{X}}{}_n^T]$ be an $n \times k$ matrix to shift the past domain, and $\Omega_{\boldsymbol{X}} = diag\,[\boldsymbol{\omega}_{\boldsymbol{X}}]$ be a $k \times k$ diagonal matrix to rescale $\tilde{\mathcal{X}}$, where k is the dimension of the design space and the colon denotes separate rows of the matrix. Define $g(\mathcal{X}, \boldsymbol{\beta}) = [g(\boldsymbol{X}_1^T, \boldsymbol{\beta}); \ldots; g(\boldsymbol{X}_n^T, \boldsymbol{\beta})]$, an $n \times 1$ dimensional vector. The new "data"–the design matrix and response vector–is given by:

$$\begin{aligned} \mathcal{X} &= \tilde{\mathcal{X}}\Omega_{\boldsymbol{X}} + \boldsymbol{\Lambda}_{\boldsymbol{X}} \\ \boldsymbol{Y} &= \tilde{\boldsymbol{Y}} + (g(\mathcal{X}, \boldsymbol{\beta}) - g(\mathcal{X}, \tilde{\boldsymbol{\beta}})). \end{aligned}$$

By considering the design points as fixed, a response from the past experiment will be converted by adding the difference,

$$g(\boldsymbol{X}, \boldsymbol{\beta}) - g(\boldsymbol{X}, \tilde{\boldsymbol{\beta}})) \approx \bigtriangledown_{\boldsymbol{\beta}} g(\boldsymbol{X}, \tilde{\boldsymbol{\beta}})(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \frac{1}{2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})^T H_{\boldsymbol{\beta}}[g(\boldsymbol{X}, \tilde{\boldsymbol{\beta}})](\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}),$$

where we explicitly write a Taylor approximation to emphasize that the data response shift is coming from the difference in the conjectured and past model parameters. The Taylor approximation is about the point $\tilde{\boldsymbol{\beta}}$ and the gradient, $\bigtriangledown_{\boldsymbol{\beta}}$, and Hessian, $H_{\boldsymbol{\beta}}$, are taken with respect $\boldsymbol{\beta}$.

The parameter $\boldsymbol{\beta}$ and design points $\boldsymbol{X}_i$ are subject to misspecification. If the experimenter specifies the new parameter as $\boldsymbol{\mu}_\beta$, then the parameter is modeled as $\boldsymbol{\beta} = \boldsymbol{\mu}_\beta + \boldsymbol{\delta}_\beta$, where $\boldsymbol{\delta}_\beta \sim N(\boldsymbol{0}, \sigma_\beta^2)$; equivalently, $\boldsymbol{\beta} \sim N(\mu_{\boldsymbol{\beta}}, \sigma_\beta^2 \boldsymbol{\Sigma}_\beta)$. By the same argument, if an experimenter specifies the new design point as $\boldsymbol{\mu}_{\boldsymbol{X}_i} \equiv \boldsymbol{\omega}_{\boldsymbol{X}} \tilde{\boldsymbol{X}}_i + \boldsymbol{\lambda}_{\boldsymbol{X}_i}$,the design point is modeled as $\boldsymbol{X}_i = \boldsymbol{\mu}_{\boldsymbol{X}_i} + \boldsymbol{\delta}_{\boldsymbol{X}i}$, where $\boldsymbol{\delta}_{\boldsymbol{X}i} \sim N(\boldsymbol{0}, \sigma_{\boldsymbol{X}}^2 \boldsymbol{\Sigma}_{\boldsymbol{X}_i}), ;$ equivalently, $\boldsymbol{X}_i \sim N(\boldsymbol{\mu}_{\boldsymbol{X}_i}, \sigma_{\boldsymbol{X}}^2 \boldsymbol{\Sigma}_{\boldsymbol{X}_i})$. Lastly, when a model for the expected value of the responses

is conjectured, it is also subject to misspecification, denoted $\boldsymbol{\delta}_y$. Even if the mean response at a specified design point is conjectured to be $g(\boldsymbol{X}, \boldsymbol{\beta})$, the true process mean could conceivably behave like $g(\boldsymbol{X}, \boldsymbol{\beta}) + \boldsymbol{\delta}_y$. The addition of bias terms to the model adds flexibility to the model by not relying on the conjectured quantities as if they were truly known. The approach we use is similar to the elicitation approach in Reese *et al.* [44]. Similar to their approach, we assume the bias term, $\boldsymbol{\delta}_y$, is independent of $\boldsymbol{X}$ since no observed data is available to estimate the bias term as is done in model calibration/adjustment literature [54].

To model the above process, or more specifically, simulate predictions from the process by Gibbs sampling, we utilize the following hierarchical Bayes model:

$$\begin{aligned}
\mathbf{Y} &\sim N(g(\mathcal{X}; \boldsymbol{\beta}) + \boldsymbol{\delta}_y, \sigma_y^2 \boldsymbol{\Sigma}_Y) \\
\boldsymbol{\beta} \mid \sigma_y^2 &\sim N(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \sigma_\beta^2 \boldsymbol{\Sigma}_{\boldsymbol{\beta}}) \\
\sigma_y^2 &\sim IG(a, b) \\
\boldsymbol{X}_i &\overset{independent}{\sim} N(\boldsymbol{\mu}_{\boldsymbol{X}_i}, \sigma_{\boldsymbol{X}}^2 \boldsymbol{\Sigma}_{\boldsymbol{X}_i}) \\
\boldsymbol{\delta}_y &\sim N(\theta_{\delta_y}, \xi_{\delta_y}^2 \boldsymbol{I}),
\end{aligned} \tag{1}$$

where $IG(a, b)$ denotes the inverse gamma distribution with mean $b/(a-1)$, $\boldsymbol{I}$ is the identity matrix, and g is the translated version of the past response function f. Additional priors can be placed on the hyperparameters as desired. For clarity, consider the following example.

Example 2.1. Zong and Watkins [59] showed how the models for the film growth rate of a CVD process could be appropriately adjusted to capture the overall trend of the film growth rate for a SFD process, providing the following analytical, engineering based model (known as the *rate expression*) for the growth rate R :

$$R = \frac{C_{t\otimes} C_{t\oplus} k_4 \sqrt{K_1 K_2 K_3} \cdot [\mathrm{CuL}_{2(g)}]^{1/2} \cdot [\mathrm{H}_{2(g)}]^{1/2}}{2 \cdot \left[1 + \left(\sqrt{\frac{K_1}{K_2}} + \sqrt{K_1 K_2}\right) \cdot [\mathrm{CuL}_{2(g)}]^{1/2} + \frac{1}{K_5}[\mathrm{LH}_{(g)}]\right] \cdot \left[1 + \sqrt{K_3} \cdot [\mathrm{H}_{2(g)}]^{1/2}\right]},$$

where the C, k, and K values are parameters representing the active sites for adsorbed compounds, reaction constants, and equilibrium constants, respectively, and the compounds in brackets represent each compound's concentration. For example, $[CuL_{2(g)}]$ represents the concentration of $CuL_{2(g)}$. The parameters for the process are dependent upon the system under study and are often unknown and need to be estimated. For simplicity of presentation, we group the terms together, resulting in the following model to relate one input variable, x, either $[CuL_{2(g)}]$ or $[H_{2(g)}]$, to the growth rate:

$$R = \frac{\beta_0\sqrt{x}}{\beta_1 + \beta_2\sqrt{x}},$$

where we will restrict β_0 for model identifiability. Equivalently, dividing both the numerator and denominator by β_0 would have the same effect, but we keep the following format to better reflect the original model structure. In this setup, $[LH_{(g)}]$ is set at a nominal, fixed value. Suppose experimental data for a CVD process is available for studying the relationship between the growth rate and the $CuL_{2(g)}$ precursor concentration, while all other concentrations are fixed at nominal levels. Additionally, suppose a future SFD experiment is to be designed for the purpose of studying how the concentration of $H_{2(g)}$ relates to the growth rate while all other precursor levels are fixed at a nominal level.

Recall, $H_{2(g)}$ plays two roles in the SFD process: it serves as a reducing agent for metalization as well as to reduce copper oxides. Since both of these roles are not present in the CVD process, to be accurate, the growth rate model needs to be shifted in the experimental domain to account for the higher $H_{2(g)}$ concentration that is being used in the SFD process. In addition, the model parameters need to be updated to reflect the change from studying the effect of $[CuL_{2(g)}]$ versus $[H_{2(g)}]$. The process of building the conjectured model for studying the effects of $[H_{2(g)}]$ in the SFD system would play out as follows:

1. Based on the understanding of the double role that $H_{2(g)}$ plays in the SFD process, the needed amount of $H_{2(g)}$ concentration for the model to be accurate needs to increase. Therefore, a new experimental domain is specified by a shift to the right by a factor of λ_X so that each $\tilde{x} \in D_p$ is converted to $x = \tilde{x} + \lambda_x \in D_c$. Note, the actual point value of the shift may not be specified; an interval for the amount of shift in the domain may be specified instead. In this case, the midpoint will be taken as λ_X, and the width of the interval–the expressed uncertainty–can be used to specify σ_X^2 and $\mathbf{\Sigma}_X$.

2. Based on the difference in reaction rates when varying $CuL_{2(g)}$ versus $H_{2(g)}$, a new parameter value, $\boldsymbol{\beta}$, will be conjectured for the rate expression when studying the $H_{2(g)}$ concentration, thus giving $g(X, \boldsymbol{\beta})$. Typically, an exact value for the new parameter will be unknown, so an interval containing $\boldsymbol{\beta}$ will be given. In this case, the midpoint can be taken as the value $\boldsymbol{\beta}$, and the width of the interval can be used to specify σ_β^2 and $\mathbf{\Sigma}_\beta$.

3. The difference in the two model predictions is calculated on the comparable domain as $g(X, \boldsymbol{\beta}) - g(X, \tilde{\boldsymbol{\beta}})$.

4. The data from the prior experiment is adjusted based on the calculated difference in the two models; each response value is converted to a new response $y \equiv \tilde{y} + (g(X, \boldsymbol{\beta}) - g(X, \tilde{\boldsymbol{\beta}}))$.

5. The *pseudo-data* can then be modeled by Model (1). More importantly to our methodology, predicted values can be simulated.

A numerical application of the above process is provided in Section 6.

If the new process is well understood in relation to the old system it may be possible to use a more precise mapping function to map the previous experimental domain to the new one. In particular, a function $g : \xi_p \mapsto \xi_c$ could be specified

to derive the expected value of $\boldsymbol{X}_i$. Additionally, a new model form of the response could be used by adding new components that are found in the new system and not in the old system or by scaling the previous model form. For example, the new mean response model could be $\gamma(\boldsymbol{X}, \boldsymbol{\beta}) f(\boldsymbol{X}, \boldsymbol{\beta}) + h(\boldsymbol{X}, \boldsymbol{\beta})$.

2.3.2 Spatially Weighted Averaged Model

All sources of information, whether coming from expert opinion, engineering models, or past experiments, need to be combined into one unified model to drive the experimental design. The methodology to combine the models must address the following issues. First, different amounts of trust are put into each model, where the trust can vary with the design space. For example, an experimenter may be confident in a model at higher settings of the input variables but less confident in the model at lower values of the input variables. Second, models do not have to be defined over identical regions of the design space. For example, at high temperatures, chemical reactions tend to happen fast, where the limiting step is often the time it takes for molecules to diffuse together. At low temperatures, however, the reactions tend to be very slow, so the limiting step is the chemical reaction. At intermediate temperatures, both steps may take a similar amount of time, so to accurately predict the process, a more complex model is needed that includes both phenomena [21]. Lastly, the design points in each of the models need not be the same. A spatially-weighted averaged model–where model weights are dependent on the location in the design space–is proposed to deal with these issues.

Formally, suppose we have a total of M competing models. These M models can be a combination of many expert opinion models, multiple models derived from many past experiments, or numerous engineering models. See the appendix for possible methods of incorporating expert opinion and analytical, physics-based models. Let $u_j(\boldsymbol{x}_i^j)$ be the trust assigned to model j, $j = 1, \ldots, M$, at design point $\boldsymbol{x}_i^j$, $i = 1, \ldots, n_j$,

where n_j is the total number of conjectured design points associated with model j. Note, for engineering models not built from past experiments, no specific design points are conjectured. Instead, the model is relevant for a complete subset of the current design space, so trust values can be assigned for a sparse griding of the experimental region. To extend the "observed" trusts to the complete design space, inverse distance weighting is used. For models pertaining to the complete design space, the prediction of trust at an arbitrary design point $\boldsymbol{x} \in D_c$ is given by:

$$\hat{u}_j(\boldsymbol{x}) = \frac{\sum_{l=1}^{n_j}(d(\boldsymbol{x}, \boldsymbol{x}_l^j))^{-\kappa} u_j(\boldsymbol{x}_l^j)}{\sum_{l=1}^{n}(d(\boldsymbol{x}, \boldsymbol{x}_l^j))^{-\kappa}},$$

where d is any distance metric and κ is the *power parameter*. The choice of κ affects the influence of the closest observed response on predicting a new value. For larger values κ, the estimated function acts almost like a step function, while for smaller values, the estimated function varies more smoothly. A higher value of $\kappa = 6$ is used here, resulting in a step-like function of trust. By using a lower value of κ, the estimated function can vary too much in between observed trusts. As a result, even if the trusts of two neighboring design points are the same, in between the design points, the estimated trust could be higher or lower, which is not intuitive.

For local models–models that pertain to only a subset of the design space–extra care needs to be taken. Suppose model j^* is a local model defined on a subset S_{j*}. In addition to specifying a trust for each design point associated with model j^*, namely $\boldsymbol{x}_i^{j^*}$, $j = 1, \ldots, n_{j^*}$, a set of low or zero trust values needs to be assigned to design points outside of S_{j^*} in order to extend the trust model outside of S_{j^*}. These extra points can be taken from a set of design points of one of the models pertaining to the entire design space. With this extension, inverse distance weighting can be used to estimate the trust for any $\boldsymbol{x} \in D_c$.

With the predicted trusts for each model available over the entire design space,

the model weight for the j-th model at design point $\boldsymbol{x} \in D_c$ is defined as

$$w_j(\boldsymbol{x}) = \frac{\hat{u}_j(\boldsymbol{x})}{\sum_{j=1}^{M} \hat{u}_j(\boldsymbol{x})}.$$

The unified model is defined as

$$\boldsymbol{Y}(\boldsymbol{x}) = \sum_{j=1}^{M} w_j(\boldsymbol{x}) \boldsymbol{Y}_j(\boldsymbol{x}).$$

From the above model formulations, we can simulate predictions from the unified model to use in the experimental design.

2.4 Experimental Design

An objective-oriented initial experimental design must achieve two goals: effectively explore the entire design space while at the same time, not place too much emphasis on design regions where the process optima is not likely to exist. By exploiting the conjectured prior model, these two goals can be balanced by using a *minimum energy design* [17].

Minimum energy designs are based on the following principle. If n charged particles, all with the same sign of the charge, are placed into a box, then the particles will spread throughout the box before occupying a final position that minimizes the total potential energy. In the experimental design setting, the "box" is the design space and the particles are the n potential design points. The total potential energy of an experimental design with n points is defined as

$$E_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{q(\boldsymbol{x}_i) q(\boldsymbol{x}_j)}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)},$$

where $d(\cdot, \cdot)$ is the Euclidean distance and $q(\boldsymbol{x}_i)$ is the charge of the design point $\boldsymbol{x}_i$. Under this criteria, points with smaller charges are more likely to be included into the experimental design and are less likely to repel other points away from it.

A similar type of design known as *minimum potential designs* has been considered by Kessels *et al.* [31] and implemented in the statistical software JMP. However, minimum potential designs are restricted to creating space filling designs. Dasgupta's [17]

sequential minimum energy designs extended the idea behind these designs to allow the charges to vary with the location in order to reflect an optimality criteria. In particular, Dasgupta provided a purely sequential (one point at a time) experimental design procedure for global optimization, where the experimental designs sequentially search for a global optimum and explore the design space. Besides incorporating a different design objective, the main novelty of our design criteria is the direct incorporation of the conjectured nature of the prior model into the design stage. In addition, we provide details on extending the minimum energy design procedure provided by Dasgupta to allow a simultaneously selection of n design points.

2.4.1 Assigning Charges

The charges for each design point must be assigned to reflect the overall goal of the experiment. For example, if the goal is to maximize a response, then the energies will be a function of the response. Similarly, if the goal is to produce a product at a nominal value and to minimize costs, then the energies will be a function of both the response and the costs. We focus on experiments where the objective is to produce a product within a certain tolerance range of a target value. Therefore, charges are assigned to be a function of

$$p(\boldsymbol{x}) \equiv \mathcal{P}\left(T - d \leq Y^*(\boldsymbol{x}) \leq T + d\right),$$

where $Y^*(\boldsymbol{x})$ is the prior model prediction at design point $\boldsymbol{x}$, T is the nominal target value, and d is a tolerance requirement. Intuitively, $p(\boldsymbol{x})$ is a spatially weighted model average that our predicted response value will meet our engineering tolerance. Rajagopal and Castillo [43] consider a similar model averaged prediction probability for process optimization, yet our objective function extends their work in that we average models locally instead of globally, allowing more flexibility in combining statistical models.

Define the energy at $\boldsymbol{x}_i \in D_c$ as $q(\boldsymbol{x}_i) = (1 - f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i)))^\gamma$, where $\mathbf{u}(\boldsymbol{x}_i)$ is

a vector with the j-th position containing the amount of trust we put into the j-th model at design point $\boldsymbol{x}_i$, γ is a tuning parameter, and f is a function containing the optimality criteria. The value of $q(\boldsymbol{x}_i)$ plays an important role in the selection of the current design. Smaller values are more likely to be selected for the experiment, and once included in the design, they are less likely to repel other points away from it. Therefore, the goal is to choose $f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i))$ so that it assigns potential optima (after accounting for the trust) a lower value of q. Defining $q(\boldsymbol{x}_i)$ in such a way, as will be seen, allows for $f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i))$ to locally control the charges to reflect the prior model and the trust in the model and γ to control the overall model-based versus space-filling properties of our design. To choose the function f, the ideal properties that f should exhibit are considered. To be precise, the following definitions are introduced.

Definition 1. For each i, define $\tilde{p}(\boldsymbol{x}_i) = p(\boldsymbol{x}_i)/\max_j p(\boldsymbol{x}_j)$ to be the *rescaled* probability that our response surface prediction lies within the tolerance limits.

The purpose of rescaling is to spread the probabilities throughout the entire unit interval as to be able to distinguish them better than if they were clustered into a small region.

Definition 2. Let m be the median value of the unique set of all $\tilde{p}(\boldsymbol{x}_i)$. The experimental design is said to *encourage* the selection of design point $\boldsymbol{x}_i$ if $f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i)) > m$. Otherwise, it is said to *discourage* the selection of the design point $\boldsymbol{x}_i$.

Definition 3. Let $h^j(\boldsymbol{x}_i)$ be the distribution of the response surface at design point $\boldsymbol{x}_i$ under model j, $j = 1, 2, \ldots, M$, and denote the Hellinger distance between the two distributions given by models j and k at $\boldsymbol{x}_i$ by $H(h^j(\boldsymbol{x}_i), h^k(\boldsymbol{x}_i))$. Define $\tilde{u}(\boldsymbol{x}_i) =$

$\max_j u_j(\boldsymbol{x}_i) - \alpha \sum_{j<k} H(h^j(\boldsymbol{x}_i), h^k(\boldsymbol{x}_i))/\binom{M}{2}$, where α is a tuning parameter chosen to satisfy

$$0 \leq \alpha \leq \frac{\max_j u_j(\boldsymbol{x}_i)}{\sum_{j<k} H(h^j(\boldsymbol{x}_i), h^k(\boldsymbol{x}_i))/\binom{M}{2}}.$$

The purpose of $\tilde{u}(\boldsymbol{x}_i)$ is to reduce the amount of trust we put into the overall model if the individual models are in disagreement. Intuitively, the amount of trust put in the overall model is the maximum observed trust over all the individual models penalized by a measure of average discrepancy between the sub-models. The tuning parameter α will allow engineers to control the magnitude of the penalty given for disagreement between the individual models.

Lemma 2.1. For a design point $\boldsymbol{x}_i$, $0 \leq \tilde{u}(\boldsymbol{x}_i) \leq 1$.

Lemma 2.1 follows directly from the inequality restrictions in Definition 3 along with the fact that the Hellinger distance is always between 0 and 1. Lastly, we have the following definition:

Definition 4. We have *sufficient trust* in our model at the point $\boldsymbol{x}_i$ if $\tilde{u}(\boldsymbol{x}_i) > 0.5$. Otherwise, we say we have *insufficient trust* in our model at the point $\boldsymbol{x}_i$.

Keeping in mind that minimum energy designs favor points with smaller charges, f should be a function satisfying the following properties:

(P1) f should be bounded between 0 and 1 as to keep the charges positive;

(P2) if $\tilde{p}(\boldsymbol{x})$ is greater (less) than m, $f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i))$ should *encourage* (*discourage*) the selection of $\boldsymbol{x}_i$ as long as there is *sufficient trust* in our prior model at $\boldsymbol{x}_i$;

and

(P3) as $\tilde{u}(\boldsymbol{x}_i)$ increases to 1, $f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i))$ should reflect the information given by the prior model. As $\tilde{u}(\boldsymbol{x}_i)$ decreases to 0, $f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i))$ should serve to normalize the charge to m, the reference value.

Property (P1) is to ensure that the charges remain between 0 and 1 for simplicity and to ensure that all charges have the same sign. The purpose of property (P2) is to select design points that will potentially lead to the optimum process condition and to not select point that are likely to not produce the optimum process condition. However, the design criteria will only promote this behavior if we have sufficient trust in our model. If we do not have sufficient trust in our model, then property (P3) says that the charge should be normalized as to neither encourage nor discourage selection directly. Instead, these points will be selected in the experimental design depending on the amount of space-filling properties that are desired.

Proposition 2.1. Let m be the median of the set of all unique values of $\tilde{p}(\boldsymbol{x}_i)$. The function

$$f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i)) = \tilde{p}(\boldsymbol{x}_i) + \operatorname{sgn}\left(m - \tilde{p}(\boldsymbol{x}_i)\right)\left(m - m\tilde{u}(\boldsymbol{x}_i)\right)\left|\frac{\tilde{p}(\boldsymbol{x})}{m} - 1\right| \tag{2}$$

satisfies properties (P1)-(P3).

Proof. The proof of Proposition 2.1 can be found in the appendix. □

Equation (2) is not the unique function that satisfies (P1)-(P3). This function is chosen because it applies a simple, linear penalty to the re-scaled probabilities. Other definitions of f could be derived to incorporate, for example, a quadratic penalty which would penalize lower trust values more heavily. The intuition behind equation (2) is as follows. At first, $\tilde{p}(\boldsymbol{x}_i)$ is considered and then readjusted based on

the amount of trust given at $\boldsymbol{x}_i$. As the trust level decreases, the second half of the sum in equation (2) readjusts $\tilde{p}(\boldsymbol{x}_i)$ by linearly "pushing" it closer to m, the reference value. For example, if $\tilde{u}(\boldsymbol{x}_i) = 1$, when the model is completely trusted, equation (2) equals $\tilde{p}(\boldsymbol{x}_i)$. On the other hand, if $\tilde{u}(\boldsymbol{x}_i) = 0$, no trust in put in the model, equation (2) always equals m. As a result of Proposition 2.1, the charges can be defined as $q(\boldsymbol{x}_i) = (1 - f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i)))^{\gamma}$, where $f(p(\boldsymbol{x}_i), \mathbf{u}(\boldsymbol{x}_i))$ is given by equation (2). Consequently, the experimental design targets regions where the predicted response is most likely to fall into a pre-specified tolerance interval, yet explores the design space if an insufficient amount of trust is present.

2.4.1.1 The Choice of γ

The value of γ plays an important role in the design process. It is incorporated into the design metric so that researchers have the freedom to balance space-filling properties with the optimality criteria. When γ is set to 0, all the energies are equal, essentially resulting in a maximin space-filling design [17]. By setting γ to a higher value, smaller energies shrink to zero faster than larger ones, resulting in completely object oriented design. Illustrations of the effect of changing γ are presented in Section 2.5. The magnitude of γ is likely to change on a problem-to-problem basis, so a universal magnitude to set its value at may not be appropriate. However, some discussion on selecting γ is provided.

When deciding which γ value to choose, researchers can view different experimental designs for different values of γ, and based on the results, the "best" γ can be chosen in an ad-hoc fashion. For one and two dimensional design spaces, the potential designs can be viewed in their entirety. In higher dimensions, all the combinations of the input variables can be plotted against each other. The objective is to find the design that spreads out design points to reflect the overall goals and prior knowledge of the experimenter. If a researcher is confident in the prior knowledge, then a higher

γ value can be chosen. For example, if the current system is very similar to a previous system or if an engineering model is posed based on well known physical phenomena, then researches can feel more comfortable with a higher γ value. Alternatively, if the current system is only relatively related to past systems, or if engineering models have been proposed based on many simplifying assumptions, a lower level of γ can be chosen to balance the two design objectives.

An additional concern when choosing γ is to make sure that the design is feasible and practical. For example, a high γ value could put two points so close together that they result in essentially the same experimental run. For example, if the temperature of a system can only be controlled within ± 5 degrees centigrade of the actual temperature, then two experimental runs that are less than 10 degrees centigrade apart are not distinguishable. In such a scenario, γ should be chosen under the constraint that no two points are too close together. Similarly, a high value of γ may result in two points that are distinguishable with regards to the actual experiment but that may still be neighboring points. As a result, the experiments conducted at these two points could potentially provide the same information. Note, this "distinguishability" of potential design points should be addressed when selecting the grid (e.g., orthogonal array [52]) for potential design points.

The advantage of using a higher value of γ is that the optimal system operating condition could be found by using less resources, provided that the prior model is relatively accurate. However, if the prior model is inaccurate, a high value of γ could lead to a design that places too many points in the wrong region, requiring additional resources for follow-up designs. The advantage of using a small value of γ is that the design points will fill the design space like a traditional space-filling design so that the process can be characterized as a whole. However, if the prior knowledge is reasonably accurate, the experimenter would have missed out on an opportunity to find the optimal system setting with fewer resources. This balancing act needs to be

considered on a case-by-base basis.

2.4.2 Experimental Design Search Region

The optimal experimental designs discussed here are found by scaling all input factors to the unit interval, thus transforming the experimental region into a unit hypercube. In addition, further constraints will be placed on the experimental region to encourage better space-filling properties. The following constraints may be too restrictive for physical experiments where the experimental design is chosen, for example, from an orthogonal array [52] where factors are not assigned a large number of levels, so caution should be exercised when applying these constraints on an ad hoc basis. In scenarios when design factors can have a large number of levels, often occurring in computer experiments, these constraints are potentially useful.

Minimum energy designs are related to minimum potential and maximin designs, where both encourage the minimum distance between any two design points to be, in some sense, maximal. This is one (of many) desirable property for space-filling designs, yet under smaller sample sizes, points are forced to the boundary of the design region, which is hardly space-filling at all. Consequently, minimum potential designs are restricted to the unit hypersphere to remedy this issue. Similarly, the search region for the experimental designs here is first restricted to the unit hypersphere and then adjusted to ensure that potentially optimal design regions have not been removed from the search region. The search region is found as follows:

1. Grid the unit hypercube into $g \times g$ squares;

2. Restrict the set of points in step 1 to the unit hypersphere, resulting in the removal of 2^d disjoint subsets, where d is the number of design factors. Denote these subsets by R_i, $i = 1, \ldots, 2^d$;

3. For each R_i, find the number of points in R_i that satisfy $f(\tilde{p}(\boldsymbol{x}_i), \tilde{u}(\boldsymbol{x}_i)) > m$. Denote the set of points in R_i satisfying this property by S_i;

4. If S_i contains multiple, non-neighboring points of the boundary of the hypersphere, then include R_i back into the search region.

This choice of the search region is hardly restrictive: it can very quickly and easily be found and is in agreement with the desired properties of the design. In particular, design points where $f(\tilde{p}(\boldsymbol{x}_i), \tilde{u}(\boldsymbol{x}_i)) > m$ are to be encouraged for selection. Regions with points satisfying this requirement are added back to the search region if they were removed when restricting the unit hypercube to the unit hypersphere. In regions not containing design points satisfying this property, space-filling properties are to be emphasized when resources are available. The restriction to the hypersphere forces design points off the boundary of the unit hypercube. As a result, the whole region can be characterized by moving points off the boundaries. Lastly, note that by setting $\gamma = 0$, the goal is to produce a completely space-filling design. The search space should always be restricted to the hypersphere in this case. Similarly, if zero trust is given to a region near a corner of the design space, the hypersphere restriction can be used in that corner.

2.4.3 Exchange Algorithm

An exchange algorithm [19] is used to find the design the minimizes the total energy. For the current design region, D_c, the goal is to find n points, $x_1, \ldots, x_n \in D_c$ to minimize

$$E_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\hat{q}(\boldsymbol{x}_i)\hat{q}(\boldsymbol{x}_j))}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)}.$$

It should be understood that here, D_c only consists of the points in the search region as discussed in the previous section. The general idea behind exchange algorithms is that a proposed solution is updated by exchanging a current design point with a potential design point, given that the potential design point (in this case) reduces the total energy. An exchange algorithm for minimizing the total potential energy works as follows:

1. Denote the current design region by D_c. Randomly select n points from D_c, denoting these points by the vector $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)$;

2. Set $i = 1$;

3. Calculate the current total energy for the vector $\boldsymbol{d}$, and denote its value by E^c;

4. For each point x_j in the design space $D_c \backslash \boldsymbol{d}$, replace d_i with x_j and calculate $E_{i,j}$, the total energy when d_i is replaced by x_j. Let $E_i^* = \min_j E_{i,j}$ and x^* be the corresponding design point to E_i^*. If $E_i^* < E^c$, exchange d_i with x^*, i.e., the i-th element of $\boldsymbol{d}$ is replaced with x^*. Set $i = i + 1$;

5. Repeat steps 3 and 4 until $i = n + 1$, at which point, go to step 6;

6. Repeat steps 2 through 5 until no more exchanges are made or the maximum allowable iterations is achieved. The final value of $\boldsymbol{d}$ is used as the initial design.

Step 4 can efficiently be calculated by noting that when replacing d_i with candidate point x_j, $E_{i,j}$ can be decomposed as

$$E_{i,j} = E^c - \sum_{k=1,k\neq i}^{n} \frac{\hat{q}(d_i)\hat{q}(d_k))}{d(d_i, d_k)} + \sum_{k=1,k\neq i}^{n} \frac{\hat{q}(x_j)\hat{q}(d_k))}{d(x_j, d_k)}.$$

Therefore, the total potential energy of the proposed design doesn't have to be completely re-calculated at each step, but instead, can just be updated by removing the energies corresponding to d_i and adding the energies corresponding to x_j. Since exchange algorithms are dependent on the starting design, it is recommended to run the algorithm multiple times with different starting points. The design from these repetitions with the minimum energy should be selected.

2.4.4 Modifications to Address Poor Prior Information

The experimental design is predicated on the fact that reasonably accurate prior information is available to build an initial model, however, this model cannot be

validated until data is present. It is therefore important to be able to make the design plan more robust in cases where the prior information is potentially inaccurate. This goal can be achieved by reducing the number of design points invested into the minimum energy design plan (discussed thus far), followed by the incorporation of additional points that satisfy some space-filling criteria. The idea is to fill in the "gaps" of the optimization-based design with a space-filling design. In particular, to create an initial design with n design points, the modified design works as follows:

1. Choose $n_1 < n$ points based on the optimal design criteria discussed thus far. Denote this design by ξ_1;

2. Choose the remaining $n_2 = n - n_1$ design points from $D_c \backslash \xi_1$ to minimize the space-filling criteria:

$$\left(\sum_{\boldsymbol{x}_i, \boldsymbol{x}_j \in D_c \backslash \xi_1} \frac{1}{d^m(\boldsymbol{x}_i, \boldsymbol{x}_j)} + \sum_{\boldsymbol{x}_i \in \xi_1, \boldsymbol{x}_j \in D_c \backslash \xi_1} \frac{1}{d^m(\boldsymbol{x}_i, \boldsymbol{x}_j)} \right)^{1/m},$$

   where d^m is a distance metric. For example, for $m = 1$ and $m = 2$, the distance corresponds to rectangular and Euclidean distances, respectively.

The metric in step 2 is a common metric to measure space-filling properties [45] and was used by Ba [3] for similar purposes. Notice when $m = 2$, the metric in step 2 is equivalent to the minimum energy design criteria when all the charges are set to 1. By splitting the design into these two steps, part of the resources are explicitly dedicated to the optimality criteria while the remaining resources are explicitly dedicated to space-filling properties. As a result the initial design can still be efficient while at the same time, have an "insurance policy".

2.5 Numerical Properties

We illustrate numerical properties of our procedure on one and two-dimensional simulated datasets. In particular, the effects that γ and $\tilde{u}(\boldsymbol{x})$ have on the space-filling properties are explored. Additionally, the role of sample size is examined.

2.5.1 One-dimensional Example with Known Model

Consider a model in the unit interval given by

$$g(x) = 10 + 20(x - 0.5)^3 - x + \epsilon,$$

where $\epsilon \sim N(0, 1)$. For illustration, suppose the goal is to find the optimal input to produce a product within the interval [9.5, 10.5], i.e, $T = 10$ and $d = 0.5$, noting that the optimal input occurs at approximately $x = 0.85$. For this example, the response surface is considered as known, $n = 6$ and the design space is gridded into intervals of length 0.001. Figure 1 shows how the experimental designs change when changing γ and $\tilde{u}(\boldsymbol{x})$. When studying the properties of γ, $\tilde{u}(\boldsymbol{x}) = 0.75$ for the entire design space, and when studying $\tilde{u}(\boldsymbol{x})$, we consider constant levels of trust for the whole region while γ is fixed at 1. As γ increases, the space-filling properties decrease, resulting in more points being placed near the optimal region. Similarly, the space-filling properties of the design decrease as $\tilde{u}(\boldsymbol{x})$ increases (and γ is held constant). As the model is trusted more, the design points are placed to target the region in which the optimal value is expected.

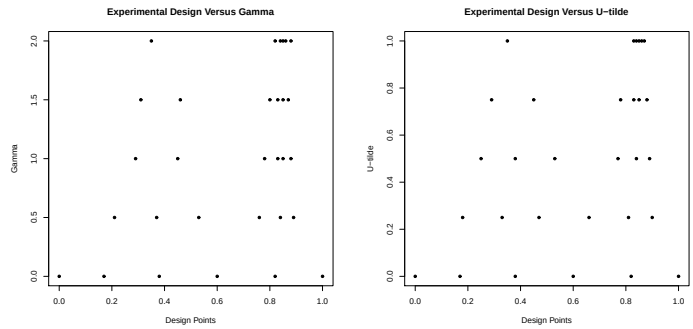

Figure 1: This figure shows *(left)* the wffect that changing γ has on the experimental design and *(right)* the effect that the amount of trust has on the experimental design (holding γ constant).

The trust is a flexible measure that does not have to (and typically won't) remain constant over the entire design region. To illustrate the effect of varying trust over

the design region, consider the same scenario as above, but where the level of trust is split at the point $x = 0.5$. In Figure 2, different designs are shown where $\tilde{u}(\boldsymbol{x})$ equals 0 for $x < 0.5$ and varies (as labeled) for $x \geq 0.5$. In the center and right plots, γ equals 0.5 and 1, respectively. For the lower input values, the design focuses on space-filling properties, while when $x \geq 0.5$, the design places more emphasis on the optimality criteria as the trust increases. As a result, the points cluster around the true optimal input with high level of trust. When $\gamma = 1$, the points tend to cluster faster than when $\gamma = 0.5$, resulting on less emphasis on the space-filling properties.

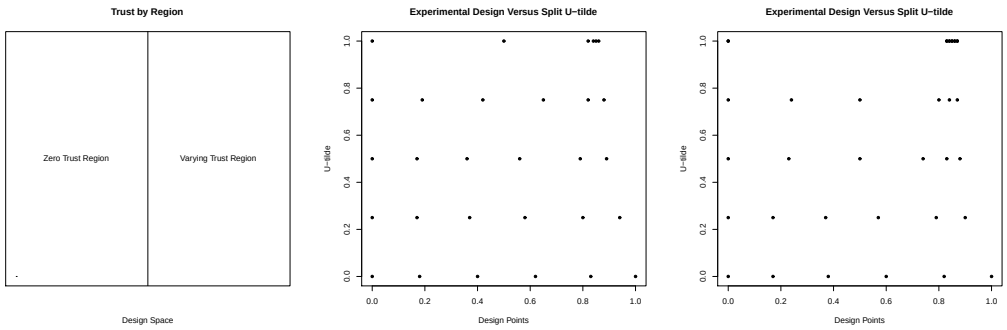

Figure 2: This figure shows the effect of splitting the trust model at the point $x = 0.5$. The left plot displays how the trust is split. For the center plot, $\gamma = 0.5$, and for the right plot, $\gamma = 1$.

Alternatively, suppose for lower input values ($x < 0.5$) the trust varies and for higher input values ($x \geq 0.5$) no trust is put in the model. Figure 3 depicts the results. In the center and right plots, γ equals 0.5 and 1, respectively. Since the true optimal input is greater than 0.5, but no trust is given to the model in that region, only space-filling properties are emphasized. Note, the function g never equals 10 when $x < 0.5$. As a result, the design points only converge slowly to the best input for $x < 0.5$.

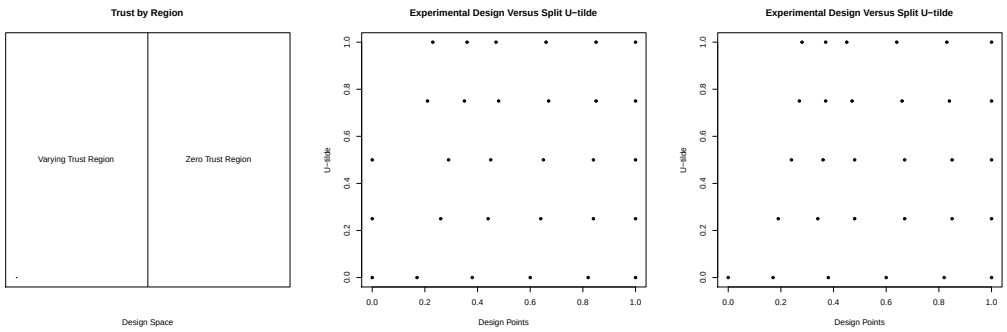

Figure 3: This figure shows the effect of splitting the trust model at the point $x = 0.5$. The left plot displays how the trust is split over the design region. For the center plot, $\gamma = 0.5$, and for the right plot, $\gamma = 1$.

2.5.2 Two-dimensional Example with Known Model

Consider the Branin function given defined by:

$$g^*(x_1^*, x_2^*) = \left(x_2^* - \frac{5(x_1^*)^2}{4\pi^2} + \frac{5x_1^*}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1^*) + 10,$$

where $-5 \leq x_1^* \leq 10$ and $0 \leq x_2^* \leq 15$. Suppose the true system behavior is defined by the Branin function rescaled to the unit square and modified resulting in

$$g(x_1, x_2) = \frac{400 + g^*(x_1, x_2)}{20} + \epsilon,$$

where $x_1 = 15x_1^* - 5$, $x_2 = 15x_2^*$, and $\epsilon \sim N(0, 1)$. For illustration, suppose the goal is to find the optimal input to produce a product within the interval [19, 21], i.e, $T = 20$ and $d = 1$, and note that the optimal input occurs at approximately $(x_1, x_2) = (0.55, 0.15)$. Figure 4 shows the effect of changing γ on the proposed minimum energy design. In particular, the space-filling properties decrease as γ increases. For this example, the response surface is considered as known, the design space is gridded into 0.05×0.05 squares, $n = 20$ and $\tilde{u}(\boldsymbol{x}) = 0.75$. Considering the same setup, the effect of different values of trust are studied with a fixed value of γ. As seen in Figure 5, as the trust increases, the optimality criteria is emphasized.

Figures 6 and 7 display the effects of when the levels of trust are not equal over the entire design space. In each figure, the assignment of trust levels is specified in

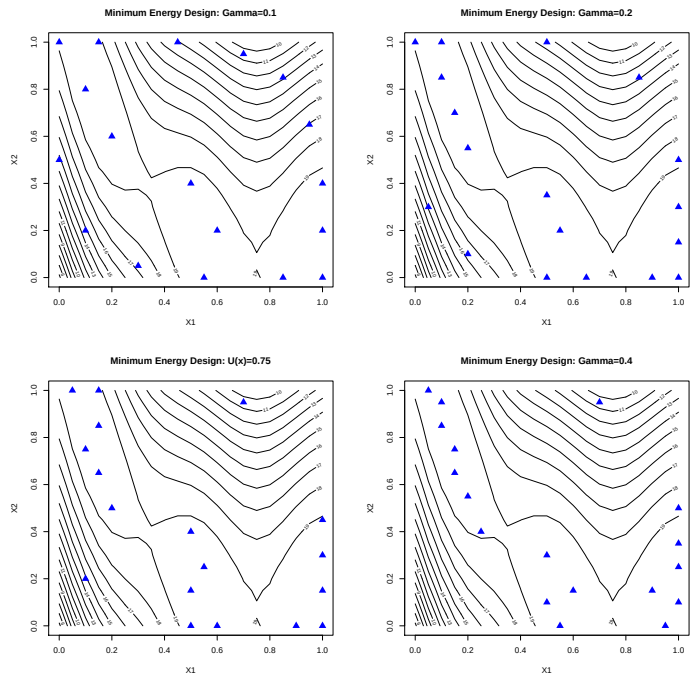

Figure 4: Experimental designs for varying values of γ overlaying the true response surface are shown. For the designs shown here, $\tilde{u}(\boldsymbol{x}) = 0.75$ for all $\boldsymbol{x}$.

the upper left plot, and the effects of changing γ are shown in the remaining three plots. In Figure 6, the lower two quadrants have a higher level of trust than the upper two. Since the lower two quadrants (especially, the lower right quadrant) contain near optimal system inputs and because a higher level of trust is given to these regions, more design points migrate to these regions as γ increases. Alternatively, in Figure 7, the upper two regions are assigned higher trust levels. Since the upper left quadrant contains a near optimal input and a higher level of trust, points migrate to that region as γ increases, lending fewer design points to the lower two quadrants.

2.5.3 Role of Sample Size

The sample size plays an important role in designing experiments and has been addressed (for computer experiments) by Loeppky, Sacks, and Welch [36] who provide the following rule of thumb: 10 runs per dimension. Up until this point, samples

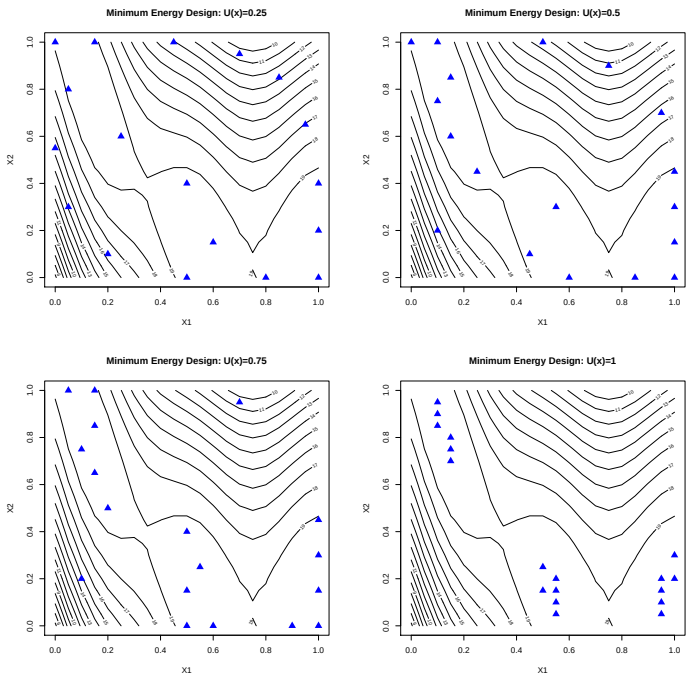

Figure 5: Experimental designs for varying levels of trust overlaying the true response surface are shown. For the designs shown here, $\gamma = 0.4$.

of size 18 have been used for the two-dimensional examples. Since the sample size can play an important role in the effectiveness of exploring the design space, smaller sample sizes are considered here. Figure 8 shows experimental designs using 8 design points with varying values of γ and $\tilde{u}(\boldsymbol{x})$ (as labeled on the plots). The main deficiency in using a small sample size occurs when low values for both γ and $\tilde{u}(\boldsymbol{x})$ are present. In such a scenario, the goal is to emphasize the space-filling properties of the design, but with a small sample size, all the points are pushed towards the boundary. A traditional space-filling design (e.g., minimax, Latin hypercube, etc. [45]) would be more appropriate in this case. The use of a larger γ in conjunction with a larger value of trust results in placing few or no points in potentially non-optimal region. However, this effect also presents itself in larger samples and by the choice of γ, can be subdued in larger sample sizes when enough resources are available to divide amongst the two design objectives.

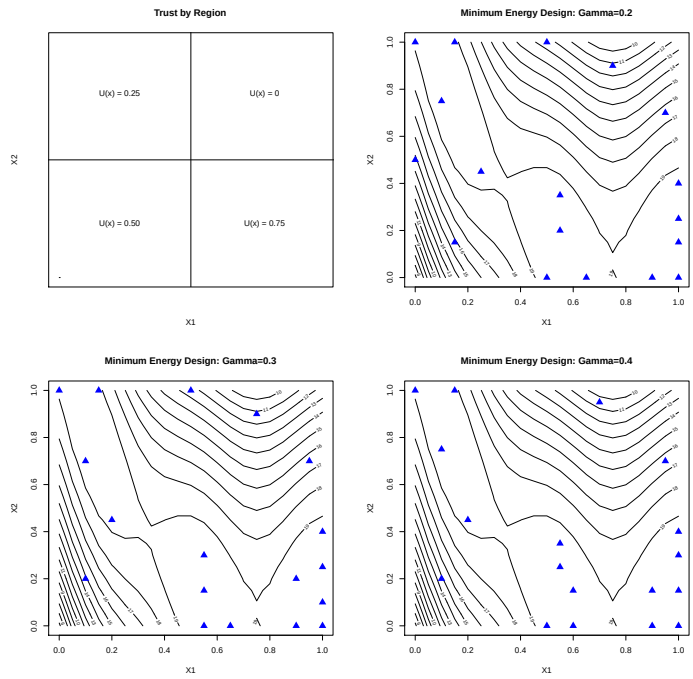

Figure 6: This figure displays the results of splitting the trust over the design space as shown in the top left plot. The remaining plots show the designs for differing values of γ.

Figure 9 shows four designs with varying levels of trust and values of γ (as labeled on the plots) for samples of size 12. By increasing the sample size by four points (12 points total), the space-filling properties are satisfied much better for both low-trust cases. For the higher trust cases, the designs only focus on the optimality criteria. By tuning γ to a lower value, it is likely that a balance can be struck between the two design objectives (as seen in the upper left plot in Figure 9).

2.6 Application to Thin Film Deposition Process

Zong and Watkins studied the thin film growth rate from a supercritical fluid deposition (SFD) process, and related this rate to that of the much better studied chemical vapor deposition (CVD) process. In their SFD process, a substrate and precursor are loaded into a closed reactor, followed by the introduction of supercritical carbon

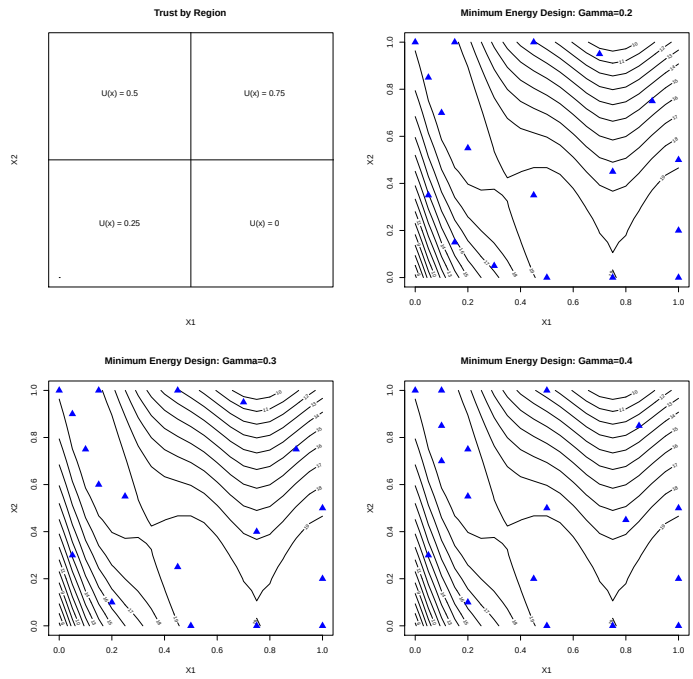

Figure 7: This figure displays the results of splitting the trust over the design space as shown in the top left plot. The remaining plots show the designs for differing values of γ.

dioxide. After the system conditions are maintained for one hour, ensuring the dissolution of the precursor into the carbon dioxide, hydrogen was loaded into the reactor. The reactor was then heated to 250 $^{\circ}C$ and allowed to react for five minutes. One of several quantities of interest, the film growth rate, was measured. As discussed in Example 2.1, the better understood kinetics of a CVD process was used to propose a growth rate model for the SFD process and was shown to be approximately accurate under some circumstances, while in others, adjustments needed to be made based on known differing aspects between the CVD and SFD processes. This link could allow previously collected data on a CVD system to guide future experimentation for the SFD system. For our application, however, we consider a simpler case based on the data provided in Zong and Watkins [59].

Consider the "past" experiment to be the experiment conducted in Zong and

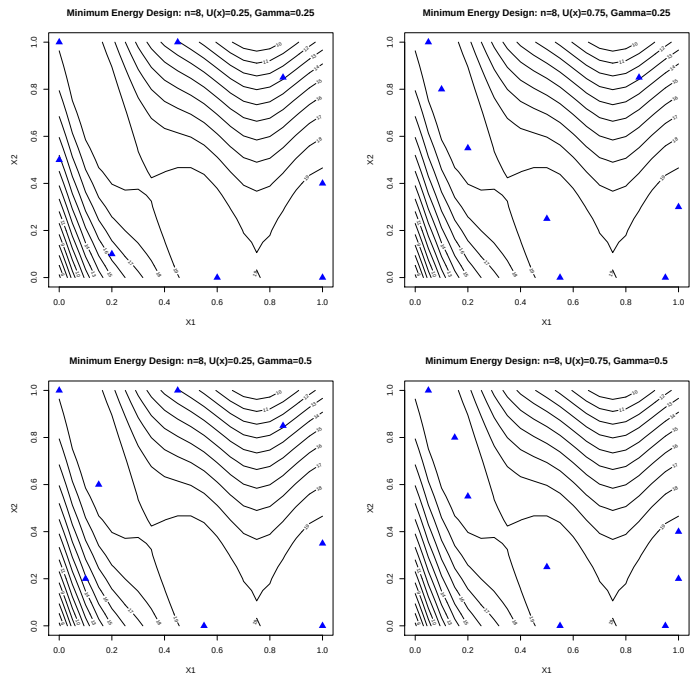

Figure 8: This figure shows the effect that sample size has on design space exploration. Samples of size 8 are considered.

Watkins [59] which studies the effects of the $\mathrm{CuL}_{2(g)}$ concentration on the thin film growth rate, R, in a SFD system. An analytical model was proposed and shown to approximate the behavior of the growth rate versus $\mathrm{CuL}_{2(g)}$ concentration. In particular, the following model is used:

$$R = \frac{C_{t\otimes} C_{t\oplus} k_4 \sqrt{K_1 K_2 K_3} \cdot [\mathrm{CuL}_{2(g)}]^{1/2} \cdot [\mathrm{H}_{2(g)}]^{1/2}}{2 \cdot \left[1 + \left(\sqrt{\frac{K_1}{K_2}} + \sqrt{K_1 K_2}\right) \cdot [\mathrm{CuL}_{2(g)}]^{1/2} + \frac{1}{K_5}[\mathrm{LH}_{(g)}]\right] \cdot \left[1 + \sqrt{K_3} \cdot [\mathrm{H}_{2(g)}]^{1/2}\right]}, \quad (3)$$

where the C, k, and K values are parameters for reaction rates and the compounds in the brackets, for example, $[\mathrm{CuL}_{2(g)}]$, represent the compound's concentration. In this example, the concentration of $\mathrm{LH}_{(g)}$ is set a nominal, fixed value. For ease of illustration, we group terms together, resulting in the following nonlinear model to relate one input variable x, either $[\mathrm{CuL}_{2(g)}]$ or $[\mathrm{H}_{2(g)}]$, to the growth rate:

$$R = \frac{\beta_0 \sqrt{x}}{\beta_1 + \beta_2 \sqrt{x}}. \quad (4)$$

For model identifiability and to keep the same model structure, we fix $\beta_0 = 100$ for

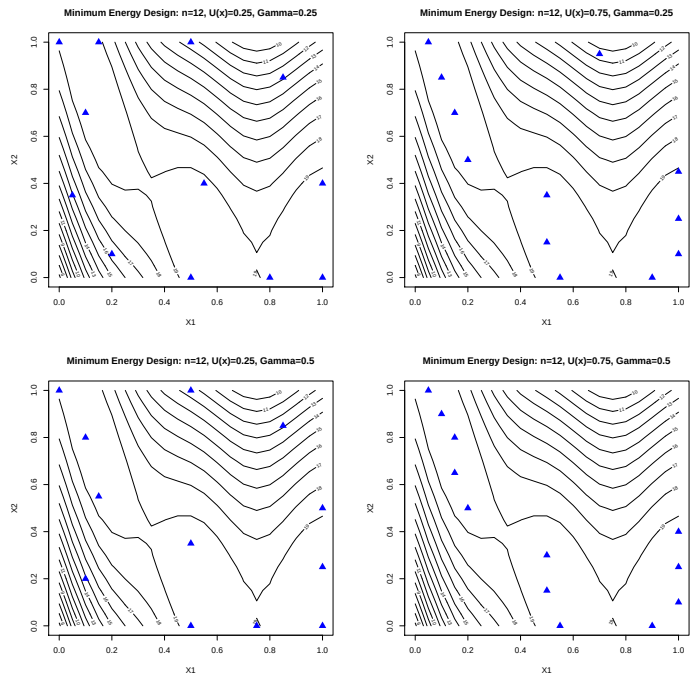

Figure 9: This figure shows the effect that sample size on design space exploration. Samples of size 12 are considered.

illustrative purposes. Alternatively, the numerator and denominator can be divided by β_0 to reduce the number of parameters. In reference to the knowledge of this past experiment, suppose we want to design a future experiment on the SFD system to study the relationship between the concentration of $H_2(g)$ and the growth rate, where it is known that Model (3) may be accurate after some needed adjustments. Using our methodology, the data for the past experiment–when studying $[CuL_{2(g)}]$–can be linked to the future experiment–when studying $[H_{2(g)}]$–through updating the parameters and design space of Model (3). The data (as given in Zong and Watkins) for the past experiment studying the effect of the $CuL_2(g)$ concentration is shown in Table 1. Figure 10 shows the plotted data along with the nonlinear least squares estimated model–the parameter values being $\tilde{\beta}_1 = 0.1968$ and $\tilde{\beta}_2 = 3.2844$.

The goal of the new experiment to be conducted is to relate the $H_{2(g)}$ concentration to the growth rate, R. In particular, suppose a target value for growth rate is

Table 1: Data relating the $CuL_2(g)$ concentration to R is shown. Data is taken from Zong and Watkins [59].

$[CuL_2(g)]$ (wt.%)	R (nm/min)
0.010	17.429
0.012	18.722
0.015	21.157
0.022	23.663
0.057	25.997
0.110	27.175
0.121	23.208
0.171	26.215

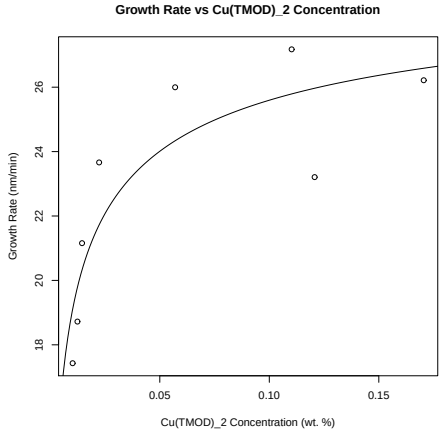

Figure 10: The past experimental data on studying the effect $[CuL_{2(g)}]$ on R is shown along with the estimated model. Data points are from Zong and Watkins [59].

set at $T = 21$ nm/min with a tolerance of $d = 0.5$ nm/min. Model (3) was based on the kinetics of a similar system, a CVD process, yet it is known that $H_{2(g)}$ plays an additional role in the SFD process which doesn't exist in the CVD process. So although it is believed that the functional form will be similar, it can be conjectured a priori that the model's domain needs to be changed to reflect this difference. Additionally, new parameter values will have to be conjectured based on differing reaction rates that will occur when the concentration of $H_{2(g)}$ is varied as opposed to that of $CuL_2(g)$. In particular, suppose an experimenter conjectures that the double role the

Table 2: Conjectured data relating the $H_{2(g)}$ concentration to R are listed.

$[H_{2(g)}]$ (wt.%)	R (nm/min)	$u(x)$
0.0352	9.5882	0.75
0.0374	11.1111	0.75
0.0395	13.7833	0.8
0.0473	17.1949	0.8
0.0820	22.7729	0.8
0.1352	27.1316	0.8
0.1457	23.6480	0.4
0.1955	28.5641	0.4

$H_{2(g)}$ plays in the reaction will cause the previous model to be shifted by a factor between [0.02, 0.03], i.e., $\lambda_{X_i} = .025$ for all i. Based on a postulated difference in the reaction rates that is to be expected when studying $[H_{2(g)}]$ instead of $[CuL_{2(g)}]$, suppose it is conjectured that $\beta_1 \in [.65, .80]$ and $\beta_2 \in [1.5, 1.9]$, i.e. $\mu_{\beta_1} = 0.725$ and $\mu_{\beta_2} = 1.7$. Note, we present the conjectured parameters in terms of intervals, which is often the case in which conjectures arise due to experts's uncertainties. The intervals also give additional information about how we may set the variance parameters in Model (1) by interpreting the endpoints as the values giving the inner 95% area of a normal probability distribution. The augmented (postulated) data is shown in Table 2 along with the specified levels of trust at each design point. With the conjectured data at hand, the parameters in Model (1) can be simulated, in turn, allowing for predicted values to be drawn. Table 3 lists the parameter values used in Model (1). The estimated model parameters are $\beta_1 = 0.8$ and $\beta_2 = 1.8$.

To more completely illustrate our methodology, suppose, in addition to the past data, expert opinion data was given as in Table 4. The data is in a similar format as to that in Reese *et al.* [44], from which we will use their technique to model such data. See Reese *et al.* for details and notations. From this data, the following linear model is selected as the best model: $y_i^* = 0.8148\ x_i^* + 0.1033\ (x_i^*)^4 + \epsilon_i$, where $\epsilon_i \sim N(0, 0.3443)$ and y^* and x^* are the normalized values of y and x, respectively.

Table 3: Parameter values used for Model (1) are listed.

Parameter	Value
$\boldsymbol{\mu_\beta}$	(0.725, 1.700)
$\boldsymbol{\sigma_\beta}^2$	(0.0014, 0.0104)
$\boldsymbol{\Sigma_\beta}$	$\boldsymbol{I}_{2\times 2}$
$\boldsymbol{\mu_{X_i}}$	$\tilde{\boldsymbol{X}} + (0.025)\mathbb{1}_{8\times 1}$
$\sigma^2_{\boldsymbol{X}}$	0.003
a	.0001
b	.0001
θ_{δ_y}	0
$\xi^2_{\delta_y}$	1

Table 4: Sample expert opinion data are shown.

$x_{o,1}$	y_o	q_ξ	u_o
.025	20.5	21	.5
.05	21	21.5	.8
.075	20	21	.8
.1	23	24	.8
.125	23.5	24.5	.65
.15	24	25.5	.5
.175	26	27.5	.4
.2	.5	30.5	.4

Using the two models–the past experimental and expert opinion models–we calculate $\tilde{u}(\boldsymbol{x}_i)$ by calculating the Hellinger distance by using Gaussian approximations to the posterior predictive distributions and using $\alpha = 0.1$. Figure 11 shows the estimated mean models based on the past experimental data, expert opinion data, and the unified model. In addition, the "unknown" data relating $[H_{2(g)}]$ to the growth rate–also given in Zong and Watkins [59]–is plotted.

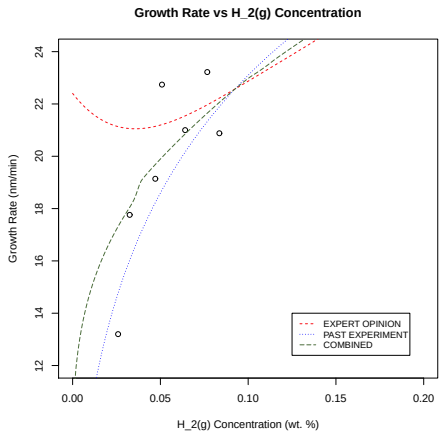

Figure 11: Past experiment, expert opinion, and unified models (along with the "unknown" data) for describing the growth rate versus $H_{2(g)}$ concentration are shown. Data points come from Zong and Watkins [59].

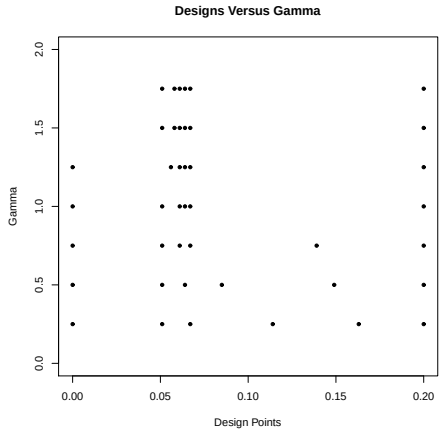

Figure 12: A plot of experimental designs for different γ values is shown.

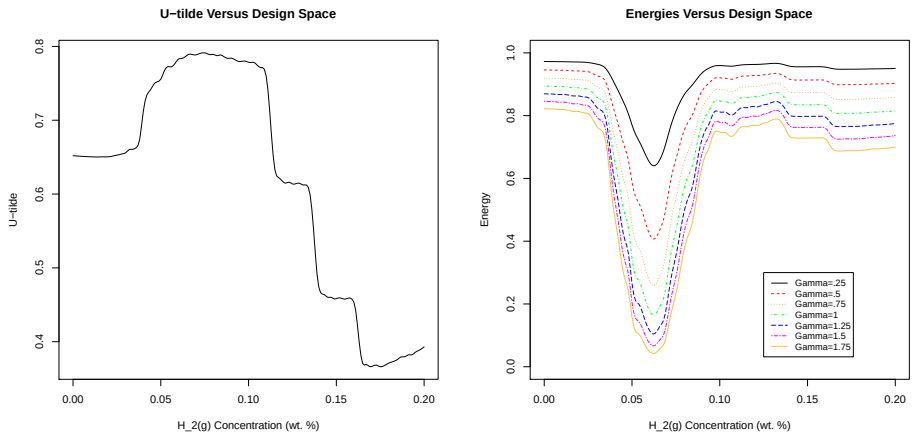

Figure 13: Smoothed plots of $\tilde{u}(\boldsymbol{x})$ and of the energy versus $H_{2(g)}$ concentration are shown for varying values of γ.

The experimental designs for different levels of γ are shown in Figure 12. It is clear that the designs focus more on the optimality criteria as γ increases, which is seen by the clustering of points near 0.6–the location that the model predicts a response at the target value. To understand how the energies and adjusted trusts affect the experimental design, consider the plots of the energy and $\tilde{u}(\boldsymbol{x}_i)$ versus the input variable, $[\mathrm{H}_{2(g)}]$, shown in Figure 13. From the plot of $\tilde{u}(\boldsymbol{x})$, we see that there is a low level of trust for higher values of $[\mathrm{H}_{2(g)}]$. In terms of the effect that this property has on the design, we see that the experimental design–especially for lower values of γ–tends to focus on space-filling properties for the higher input range (where the level of trust is low). Additionally, the plot of energy versus the design space, with varying values of γ, displays distinct regions with relatively low energy values, resulting from the combination of both high trust and a "good" value of the unified model. On the other hand, even with sufficient trust in the unified model for low values of $[\mathrm{H}_{2(g)}]$, the energy values for this region are high due to the model predictions being far from the target value. As a result, the experimental design, no matter the value of γ, places at most one design point in that region. Notice that as γ increases, regions with lower energy values are more pronounced, illustrating how γ plays a role in

globally controlling the design. Selecting which value of γ to use is the decision of the experimenter and can be chosen in accordance with the discussion we provided.

For comparative purposes and to emphasize the importance of using good prior information, we consider the experimental designs when the past experimental data is not converted to reflect the behavior of the current system. All other aspects of the modeling, e.g., expert opinion and trust models, are kept the same as before. We still consider the past experiment to be the one when studying the effects of $[\mathrm{CuL}_2(g)]$ on the growth rate of the film and the experiment to be conducted to be the one when studying the effects of $[\mathrm{H}_{2(g)}]$ on the growth rate of the film. Figure 14 show the initial designs versus varying γ. Recall that the optimal value, based on the limited data provided in Zong and Watkins [59], is around 0.06. In this case, as γ increases, the design points cluster more and more around 0.35, the wrong optimal value. As seen, for higher values of γ, not spending time to carefully conjecture a good prior model can lead to bad consequences. The designs with low levels of γ are more along the lines of traditional space-filling designs which are typically used when there is no consideration to build a prior model.

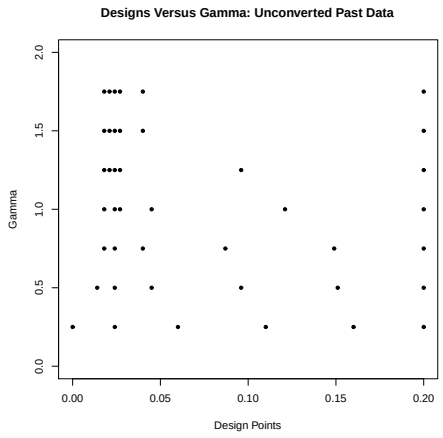

Figure 14: A plot of experimental designs for different γ values are shown for the case when the past experimental data is not converted.

2.7 Conclusion

In this work, a framework for combining data from multiple sources of information for designing an initial experiment for new and advanced technology is provided. Such initial experimental design plans can be used as stand alone designs or as an initial experiment to jump-start a sequential design scheme. Differing from previous works, we provide a methodology to incorporate data coming from previous experiments on non-identical systems and discuss real scenarios in a nano-manufacturing context to illustrate the applicability and potential benefit of our methodology. Additionally, we provide a new flexible method for local model averaging by incorporating a spatially varying trust model, allowing for the emphasis and deemphasis of each model in regions where the model is conjectured to be accurate or inaccurate, respectively. Lastly, we provide a new minimum energy design criteria to incorporate both a response surface and a trust model. Through simulations, we show that our design criteria can lead to efficient initial experiments by targeting regions that possibly contain a potential optimal value and by placing fewer design points in regions where an optimal value is less likely to occur. The balance between the space-filling and optimal design criteria are dependent upon the trust model (at a local level) and a tuning parameter (at the global level), giving experimenters the freedom to control the trade-off of these properties. Concluding, we illustrate our methodology for designing an initial experiment where the end goal of experimentation is to find the optimal growth rate in a supercritical fluid thin film deposition process. The experimental designs change based on the trust model and global tuning parameter, and, moreover, when reasonably accurate updates to past experimental data are possible, our methodology for modeling past experimental data on non-identical systems can give better results than ordinary space-filling designs. Consequently, in the presence of expert knowledge, our initial design plan potentially saves resources in the experimental design process.

CHAPTER III

MODEL ROBUST PROCESS OPTIMIZATION THROUGH BATCH SEQUENTIAL DESIGNS: A LOCAL APPROACH

3.1 Introduction

When studying new, advanced technologies, new statistical challenges can arise, resulting in the need for new methodology to accurately and efficiently address these challenges. Nanotechnology is no exception. As nanoscience is becoming important in both scientific and non-scientific communities alike, new statistical challenges are being introduced to accommodate the needs of this emerging field. These challenges include, yet are not limited to, large unknown design regions and highly non-linear response surfaces with multiple local optima and non-convex failure regions [18]. These attributes present challenges in both the design and analysis of such systems. In order to expand nanoprocesses to large scale production, new statistical advances to accommodate these statistical challenges need to be developed. For more on statistical methods in nanoscience, see Lu *et al.* [37].

The experimental design and analysis challenges introduced in nanotechnology are not adequately addressed by existing design literature. Fractional factorial designs and orthogonal arrays [52] can require a prohibited number of runs which prove too costly for process optimization. Space filling designs [45], although allowing for flexible modeling, can also require too many runs in addition to wasting too many resources on no yield regions. Casciato *et al.* [10] and Kim [32] addressed designing experiments in a nanomanufacturing scenario for modeling both the response and failure region of a system. Additionally, Kim [32] and Casciato *et al.* [9] provided a flexible batch sequential methodology, termed *Layers-of-Experiments* (LOE), for

optimizing a nanoparticle deposition process. Woo [51] provided further study on the LOE procedure in addition to provided a Gaussian process modeling approach where the design region contains a failure region. Other advancements include using statistically designed experiments to optimize the growth quality of ZnO nanowires [55] and new methodology in modeling the synthesis of nanostructures [18]. In addition, Dasgupta [17] and Joseph, Dasgupta, and Wu [27] provided a sequential design procedure, termed *sequential minimum energy designs* (SMED), for optimizing the yield in a nanowire synthesis system characterized by a large, no yield regions along with a highly, non-linear response surface with multiple local optima.

A good design procedure may not be enough to efficiently optimize complex systems as found in nanoscience. Alternatively, a flexible, adaptive modeling procedure, along with a sequential design, can provide a more powerful and efficient process of study. Under limited resources, however, a purely data-driven statistical approach may not be enough for efficiently optimizing a complex, high-cost process, so it is important that engineering knowledge can easily be incorporated into the modeling and design methodologies. To this end, we provided a complete and efficient batch sequential experimental design and analysis procedure that easily allows the incorporation of expert knowledge in the form of engineering models, expert opinion, and past experimental data on similar, non-identical systems (as discussed in Vastola *et al.* [46]).

Two new approaches in incorporating engineering knowledge into the modeling process are introduced: local model calibration and local model averaging. By taking a local approach to modeling, both statistical models and engineering information can be combined, even in the scenario that each piece of information only pertains to a subregion of the complete design region. In addition, a new batch sequential experimental design procedure, motivated by the principles of the *Layers-of-Experiments* approach in Kim [32], is provided in which both an adaptive design region and combined-design

criteria are introduced.

3.2 Literature Review

The methodology we propose is a complete process optimization scheme that includes new methodologies in both experimental design and statistical modeling. We review the relevant literatures separately below.

3.2.1 Review of Sequential Designs for Optimization

A good review of sequential designs for optimization in the nanotechnology setting is given in Dasgupta and Joseph, Dasgupta, and Wu. In particular, arguments for the necessity for designs to be able to deal with systems containing non-convex failure regions and highly complex response surfaces are given. We briefly review the literature here and refer readers to Dasgupta [17] and Joseph, Dasgupta, and Wu [27] for further details.

Arguably, the most popular methods for sequential optimization include *probability of improvement* (PI) and *expected improvement* (EI) criteria [26]. Both methods rely on first using a space-filling design to estimate a kriging model [45] and then based on this model, choose the next design point to maximize the PI or EI. Dasgupta [17] and Joseph, Dasgupta, and Wu [27] point out that although PI and EI algorithms have some nice theoretical properties, they may suffer in nanotechnology research due to two reasons. First, the initial kriging model requires a moderate amount of initial experimental runs to accurately represent the response surface, many of which would be placed in no-yield regions. By incorporating engineering models, our proposed methodology can efficiently estimate an initial model under a sparse allocation of resources to the initial design. Second, they point out that the accuracy of the kriging model can suffer from the amount of no-yield (or non-response) regions, as a result, misleading the follow-up experimental designs.

Other types of sequential experimental design procedures for optimization use an

adaptive design region. Response surface methodology (RSM) [8] is a traditional approach using a gradient based method to search for the optimal value. As pointed out in Dasgupta *et al.* [18], under highly complex surfaces it is unlikely that the simple modeling structures used in RSM will be able to capture the true response behavior. Other modern gradient based methods are the STRONG [14] and STRONG-X [15] procedures which follow the *trust region* approach in simulation optimization literature. These approaches, however, require far too many runs for space exploration and optimization in high-cost, low-resource experiments. Other procedures (SEL [53], SELC [39], and G-SELC [38]) focus on eliminating regions (or points) from the design space, yet, as discussed in Dasgupta [17] and Joseph, Dasgupta, and Wu [27], these methods require too many experimental runs in a high-cost, low-resource scenario. Another approach is the Grid Algorithm of Wissmann and Grover [49, 50] in which the experimental domain is updated based on the most probable model amongst competing models. Reducing the design space resulted in more efficient designs, however, their proposed design criteria was sensitive to model misspecification, and their method didn't account for the uncertainty in model selection (e.g., through model averaging).

Motivated by the Grid Algorithm of Wissmann and Grover [49, 50], a recent approach that uses an adaptive region and adaptive design criteria is the Layers-of-Experiments (LOE) methodology provided in Kim [32]. The key principles in the LOE approach are:

1. sequentially update the design region in a flexible manner so that it can contract or expand to include potential optimal regions;

2. sequentially update a compound design criteria to emphasize exploration versus the optimality criteria; and

3. sequentially build local models on updated design regions.

Although an effective procedure, no methodology for directly incorporating multiple sources of information is provided, the adaptive design criteria relies on D-optimality, and the models are restricted to be linear, or at least, linearly adjusted (when considering engineering models). The methodology presented here, based on the LOE principles, provides a more flexible LOE procedure that directly incorporates multiple, competing sources of engineering knowledge.

3.2.2 Review of Model Averaging and Model Calibration

Model averaging is a technique that has frequently been used in order to obtain "better" models by taking into account the uncertainties in model selection. For a more complete review, refer to Hoeting *et al.* [25]. Traditionally, model averaging is applied in a global setting, in which a model's "usefulness" is assessed over the entire input domain [25, 43]. Local approaches in model averaging have recently been considered [33, 41], where in the former approach, model averages are taken point-by-point in a spatial setting and in the latter approach, data is binned into regions where model averaging is done in each bin. Considering model averaging on a point-by-point basis only allows for weights to be estimated where data are observed, and to estimate weights at a given design point, only the data at the specified design point are used. The binning procedure requires lots of data and is useful in a non-parametric setting, however, many physical and computer experiments are expensive or time consuming to run and, thus, only limited data are available. Differing from these works, we propose a kernel based method to calculate model weights (posterior model probabilities) in the presence of competing statistical, computer, and/or analytical engineering models. The advantage of our approach is two-fold: the local strengths of competing models are emphasized in relation to other models and model weights can be directly estimated over the entire domain with a limited amount of data.

Model calibration and adjustment is a popular field in computer experiments

where computer models are adjusted in reference to physical data that has been collected. Most work stems from the seminal paper by Kennedy and O'Hagan [30], in which a flexible Gaussian process model is used to describe the discrepancy between the computer model and the physical data. Similar methodologies can be found in Bayarri *et al.* [6], Qian and Wu [42], Wang, Chen, and Tsui [48], Chang [13], and many others. A review can be found in Xiong *et al.* [54]. Joseph and Melkote [29] argue that for physical experiments, the "curvature" of the Gaussian model may be inappropriate for calibrating engineering models to physical data. As a result, they provide an adjustment to engineering models by using multiple regression. Both methods assume engineering or computer models are appropriate for the whole design space. However, it is not always the case that a model is appropriate for the entire experimental domain. For example, in a chemical vapor deposition (CVD) process, at high temperatures a process may be described by a *diffusion limited* model, while at low temperature the process may be described by a *reaction limited* model [21]. In many instances, it is not clear where one model stops and another starts. In another scenario, a metamodel may be developed based on experimentation from only a subregion of the entire current experimental region. After calibration, the metamodel could be useful on the subregion it was developed for, in addition to regions lying just outside the subregion, yet it is unlikely to be valid for the whole design space. When locally calibrating the model, only data in the appropriate subregion and areas lying just outside the subregion should be included. The direct application of the calibration procedures above can result in poor model calibration in the above scenarios. Alternatively, we provide a very general approach to locally calibrate models by assigning importance to each data point for the calibration process.

3.3 Modeling Methodology

The first step in our methodology is to combine all sources of information to build a unified model to describe the experimental process. In particular, all models–local, global, detailed, approximate, computer-based metamodels, and physics-based analytical models–are locally adjusted to the physical data before creating a unified global model through local averaging. We emphasize a local approach since we do not require that all models be global models. Instead, a model can describe the process on a subset of the entire design space. For example, engineers may be able to postulate an analytical model under restricted experimental settings or previous experimentation may been conducted on a fraction of the complete current experimental space, yielding a statistical model that is relevant to only a part of the complete design space. In these scenarios, it makes sense to adjust a model on its relevant domain–the domain for which the model was built. After adjustments are made, local averaging is used to unify all the models.

3.3.1 Local Model Adjustment

Before experimentation on the current system, many times engineers can provide analytic or computer models based on the physical understanding of the process. After an initial experiment is conducted, the validity of these models can then be evaluated. Departures of the collected data from the proposed model can potentially lead engineers to update the model in attempt to explain the model mismatch. The goal of this updating procedure is to create a dialog between statisticians and engineers. This interaction process could lead to engineers updating models based on information that the initial experiment provided instead of relying on a purely statistical approach to adjusting models. In turn, the updated engineering models have all the advantages of engineering models, in particular, the power to extrapolate the model to more accurately guide experimentation with limited data. There is no guarantee that

the new model will match the process or that an updated engineering model could be proposed at all. Therefore, after interacting with engineers, a statistical calibration should be performed if the updated engineering model is still deemed inadequate. As models may only pertain to subsets of the complete experimental domain, we propose using a local model adjustment approach.

To better motivate the use of locally calibrating models, consider the following scenario. A CVD process can be *reaction-limited* at low temperatures, where the reaction tends to be very slow. At high temperatures, when chemical reactions tend to happen quickly, the same process can be *diffusion-limited* [21]. When it comes time to calibrate each model, we do not want to bias the calibrations by incorporating irrelevant data; we don't want to use data at low temperatures to calibrate the diffusion-limited model (or *vice versa*) since the model is not expected to explain the low temperature phenomenon. We therefore propose to calibrate the models by implicitly relying on weighted least squares, where higher weights are given to the relevant data.

The adjustment/calibration weights can be determined with direct interaction with experts in the field. Take for example the following weighting scheme. For model j, $j = 1, \ldots, m$, let $\mathcal{B}^j = \left[\boldsymbol{b}_1^j, \boldsymbol{b}_2^j\right]$ be the hypercube in the experimental domain where it is known that model j is appropriate, and let $\mathcal{A}^j = \left[\boldsymbol{a}_1^j, \boldsymbol{a}_2^j\right]$ be the hypercube in the domain where it is known that model j is not suspected to be useful for $(\mathcal{A}^j)^c$, the complement of $\mathcal{A}^j$. Then, the data in region $\mathcal{B}^j$ can be weighted equally with (non-normalized) weights equal to 1, the data in region $(\mathcal{A}^j)^c$ can be weighted 0, and the data in region $\mathcal{A}^j \backslash \mathcal{B}^j$ can be weighted such that the (non-normalized) weights decrease from 1 to 0 as the distance from the set $\mathcal{B}^j$ increases. For example, in the

one dimensional case, a linear weighting function for model j could look like:

$$v_i^j(x_i) = \begin{cases} \frac{b_1^j - x_i}{a_1^j - b_1^j} + 1, & \text{if } x_i \in \left[a_1^j, b_1^j\right) \\ 1, & \text{if } x_i \in \left[b_1^j, b_2^j\right] \\ \frac{b_2^j - x_i}{a_2^j - b_2^j} + 1, & \text{if } x_i \in \left(b_2^j, a_2^j\right] \\ 0, & \text{otherwise} \end{cases}.$$

For example, let the complete design space be the unit interval, and consider a model, denoted $\mathcal{M}_j$, that is proposed for the region $[\frac{3}{4}, 1]$. Furthermore, suppose it is known that the model is not appropriate for the region $[0, \frac{1}{2}]$. Then, $\mathcal{B}^j = [\frac{3}{4}, 1]$, $(\mathcal{A}^j)^c = [0, \frac{1}{2}]$, and $\mathcal{A}^j \backslash \mathcal{B}^j = (\frac{1}{2}, \frac{3}{4})$. When adjusting/calibrating model $\mathcal{M}_j$, the available data would be given (unnormalized) weights as follows: $v^j(x_i) = 1$ if $x_i \in [\frac{3}{4}, 1]$; $v_i^j(x_i) = \frac{3/4 - x_i}{1/2 - 3/4} + 1$ if $x_i \in [\frac{1}{2}, \frac{3}{4}]$; and $v_i^j(x_i) = 0$ if $x_i \in [0, \frac{1}{2}]$.

With the weights specified, models can be locally calibrated. We consider the constant and functional-adjustment modeling approaches discussed in Joseph and Melkote [29], but in our local context. As a result, extra care needs to be taken in estimating the mean and prediction variances. For ease of illustration, we first discuss the constant-adjustment model where the engineering model is an analytical model, i.e., only a mean specification of the process. Afterward, we will consider the functional-adjustment model as well as approximated engineering models such as metamodels arising from computer simulations.

The constant-adjustment model for model j is given by

$$Y(\boldsymbol{x}) - f_j(\boldsymbol{x}) = \beta_0 + \beta_1(f_j(\boldsymbol{x}) - \bar{f}_j) + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$, $\boldsymbol{x}$ is the input value, $Y(\boldsymbol{x})$ is the observed response at $\boldsymbol{x}$, $f_j(\boldsymbol{x})$ is the expected value of engineering model j at $\boldsymbol{x}$, $\bar{f}_j$ is the mean of engineering model j, and β_0 and β_1 are unknown adjustment parameters. In the local context, estimates

of β_0 and β_1 can be found using weighted least squares, i.e., by minimizing

$$\sum_{i=1}^{n} v_i^j(\boldsymbol{x}_i)\left[(Y(\boldsymbol{x}_i) - f_j(\boldsymbol{x}_i)) - (\beta_0 + \beta_1(f_j(\boldsymbol{x}_i) - \bar{f}_j))\right]^2$$

with respect to β_0 and β_1. Using matrix notation, the weighted sum of squares can be expressed as

$$\left((\boldsymbol{Y} - \boldsymbol{f}_j) - \boldsymbol{F}\boldsymbol{\beta}\right)^T \boldsymbol{V}\left((\boldsymbol{Y} - \boldsymbol{f}_j) - \boldsymbol{F}\boldsymbol{\beta}\right), \tag{5}$$

where $\boldsymbol{Y}$ and $\boldsymbol{f}_j$ are vectors containing all the $Y(\boldsymbol{x})$ and $f_j(\boldsymbol{x})$ values, respectively, $\boldsymbol{V}$ is a diagonal matrix of weights, $\boldsymbol{\beta}$ is the vector of regression (adjustment) parameters, and $\boldsymbol{F}$ is the design matrix with a vector of ones in the first column and the vector $\boldsymbol{f}_j - \bar{\boldsymbol{f}}_j$ in the second column, where $\bar{\boldsymbol{f}}_j$ is a vector with all elements equal to $\bar{f}_j$.

A numerical solution to (5) can be found, but the nature of the weight matrix $\boldsymbol{V}$ leads to complications in statistical inference. Although, (5) is the same optimization problem as the typical formulation for weighted least squares, the weights have no relation to the variances of the observations [40]. In particular, $\boldsymbol{V}$ may have zeros for diagonal elements which can make the matrix $\boldsymbol{X}^T\boldsymbol{V}\boldsymbol{X}$ singular. This weighted least squares approach has been considered by Lindgren, Alberg, andn Domkin [34] for deterministic model adjustment. The weights in their case corresponded to the quality of the observed data. However, as discussed in Xiong *et al.* [54], deterministic model adjustment cannot account for the randomness in the data. We remedy this situation by not using weights explicitly, but instead, we conduct inference using implied weights from using a weighted bootstrapping procedure. As a result, standard calibration techniques can be applied after weighting each data point based on its location.

The bootstrap method is a versatile resampling procedure which can be used to estimate the parameters of a model. The power of using a weighted bootstrap allows some observations to be downplayed in the analysis [4] and is used here by assigning low, or even zero, weights to observations that are irrelevant to the model

that is being calibrated. Existing analysis procedures are applied to the weighted bootstrap samples, resulting in a simple way to locally calibrate models based on existing methodologies. Local adjustments for a single analytical model using the constant adjustment model proceeds as follows:

1. For data point i, $i = 1, \ldots, N$, define sampling weights, $v_i(x)$ according to expert knowledge. Normalize the weights to sum to 1, i.e. $\sum_{i=1}^{N} \tilde{v}_i(x) = 1$, where $\tilde{v}_i(x)$, $i = 1, \ldots, N$, denote the normalized weights.

2. Let n, $n \leq N$, be the number of non-zero sampling weights. Draw a sample with replacement of size n from the discrete distribution

$$\begin{pmatrix} (x_1, y_1) & (x_2, y_2) & \ldots & (x_N, y_N) \\ \tilde{v}_1(x) & \tilde{v}_2(x) & \ldots & \tilde{v}_N(x) \end{pmatrix},$$

   giving resampled data $D^* = \{(x_1^*, y_1^*), (x_2^*, y_2^*), \ldots, (x_n^*, y_n^*)\}$.

3. Using the resampled data, (x_i^*, y_i^*), $i = 1, \ldots, n$, estimate the constant-adjustment model parameters using the procedures discussed in Joseph and Melkote [29] (or any other model calibration procedure). In addition, estimated values such as the model variance and prediction interval widths can be stored for use in later inference.

4. Repeat Step 3 M times. For a parameter η, M bootstrap estimates exist, namely η_i^*, $i = 1, \ldots, M$, one estimate for each bootstrap sample.

5. For parameter η, the weighted bootstrap estimate is just $1/M \sum_{i=1}^{M} \eta_i^*$.

For the process optimization problem considered here, an important quantity is the 95% prediction interval for an individual response. Let $\hat{\mu}^C$ denote the mean of the constant adjustment model, noting that $\hat{\mu}^C(\boldsymbol{x}) = f(\boldsymbol{x}) + \hat{\beta}_0 + \hat{\beta}_1(f(\boldsymbol{x}) - \bar{f})$. When the

engineering model is an analytical model, the approximate 95% prediction interval for a new response at design point $\boldsymbol{x}$ is

$$\hat{\mu}^C(\boldsymbol{x}) \pm z_{\alpha/2}\sigma \left\{ 1 + \frac{1}{n} + \frac{(f(\boldsymbol{x}) - \bar{f})^2}{S} \right\}^{1/2},$$

where $S = \sum_{i=1}^{n}(f_i - \bar{f})^2$ (see [29]). To calculate the prediction interval using the bootstrapped samples, in Step 3 the value of the half-length of the above prediction interval, in addition to the β estimates, needs to be stored.

If the engineering model is an estimated model, such as a metamodel arising from a computer experiment, extra care needs to be taken to incorporate the variance of the estimated model. In this case, the constant-adjustment predictor is given by $\hat{\mu}^C(\boldsymbol{x}) = \hat{f}(\boldsymbol{x}) + \hat{\beta}_0 + \hat{\beta}_1(\hat{f}(\boldsymbol{x}) - \bar{f})$, noting that f is replaced by it approximation $\hat{f}$. Denote the estimated variance of the metamodel at point $\boldsymbol{x}$ by $\nu(\boldsymbol{x})$. As shown in Joseph and Melkote [29], the approximate 95% prediction interval is given by

$$\hat{\mu}^C(\boldsymbol{x}) \pm z_{\alpha/2}\sigma \left\{ 1 + \frac{1}{n} + \frac{(f(\boldsymbol{x}) - \bar{f})^2 + \nu(\boldsymbol{x})}{S} + (1 + \hat{\beta}_1)^2\nu(\boldsymbol{x}) \right\}^{1/2}.$$

This interval can be calculated in the same manner as in the analytical model case by using the weighted bootstrapping procedure.

Example 3.1. To provide a simple illustration of the local model calibration procedure, consider a true response surface defined by the function

$$f(x) = \begin{cases} \frac{1}{2}\sin(\pi x), & \text{if } x \leq 6; \\ \frac{\pi}{2}(x - 6), & \text{if } x > 6. \end{cases}$$

Suppose two engineering models are conjectured, one to describe the process behavior for lower inputs in the design space and the other for describing the behavior for higher inputs in the design space. Specifically, suppose the first model (for lower inputs) is conjectured as $f_1(x) = (1/2)\sin(\pi x) + .3$ and the second model (for higher inputs) is conjectured as $f_2(x) = ((\pi + 1)/2)(x - 6) - 1/2$. The plot on the left in Figure 15

shows the two conjectured models along with simulated data from the true response function. Engineering model 1 captures the overall shape of the data in the lower input region, yet overestimates the observed data. Similarly, engineering model 2 seems to capture the overall shape of the data in the higher input region, yet it is rotated, resulting in underestimating much of the data. The plot on the right in Figure 15

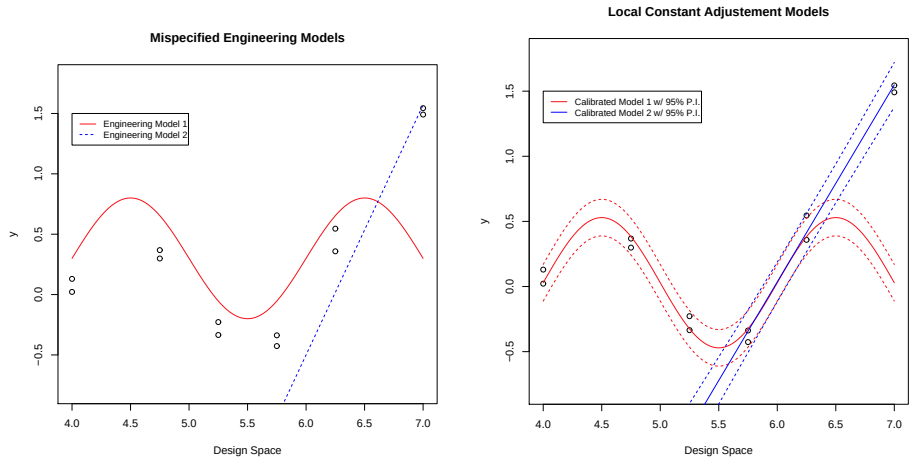

Figure 15: Example of local model calibration: (*left*) engineering models with observed data and (*right*) locally calibrated engineering models with 95% prediction intervals are shown.

shows the locally calibrated engineering models. For calibrating engineering model 1, the data inputs which are less than 5.5 are assigned (non-normalized) weights of 1, while data with inputs greater than 6.5 are assigned (non-normalized) weights of 0. Data with inputs at 5.75 and 6.25 are assigned (non-normalized) weights of .75 and .25, respectively. For calibrating engineering model 2, data with inputs less than 5.5 are assigned (non-normalized) weights of 0, while data with inputs greater than 6 are assigned (non-normalized) weights of 1. Data with inputs at 5.75 are assigned (non-normalized) weights of .75. Using the proposed weighted bootstrapping method, the constant adjustment model parameters were estimated along with the approximate 95% prediction intervals of a new response value. Note that due to the larger bootstrapped variances, the prediction intervals are wider in regions where

the estimated response has more curvature. In Example 3.2, we will continue this example to illustrate the local averaging procedure discussed next. □

If a constant adjustment model is deemed inadequate (see Joseph and Melkote [29]), then functional-adjustment models can be used to further calibrate the engineering model. The functional-adjustment model is defined as

$$Y - \mu^C(\boldsymbol{x}) = \delta(\boldsymbol{x}; \boldsymbol{\alpha}) + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$, $\delta(\boldsymbol{x}; \boldsymbol{\alpha})$ is a function to explain the discrepancy between the data and the constant adjustment model, and $\boldsymbol{\alpha}$ is a vector of unknown parameters associated with δ. For this work, δ is restricted to the set of linear predictors, namely

$$\delta(\boldsymbol{x}; \boldsymbol{\alpha}) = \alpha_0 + \sum_{i=1}^{p} \alpha_i u_i(\boldsymbol{x}),$$

where the $u_i(\boldsymbol{x})$ are known functions of $\boldsymbol{x}$. The estimated expected value of the functional adjustment predictor is given by

$$\hat{\mu}^F(\boldsymbol{x}) = \hat{\mu}^C(\boldsymbol{x}) + \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i u_i(\boldsymbol{x}),$$

where the $\hat{\alpha}_i$ are the least squares estimates of the α_i's. The 95% prediction interval is complicated due to the two-stage estimation process, however it can be derived (as shown in Joseph and Melkote [29]) as

$$\hat{\mu}^F(\boldsymbol{x}) \pm z_{\alpha/2}\sigma \left\{ 1 + \frac{1}{n} + \frac{(f(\boldsymbol{x}) - \bar{f})^2}{S} + \boldsymbol{u}(\boldsymbol{x})^T[(\boldsymbol{U}^T\boldsymbol{U})^{-1} - \boldsymbol{A}]\boldsymbol{u}(\boldsymbol{x}) \right\}^{1/2},$$

where $\boldsymbol{A} = (\boldsymbol{U}^T\boldsymbol{U})^{-1}\boldsymbol{U}^T\boldsymbol{F}(\boldsymbol{F}^T\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{U}(\boldsymbol{U}^T\boldsymbol{U})^{-1}$, $\boldsymbol{U}$ is the design matrix for the functional adjustment regression model, and $\boldsymbol{u}(\boldsymbol{x})^T = (u_1(\boldsymbol{x}), \ldots, u_p(\boldsymbol{x}))$. As was done for the constant adjustment model, the weighted bootstrap method can be directly applied to estimate the functional-adjustment parameters and prediction interval in the local sense.

The weighted bootstrap procedure introduced here for local calibration is a very general method for locally calibrating models. Its simplicity lies in the fact that the

resampled data take into account the local weighting, so that any model calibration methodologies can be applied without any additional, complicated analysis. In some instances, model selection needs to be incorporated into the calibration process. For example, in the functional-adjustment model can incorporate any number of regressors, $u(\boldsymbol{x})$. In such cases, ad-hoc analysis can be done using only the data with large weights to select which regressors to incorporate. Alternatively, a small number weighted bootstrap samples can be scrutinized to select the best adjustment model. In either case, the weighted bootstrap procedure can then be applied to the "best" model or even a select few models.

3.3.2 Local Model Averaging

The physics of many engineering process are understood to the point where analytical-engineering models can be proposed to model the mean of the true system behavior. For example, Joseph and Melkote [29] consider modeling the surface roughness in a micro-cutting process. Based on the mechanics of the process, multiple engineering models were proposed to model the process. Similarly, Zhao *et al.* [58] used a solution to a set of partial differential equations as the mean model for describing the material removal rate of a wire saw slicing process. In some cases, analytical models can be proposed over the whole design space, yet in others, analytical models may only describe the process behavior on subsets of the complete design space, resulting in *partial models*. For example, *reaction-limited* and *diffusion-limited* models could be used to describe the chemical process at low and high temperatures, respectively, yet for moderate temperatures, potentially both *limiting steps* occur. Similarly, a computer experiment may have been previously performed for a computer model describing the system behavior of a physical process on a subregion of the complete design space. The resulting metamodel is then restricted to the specific subregion and cannot be extrapolated to the complete design space. With a local model averaging

approach, we are able to combine all partial models to build a complete model over the entire design space.

To take advantage of statistical models and engineering knowledge (via potentially calibrated analytic or meta- models), we propose a locally averaged model procedure. In particular, denote the complete design space by D, consisting of points $\boldsymbol{x}_i$ with corresponding response y_i. Consider m competing models, given by $y_j(\boldsymbol{x}) = f_j(\boldsymbol{x}) + \epsilon_j$, $j = 1, \ldots, m$, where $\epsilon_j \sim N(0, \sigma_j^2)$. The locally averaged model at the point $\boldsymbol{x}$ is defined as

$$g(\boldsymbol{x}) = \sum_{j=1}^{m} w_j(\boldsymbol{x}) f_j(\boldsymbol{x}), \tag{6}$$

where we emphasize that each model weight is dependent upon the location of the design space. For each fixed $\boldsymbol{x}$, all the collected data are assigned weights using a kernel weighting method and the optimal mixture weights in (6) are estimated using the EM-algorithm. This approach allows the mixture weights to be estimated directly for any design point in the design space, not just at observed points.

To estimate the model weights at the point $\tilde{\boldsymbol{x}}$, we use local likelihood estimation procedure as follows. Assign a weight to each data point using a kernel function, i.e., assign the importance of the data point for the estimation of the mixture weights. Specifically, to estimate the mixture weights at $\tilde{\boldsymbol{x}}$, define the kernel weight for point $\boldsymbol{x}_i$, denoted $k_i(\tilde{\boldsymbol{x}})$, according to the formula

$$k_i(\tilde{\boldsymbol{x}}) = K\left(\frac{\boldsymbol{x}_i - \tilde{\boldsymbol{x}}}{h(\tilde{\boldsymbol{x}})}\right),$$

where $h(\tilde{\boldsymbol{x}})$ defines a bandwidth resulting in the *smoothing* window $(\tilde{\boldsymbol{x}} - h(\tilde{\boldsymbol{x}}), \tilde{\boldsymbol{x}} + h(\tilde{\boldsymbol{x}}))$ and $K(u)$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(u)du = 1$ and $K(u) = K(-u)$ for all u. Equipped with these kernel weights, we can then estimate the mixture weights, along with any unknown model parameters, by using the EM algorithm to maximize the local log likelihood [35]. Denote the complete vector of unknown model parameters at the point $\tilde{\boldsymbol{x}}$ as $\boldsymbol{\theta}(\tilde{\boldsymbol{x}})$ and let $\Theta(\tilde{\boldsymbol{x}}) = \{\boldsymbol{w}(\tilde{\boldsymbol{x}}), \boldsymbol{\theta}(\tilde{\boldsymbol{x}})\}$, where $\boldsymbol{w}(\tilde{\boldsymbol{x}}) =$

$(w_1(\tilde{\boldsymbol{x}}), \ldots, w_m(\tilde{\boldsymbol{x}}))$. For example, $\boldsymbol{\theta}(\tilde{\boldsymbol{x}})$ could include unknown model variances for an uncalibrated analytic engineering models. Then, for any given point in the design space $\tilde{\boldsymbol{x}}$, we seek to find

$$\begin{aligned} \arg\max_{\Theta(\tilde{\boldsymbol{x}})} l(\Theta(\tilde{\boldsymbol{x}})) &= \arg\max_{\Theta(\tilde{\boldsymbol{x}})} \sum_{i=1}^{n} k_i(\tilde{\boldsymbol{x}}) \log(g(\tilde{\boldsymbol{x}}, \Theta(\tilde{\boldsymbol{x}}))) \\ &= \arg\max_{\Theta(\tilde{\boldsymbol{x}})} \sum_{i=1}^{n} k_i(\tilde{\boldsymbol{x}}) \log(\sum_{j=1}^{m} w_j(\tilde{\boldsymbol{x}}) f_j(\tilde{\boldsymbol{x}}, \boldsymbol{\theta}(\tilde{\boldsymbol{x}})))). \end{aligned} \tag{7}$$

To find the maximum likelihood estimates in equation (7), we propose using the EM algorithm. Assume there is a set of latent variables, $z_1, \ldots, z_m$, representing the model which generated the response value at data point $\boldsymbol{x}_i$. In particular, if y_i was generated from model l, $l = 1, \ldots, m$, then $z_i = l$. Additionally, let $f(l \mid y_i, \Theta)$ denote $P(z_i = l \mid y_i, \Theta)$ and $f_l(y_i \mid \Theta))$ denote the distribution of the response given by model l. The EM algorithm then works as follows.

Proposition 3.1. Let there be m competing models, and let z_i be a latent variable representing the model class at design point $\boldsymbol{x}_i$. Additionally, let $\Theta^{(t-1)}$ be the current estimate of Θ (i.e., after iteration $t-1$) and let $f(l \mid y_i, \Theta) = P(z_i = l \mid y_i, \Theta)$. For each fixed point $\tilde{\boldsymbol{x}}$ in the design space, let k_i denote the kernel weights for each data point. Then, the *expectation step* for optimizing equation (7) is given by

$$\begin{aligned} E_{\boldsymbol{z}}\left[\log f(\boldsymbol{y}, \boldsymbol{z} \mid \Theta) \mid \boldsymbol{y}, \Theta^{(t-1)}\right] &= C + \sum_{l=1}^{m} \sum_{i=1}^{n} k_i \log(w_l) f(l \mid y_i, \Theta^{(t-1)}) \\ &+ \sum_{l=1}^{m} \sum_{i=1}^{n} k_i \log(f_l(y_i \mid \Theta)) f(l \mid y_i, \Theta^{(t-1)}), \end{aligned}$$

and the *maximization step* is given by

$$\hat{w}_l = \frac{\sum_{i=1}^{N} k_i f(l|y_i, \Theta^{(t-1)})}{\sum_{j=1}^{N} k_i}$$

and

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta} \in \Theta \setminus \boldsymbol{w}} \sum_{l=1}^{m} \sum_{i=1}^{n} k_i \log(f_l(y_i \mid \Theta)) f(l \mid y_i, \Theta^{(t-1)}).$$

For notational simplicity, we suppressed the dependence of the parameters on $\tilde{\boldsymbol{x}}$ in the notation.

Proof. See Appendix. □

Estimating the mixture weights is conditional on the given models as done. For uncalibrated engineering models, where only the mean is specified without any measure of variation, a variance estimate needs to be estimated from the data. Formally, uncalibrated/unadjusted, analytical engineering model f is included by assuming the model

$$Y \sim N(f(\boldsymbol{x}, \boldsymbol{\beta}), \tau^2),$$

where τ is treated as unknown and needs to be estimated in the optimization in equation (7).

Example 3.2. To provide a simple illustrations of the local model averaging, consider the unknown response function

$$f(x) = \begin{cases} \frac{1}{2}\sin(\pi x), & \text{if } x \leq 6; \\ \frac{\pi}{2}(x-6), & \text{if } x > 6. \end{cases}$$

The local averaging methodology using the tricube kernel, i.e., $K(x) = \{1 - |x|^3\}^3$, is applied to the calibrated engineering models from Example 3.1. The estimated model weights are also shown in the left plot of Figure 16. Both models are assumed to follow a normal distribution with constant error variance. From the plot of the estimated weights, model 1 is given sole preference at about $x \leq 5.75$ and model 2 is given sole preference at about $x \geq 6.5$. Between about 5.5 and 6.5, both models are given non-zero weights, with model 1 being preferred up until about $x = 6.25$. The plot on the right in Figure 16 show the expected value of the locally averaged calibrated engineering models along with the approximate 95% prediction intervals.

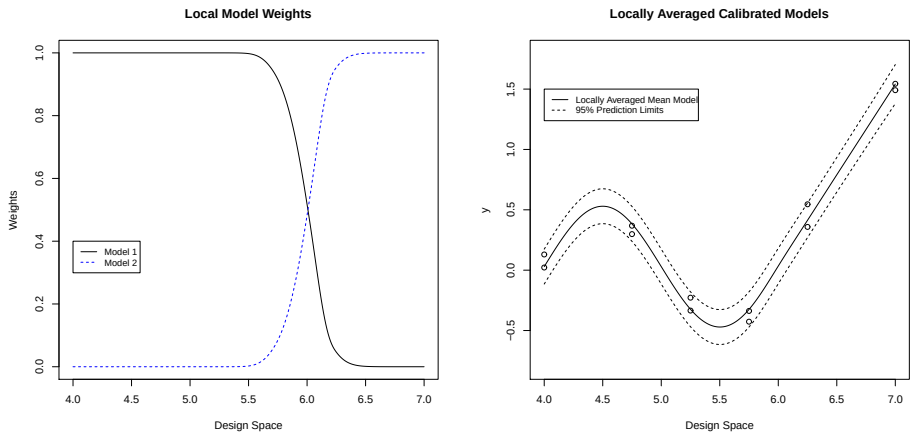

Figure 16: Example of local model averaging: (*left*) the locally estimated model weights for the calibrated models in Example 3.1 and (*right*) the overall locally averaged model are shown.

This locally averaged model pulls from the strengths of each calibrated model to create a unified model that accurately approximates the true response function. □

3.4 Design Methodology

Experimentation on new, advanced technologies can be very costly and time consuming. Even after an experiment is conducted, analyzing the output can be just as costly and time consuming as the experiment itself. For example, equipment used to characterize materials can be too expensive to purchase, resulting in the need for experimenters to pay for rental time to use the equipment. Moreover, analyzing the experimental output can be time consuming, in turn, limiting the number of outputs that can be analyzed in a given day. A good design methodology should wisely select design points to minimize this expense of resources in both the design and analysis stages while at the same time, ensure that enough information is gained to accurately address the experimental goals. We consider the case when batch experiments must be conducted due to cost constraints or physical restrictions.

Based on the principles of the Layers-of-Experiments (LOE) approach in Kim [32],

the design methodology here achieves these objectives in the following ways. First, whenever possible, statistically-adjusted engineering models are combined into a unified engineering-statistics model to drive the experimental designs. Combined engineering and statistical models can make a very efficient use of data by allowing engineering knowledge to fill in the gaps in potentially sparse data sets. In addition, a flexible local polynomial regression model is used whenever or wherever engineering knowledge is lacking. Second, a batch sequential procedure is used. Running experiments in batches can reduce the cost at the analysis stage compared to using a purely sequential approach, and in some settings, only designing experiments in batches is the only practical scenario. Using a sequential design allows models to be updated to more efficiently guide future design plans. Furthermore, we use an adaptive design region as in the LOE approach of [32], where at each stage of experimentation, the design region is adapted (contracted or expanded) to contain only regions of experimental interest. For example, if the goal is to produce a product at a certain target value and within a tolerance level, the design space is adapted to contain only regions with a "high enough" probability of producing such a product. By eliminating whole regions that are deemed unsatisfactory, experimental design plans only focus on the objective at hand without potentially wasting design points.

A powerful modeling approach has been provided in Section 3.3 by exploiting engineering knowledge. Additionally, a local polynomial regression model is used to build a model over the complete design space which is to be incorporated as a competing source of information in local model averaging. Local polynomial regression is chosen due to its flexibility to locally adjust to incoming data. Since the LOE methodology focuses on adapting the design region, it is necessary to have a model that can be updated locally. To take advantage of the efforts put into modeling, a model based experimental design is proposed.

The desired experimental design properties for this work must balance two criteria:

the search for the optimal point and spread out points to search the design space for updating models and exploration. Compound design criteria–designs to combine two or more objectives–have been considered. Typically, approaches include taking a weighted average of different criteria [28], assigning a fraction of the design runs by one criteria and the rest by another [38, 23], and combining/constraining design criteria [2, 17, 27, 32]. Kim [32] and Casciato *et al.* [9] proposed a new, adaptive combined criteria to combine D-optimal and space filling criteria by restricting the distance between any two points when producing D-optimal designs. In the LOE setting, this design criteria is adapted over time to balance the two criteria, where the space-filling properties are deemphasized through time. Kim developed this criteria to address the same goal as those described here, however, the D-optimal criteria is not developed directly for optimization and can be restrictive (allowing linear or linear approximations to models). Alternatively, we address combining space-filling and optimization design criteria in the LOE framework by modifying the SMED procedure provided in Dasgupta [17] and Joseph, Dasgupta, and Wu [27].

Our goal is not to develop a new batch SMED procedure but instead, to take advantage of its flexibility and ease of incorporation into the LOE framework. However, some modifications need to be made to address the LOE principles. The choice to adapt follow-up design regions needs to be developed, and methodology on the choice of how to balance the space-filling and optimality criteria needs to be given. To this end, a review of the SMED algorithm is discussed, followed by its direct extension to a batch scenario. Following, modifications to the batch SMED (BSMED) procedure are given to allow its incorporation into the LOE framework. We refer to this new approach as LOE-BSMED.

3.4.1 Sequential Minimum Energy Design

The sequential minimum energy design procedure is a sequential design method developed by Dasgupta [17] and Joseph, Dasgupta, and Wu [27] in order to find the global maximum of a function that lies between 0 and 1. For example, Dasgupta and Joseph, Dasgupta, and Wu considered maximizing the percent yield of nanowires from an experiment. Similarly, for this work, we will be concerned with maximizing the probability that a response value lies within some target region. The main idea behind the procedure is based on a principle in physics – if n charged particles, all with the same sign of charge, are placed into a box, then the particles will disperse throughout the box before reaching a final position that minimizes the total potential energy of the system. By considering the potential design points as charged particles– each with wisely chosen charges–a new design point is selected to minimize the total potential energy of the system. Let D denote the design space and $\mathcal{X}^t$ be the set of design points in the experiment at time t. Then at time $t+1$ the next design point is chosen as

$$\arg \min_{\boldsymbol{x}_{t+1} \in D} \sum_{\boldsymbol{x}_i \in \mathcal{X}^t} \frac{q(\boldsymbol{x}_{t+1}) q(\boldsymbol{x}_i)}{d(\boldsymbol{x}_{t+1}, \boldsymbol{x}_i)},$$

where $d(\boldsymbol{x}_{t+1}, \boldsymbol{x}_i)$ denotes the Euclidean distance between $\boldsymbol{x}_{t+1}$ and $\boldsymbol{x}_i$.

Below, we provide the main components of the SMED procedure as provided in Joseph, Dasgupta, and Wu. For more details and discussion on this procedure, refer to Dasgupta [17] and Joseph, Dasgupta, and Wu [27]. The charge of a design point $\boldsymbol{x}$ is assigned as $q(\boldsymbol{x}) = (1 - \alpha \hat{p}(\boldsymbol{x}))^\gamma$, where α is a parameter estimated from the current data, γ is tuning parameter set by the experimenter, and $\hat{p}(\boldsymbol{x})$ is the current estimated value of the response/objective function (between 0 and 1) at $\boldsymbol{x}$. The parameter α is sequentially estimated as $1/p_g^{(n)}$, where $p_g^{(n)}$ is the estimate of the global optimum after n points, $\boldsymbol{x}_i, \ldots, \boldsymbol{x}_n$, have been selected. The value $p_g^{(n)}$ is estimated as follows. Let $p_{(n)} = \max_{1 \le i \le n} p(\boldsymbol{x}_i)$, i.e., the maximum observed response, and let n_0 be the total

number of design points until termination of the algorithm. Then $p_g^{(n)}$ is estimated by fitting the model

$$p_{(i)} = p_g^{(n)} \frac{\exp(\beta_0^{(n)} + \beta_1^{(n)} i)}{1 + \exp(\beta_0^{(n)} + \beta_1^{(n)} i)}, i = 1, \ldots, n.$$

In addition to the n observed points, two more points are used to fit the above model. These points are $(0, \frac{p_{(1)}}{2})$ and $(n_0, \tilde{p}_{(n)})$, where $\tilde{p}_{(n)} = \frac{np_{(n)}+1}{n+1}$. Estimation for the above logistic model is done by minimizing the sum of squared residuals subject to $p_g^{(n)} \geq (p_{(n)} + \tilde{p}_{(n)})/2$.

3.4.2 Layers-of-Experiments Batch Sequential Minimum Energy Design

Modifications to the SMED procedure are developed in this section to incorporate the batch sequential experimental design in the Layers-of-Experiments setting. Consider selecting a batch of size b at the new design stage. After the selection of n points, the minimum energy design criteria is to find the b points $(\boldsymbol{x}_{n+1}, \ldots, \boldsymbol{x}_{n+b}) \in D$ to minimize

$$E_n = \sum_{i=1}^{n+b-1} \sum_{j=i+1}^{n+b} \frac{q(\boldsymbol{x}_j) q(\boldsymbol{x}_i)}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)}. \tag{8}$$

This batch metric was briefly mentioned in Joseph, Dasgupta, and Wu without details or illustration. Where needed, we provide details and discussion on this batch sequential minimum energy design (BSMED) provided in Joseph, Dasgupta, Wu [27], and in addition, we provide insight on how aspects of the LOE formulation can be beneficial to the BSMED procedure, even when the full LOE methodology isn't applied.

3.4.2.1 The Choice of $q(\boldsymbol{x}_i)$

Assigning the energies is an important role. For this paper, the energies are assigned as in Dasgupta [17] and Joseph, Dasgupta, and Wu [27], i.e., $q(\boldsymbol{x}_i) = (1 - \alpha\hat{p}(\boldsymbol{x}_i))^\gamma$, where $\hat{p}(\boldsymbol{x}_i)$ is the estimated objective function (restricted to functions between 0 and 1) at $\boldsymbol{x}_i$. For example, in Dasgupta and Joseph, Dasgupta, and Wu, $p(\boldsymbol{x})$ was

the response surface when considering optimizing the percent yield. In this work, where we consider producing a product at a nominal level, $p(\boldsymbol{x})$ is the probability that a product is produced within some tolerance band of the target value. In our setting, where competing sources of information are provided, the objective function will be estimated by using the locally averaged model. An alternative assignment of energies that enables the incorporation of engineering trust has also been proposed for collaborative scenarios in which engineering expertise can be incorporated in the design beyond the use of mean and variance models. See Vastola *et al.* [46] for more details.

For clarification, estimating the global maximum of the objective function, $p_g^{(n^b)}$, and hence α, has to be addressed in the batch scenario. Let n^b denote the current number of batches that have been collected (including the initial design), and let n denote the total number of points currently in the design (also including the initial design). Let $p_{(n^b)} = \max_{1 \leq i \leq n} p(\boldsymbol{x}_i)$, i.e., the maximum observed response, and let n_0^b be the total number of batches of design points (including the initial experiment) that are going to be taken until terminating the algorithm. Then $p_g^{(n^b)}$ is estimated by fitting the model

$$p_{(i)} = p_g^{(n^b)} \frac{\exp(\beta_0^{(n^b)} + \beta_1^{(n^b)} i)}{1 + \exp(\beta_0^{(n^b)} + \beta_1^{(n^b)} i)}, i = 1, \ldots, n^b.$$

In addition to the n^b observed maximum points, two more points are used to fit the above model. These points are $(0, \frac{p_{(1)}}{2})$ and $(n_0^b, \tilde{p}_{(n^b)})$, where $\tilde{p}_{(n^b)} = \frac{n^b p_{(n^b)} + 1}{n^b + 1}$. Estimation for the above logistic model is done by minimizing the sum of squared residuals subject to $p_g^{(n^b)} \geq (p_{(n^b)} + \tilde{p}_{(n^b)})/2$. Differing from the SMED procedure, we propose taking $p_{(i)}$ to be the maximum of the predicted objective function after the i-th batch experiment over the entire experimental domain instead of just the maximum at the observed points, i.e., $p_{(i)} = \max_{\boldsymbol{x}_i \in D} \hat{p}(\boldsymbol{x}_i), i = 1, \ldots, n^b$. Defining $p_{(i)}$ in this manner guarantees the energies are between 0 and 1 which is not guaranteed when choosing $p_{(i)}$ just based off of the observed values.

3.4.2.2 The Choice of γ

The choice of γ plays an important role in the design. Vastola *et al.* [46] discussed the effects of the parameter γ in designing an initial experimental design, and Dasgupta [17] and Joseph, Dasgupta, and Wu discussed it in a purely sequential setting. There is a clear trade-off: lower values of γ lead to designs emphasizing space filling properties, while higher values of γ lead to designs that aim to satisfy the optimality criteria. Instead of holding γ fixed, sequentially updating it can exploit this trade off. For the purely sequential case, Dasgupta [17] considers $\gamma = 1, 2$, or 3 and Joseph, Dasgupta, and Wu [27] recommended setting γ to 1, yet we show in Section 3.5 that in a batch setting, these values are too low, giving poor results.

When collaborating with scientists and engineers, expert knowledge may be used to set the value of γ. Potential designs with differing values of γ can be viewed before any experiments are conducted in the next batch. Experimenters can review each design and decide which γ best suites their needs. For example, an expert could believe that the model prediction is not mature enough at the current stage or the results may go against engineering intuition. In such cases, a lower value of γ could be chosen in order to encourage exploring the design space as opposed to searching for the global optimum. In some cases, choosing a large γ can result in designs that place points so close together, that for all practical purposes, it is infeasible to distinguish the points apart. For example, if a systems temperature control unit has an error of ± 5 degrees centigrade, then it is not reasonable to consider conducting physical experiments that are only a couple of degrees centigrade apart.

Another way to set γ is to use an automated procedure, but first, an idea of the magnitude of γ needs to be developed. Dasgupta [17] and Joseph, Dasgupta, and Wu [27] related the convergence of the SMED algorithm to the energies in the following manner: the sequential design converges if $q(\boldsymbol{x}_i) < \delta$, where δ is a small positive number. Using this concept, the minimum value of γ necessary to force

convergence of the algorithm (at least in the purely sequential setting) is the value γ_{max} that forces the minimum energy to be less than δ, i.e.,

$$\gamma_{max} = \frac{\log(\delta)}{\log(1 - \alpha \max_{\boldsymbol{x}_i} p(\boldsymbol{x}_i))}.$$

Here, γ_{max} is calculated only after the initial experiment. Setting $\gamma = \gamma_{max}$ forces convergence instead of exploration, when in early designs, exploration should be emphasized. Therefore, γ can be set to only a fraction of γ_{max}, namely, $\gamma = \kappa\gamma_{max}$, where $0 \leq \kappa \leq 1$. Setting κ will be discussed later, but the general idea is to increase κ as the design objective is closer to being satisfied.

3.4.2.3 Adaptive Region Choice

The concept of adapting design regions has been explored in methodology such as response surface methodology [8], the Grid Algorithm [49, 50], Layers-of-Experiments [32], SELC [39], G-SELC [38], and others. Typically, these methods take one of two approaches: (i) shrink the design region to target only promising regions, or (ii) eliminate design regions that are unpromising. Each approach has its benefits. The former approach has the potential to quickly find the optimal operating condition, while the latter approach, although slower to converge, cautiously removes design spaces as to not miss the optimal region. However, zoom-in approaches run the risk of prematurely zooming into the wrong design region, especially when dealing with complicated functions, while the latter approach can require too many experimental runs rendering it inadequate for high-cost, low-resource experiments. To reap the benefits of each methodology, the procedures are combined.

Early stages in sequential design should be used for exploring the design space and validating models before spending resources on targeting the objective. Therefore, in earlier stages, uninteresting regions of the design space will be removed, while in later stages, the design region will zoom in on optimal regions. We call a sequential design a *purely elimination design* if after each iteration, the follow-up design region

corresponding to the lower 5% of the objective function is removed. Alternatively, we consider a follow-up design a *purely zoom-in design* if at each iteration, the design region is shrunk to the region corresponding to the upper 5% of the objective function. Denote the threshold for the purely elimination design at the k-th iteration by $\xi_{.05}^{(k)}$ and the threshold for the purely zoom-in design at the k-th iteration by $\xi_{.95}^{(k)}$. In particular, $\xi_{.05}^{(k)}$ the 5% quantile of the estimated objective function (at step k) and $\xi_{.95}^{(k)}$ the 95% quantile of the estimated objective function (at step k). A new threshold is proposed as $\tau^{(k)} = (1-\kappa)\xi_{.05}^{(k)} + \kappa\xi_{.95}^{(k)}$, where κ is the same as the coefficient for γ_{max}.

The proposed threshold is used to determine the next design region for follow-up experimentation. For example, suppose an experiment is to be conducted in three batches. In the initial experiment, the design points are chosen from the entire design region. Using the data collected from the initial experiment, the objective function is estimated and $\xi_{.05}^{(1)}$ and are $\xi_{.95}^{(1)}$ calculated. The threshold is then defined as $\tau^{(1)} = (1-\kappa)\xi_{.05}^{(1)} + \kappa\xi_{.95}^{(1)}$. Setting κ is discussed later. All design points with corresponding estimated objective function less than $\tau^{(1)}$ are removed from consideration for the follow-up design. In particular, the design region for the second round of experimentation is defined as all the

All quantiles are calculated over the entire experimental region, not just in the follow-up design regions, allowing for subsequent regions to be larger or smaller than previous regions. The goal is to define κ so that it increases over time, resulting in the more cautious elimination approach in earlier stages and a more aggressive zoom-in approach in later stages.

As a side note, we discuss here the potential benefit of adapting the design region on the BSMED procedure, even when other aspects of the LOE approach are not considered. Notice that the metric in equation (8) is a function of the distance between the points. Potential design points have a natural repelling force that pushes them

away from each other unless the effect of the charges overshadows the repulsion. This fact is not an issue when using a non-batch sequential design where the model and design parameters are updated before selecting a subsequent point. In the batch setting, however, simply choosing b design points can result in the points spreading away from each other; points can be placed in regions that are highly unlikely to be optimal based solely on the distance between them. By first restricting the design space, the negative effects of repulsion are counteracted. Consider the following simple example.

Example 3.3. Let $f(x_1, x_2) = 30 - (8x_1 - 4)^2 - (8x_2 - 4)^2$, where $(x_1, x_2) \in [0, 1]^2$, and consider the random function defined by $y = f(x_1, x_2) + \epsilon$, where $\epsilon \sim N(0, 1)$. Suppose the goal of experimentation is to find the operating condition that maximizes the probability that a product is produced at a target mean of $T = 30$ and within a tolerance of $d = 1$. The optimal operating condition in this example occurs when $(x_1, x_2) = (1/2, 1/2)$. At this point, the probability that the product meets our tolerance requirements is 0.6826. We compare two design methods. In the method 1, we apply the BSMED procedure. In method 2, we apply the BSMED procedure while at the same time adapting the design region. For this example, we consider a simple procedure to adapt the design region to illustrate the point. All design points whose predicted response is not within ± 1.96 standard prediction errors of T are eliminated before the next batch is drawn. Here, γ is set to 1, local polynomial regression is used to estimate the response surface, α is updated in the same manner as in Joseph, Dasgupta, and Wu [27], and an initial experiment (using a space-filling design of size 8, with 3 repetitions) is conducted to obtain a stable approximation of the initial loess regression model. Two follow-up experiments are then conducted by selecting batch sizes of $b = 3$ with three repetitions taken at each design point. An exchange algorithm [19] (discussed in Section 3.4.2.5) is used to find the optimal designs.

Table 5 shows the results of 100 simulations. Considering only the six selected batch

Table 5: Example 3.3 Results (Average Over 100 Simulations)

	Max Probability	Mean Probability	# No Yield Pts.
Non-Adaptive Region	0.48	0.18	4.21
Adaptive Region	0.64	0.50	0.53

points (collected in two batches), the values reported in Table 5 are: (i) the maximum probability for the true response; (ii) the average probability for the true response; and (iii) the number batch points placed in zero probability regions. Each of these three values is averaged over 100 simulations. □

The overall goal of the experiments being considered is to find the optimal operating condition to meet the engineering tolerance. In Example 3.3, adapting the design region results in placing at least one point closer to the optimal operating condition on average. Moreover, the average probability of meeting the engineering tolerance for all the batch points is also higher when the design space is adapted. Lastly, the number of batch points placed in no yield regions is lower. The latter two results are especially important when engineering models are being considered to guide model-based experimentation. The models are being validated, adjusted, and/or updated after each round of experimentation. In a limited resource setting, the number of design points required to, for example, adjust an engineering model are likely be few and far between. Therefore, it is important to design experiments that place points in locations that are more likely to contain the optimal operating condition. Engineering models are then at least able to be validated and adjusted in regions that are important to achieve the overall goal.

3.4.2.4 The Choice of κ

The role of κ is two fold: it dictates the aggressiveness in setting γ and in choosing the subsequent design region. On the two extremes, setting $\kappa = 0$ results in a purely

space-filling, elimination sequential design, while setting $\kappa = 1$ results in a purely objective oriented, zoom-in sequential design. The goal is to set κ closer to 0 in early stages to encourage exploration of the design space and to set κ closer to 1 in later design stages to encourage targeting the objective.

There are many ways to choose κ. First, consider a collaborative setting where experts can view the effects of different κ values and choose the experimental design that best suites their needs. A larger value of κ may result in values of γ that are too aggressive, as discussed above, and subsequent design regions that eliminate too many design points from consideration. Alternatively, smaller values of κ can yield designs that are too passive in the search for the global optimum.

Automated approaches can instead be taken to choose the value of κ. Given the maximum allowable resources, i.e., the maximum number of batch experiments that can be run, κ can be set using a step function. Let n^b be the number of batch experiments, including the initial experiment, that are to be run. For the i-th batch experiment, set $\kappa = (i-1)/n^b$ for $i = 2, \ldots, n^b$. For a discussion on setting κ based on the information gained from experimentation, see Kim [32], who considers a similar sequential design scenario.

3.4.2.5 Exchange Algorithm

An exchange algorithm [19] is used to find the design the minimizes the total energy. Given n points in the design, the total potential energy after adding b points is

$$E_{n+b} = \sum_{i=1}^{n+b-1} \sum_{j=i+1}^{n+b} \frac{\hat{q}(\boldsymbol{x}_i)\hat{q}(\boldsymbol{x}_j))}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)}.$$

Notice that E_{n+b} can be decomposed into a fixed component containing the n existing design points and a variable component containing the additional energy introduced by adding b design points. The total energy becomes:

$$E_{n+b} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\hat{q}(\boldsymbol{x}_i)\hat{q}(\boldsymbol{x}_j))}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)} + \sum_{i=n+1}^{b} \sum_{j=1}^{n} \frac{\hat{q}(\boldsymbol{x}_i)\hat{q}(\boldsymbol{x}_j))}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)} + \sum_{i=n+1}^{b-1} \sum_{j=i+1}^{b} \frac{\hat{q}(\boldsymbol{x}_i)\hat{q}(\boldsymbol{x}_j))}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)}.$$

Since the first component is irreducible, the goal is to find the b points that minimize

$$E_b^{(n)} \equiv \sum_{i=n+1}^{b} \sum_{j=1}^{n} \frac{\hat{q}(\boldsymbol{x}_i)\hat{q}(\boldsymbol{x}_j))}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)} + \sum_{i=n+1}^{b-1} \sum_{j=i+1}^{b} \frac{\hat{q}(\boldsymbol{x}_i)\hat{q}(\boldsymbol{x}_j))}{d(\boldsymbol{x}_i, \boldsymbol{x}_j)}.$$

The algorithm works as follows:

1. Denote the current design region excluding the fixed n design points by $D^{(n)}$. Randomly select b points from $D^{(n)}$, denoting these points by the vector $\boldsymbol{b} = (d_1, d_2, \ldots, d_b)$;

2. Calculate the current energy $E_b^{(n)}$, and denote this value by E^c. Set $i = 1$;

3. For each point x_j in the design space $D^{(n)} \backslash \boldsymbol{b}$, replace d_i with x_j and calculate $E_{b,j}^{(n)}$, the value of $E_b^{(n)}$ when d_i is replaced by x_j. Let $E_b^{(n)*} = \min_j E_{b,j}^{(n)}$ and x^* be the corresponding design point to $E_b^{(n)*}$. If $E_b^{(n)*} < E^c$, set d_i equal to x_i^*. Update $\boldsymbol{b}$ to reflect this change. Set $i = i + 1$;

4. Repeat steps 2 and 3 until $i = b + 1$, at which point, go to step 5;

5. Repeat steps 2 through 4 until no more exchanges are made or the maximum allowable iterations is achieved.

Step 3 can efficiently be calculated by noting that when replacing d_i with candidate point x_j, $E_{b,j}^{(n)}$ can be decomposed as

$$\begin{aligned} E_{b,j}^{(n)} &= E_b^{(n)} - \sum_{k=1}^{n} \frac{\hat{q}(d_i)\hat{q}(\boldsymbol{x}_k))}{d(d_i, \boldsymbol{x}_k)} - \sum_{k=n+1, k\neq i}^{b} \frac{\hat{q}(d_i)\hat{q}(d_k))}{d(d_i, d_k)} \\ &+ \sum_{k=1}^{n} \frac{\hat{q}(x_j)\hat{q}(\boldsymbol{x}_k))}{d(x_j, \boldsymbol{x}_k)} + \sum_{k=n+1, k\neq j}^{b} \frac{\hat{q}(x_j)\hat{q}(d_k))}{d(x_j, d_k)}. \end{aligned}$$

Therefore, the total potential energy of the proposed design doesn't have to be completely re-calculated at each step, but instead, can just be updated by removing the energies corresponding to d_i and adding the energies corresponding to x_j. Due to the randomness in the initial design in step 1, it is recommended to run the exchange algorithm many times and choose the best design resulting from the multiple runs.

3.5 Combined Methodology: Layers-of-Experiments

Combining the modeling methodology presented in Section 3.3 with the experimental design methodology presented in Section 3.4 provides a powerful methodology for process optimization since both the modeling techniques and design methodology in and of themselves are effective techniques. Figure 17 gives a schematic of the combined methodology.

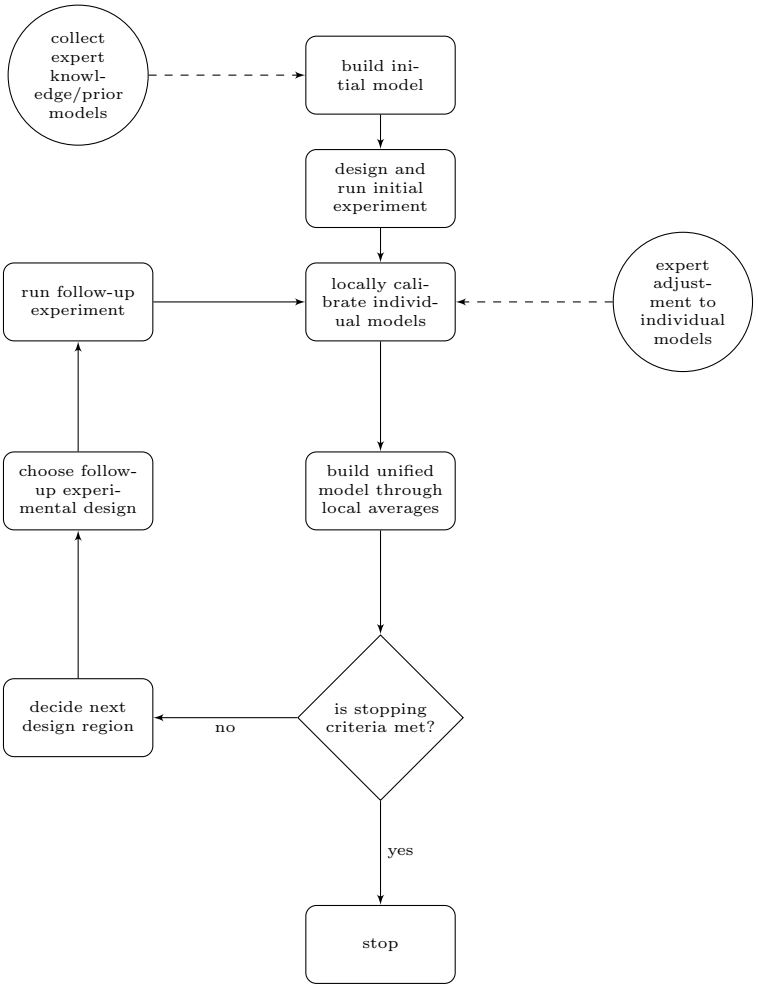

Figure 17: A schematic of design and modeling process is shown here.

The entire process optimization process begins with an initial experiment that combines all prior knowledge, e.g., engineering models, expert opinion, or past experimental results on similar, yet non-identical systems. Vastola *et al.* [46] provide

a flexible and effective procedure for combining this information to plan an initial experiments. If no prior knowledge exists, a standard space-filling design can be used for the initial experiment. The data collected in the initial experiment are used to calibrate existing engineering models based on the methodology in Section 2. In addition, the models can be adjusted by experts based on the knowledge learned from the initial experiments. For example, additional model terms can be added to account for physical processes that were unaccounted for in the original engineering models. The updated and adjusted engineering models are locally averaged into a unified model. In addition to engineering models, statistical models can be incorporated into the locally averaged model. The statistical models we consider are local polynomial regression models. These models have the advantage to adapt locally, which goes hand-in-hand with our methodology. Since the design region adapts to include only sub-regions of interest, from which subsequent design points are chosen, local polynomial regression provides a flexible procedure for adjusting the statistical model locally. If no engineering models are present, only purely data-driven models can be used.

The unified, locally averaged model is first used to select the follow-up design region and tuning parameter γ, as discussed in subsections 3.2.3 through 3.2.5. Following, the unified model is used to estimate the objective function, namely $p(\boldsymbol{x})$, which in turn, is used to assign the energies in our *Layers-of-Experiments, batch sequential minimum energy design* (LOE-BSMED) methodology. The follow-up LOE-BSMED experiment is run and the second model updating and adjustment process begins again. The whole procedure is iterated as in Figure 17 until a stopping criteria is met. Stopping criteria include using the maximum allowance of design points and/or until satisfactory model accuracy is obtained.

3.6 Applications

We consider two applications to illustrate our methodology. First, we consider the CdSe nanowire example studied by Dasgupta [17] and Joseph, Dasgupta, and Wu [27], where the goal is global optimization of the nanowire yield. This application provides a study of the LOE-BSMED design strategy without using the modeling methodology discussed in Section 2. Second, a simulation study on a modified six-hump camel back function is provided to illustrate our complete methodology, incorporating the complete LOE-BSMED methodology (with prior engineering models), to maximize the probability of producing a product within a tolerance limit around a target value.

3.6.1 Nanosynthesis Morphology

Dasgupta *et al.* [18] reported an experiment studying the effects of temperature and pressure on the yield of CdSe nanowires. Dasgupta [17] and Joseph, Dasgupta, and Wu [27] developed the SMED optimization algorithm for and studied its performance on this nanowire fabrication process. Figure 18 shows a plot of response function. The experiment consisted of 45 points, 5 temperatures and 9 pressures, transformed to the unit interval, and the response value is the percentage yield of nanowires. The maximum yield occurred at the point (0.32, 0.25).

In the simulation studies that follow, we compare the BSMED (batch-SMED) and LOE-BSMED design procedures. We mimic the setup as in Joseph, Dasgupta, and Wu, except the experimental points are taken in batches as discussed in Sections 3.4.1 and 3.4.2. Since no prior engineering models exist, the response surface is estimated using only local polynomial regression with spanning parameter $\lambda = 1$ (as it gave the best results). The simulated responses are drawn from a binomial distribution with size r and success probability equal to the observed yield in the actual physical experiment (see Dasgupta [17]). We consider sequential designs with a total number of 20 experimental runs. The 20 experimental runs are allocated in three scenarios:

initial designs of size 5, 8, and 11, with sequential samples taken in sizes of 5, 4, and 3, respectively. The performance of the algorithms are based on three measures: the average number of times that the true optimal value is sampled (over all 20 design points), the average yield over the sampled points (excluding the initial sample), and the average number of points placed in no yield regions (excluding the initial sample). The results are shown in Tables 6-8. For both procedures, the initial designs are space filling designs. For the BSMED procedure, different values of γ are considered. For the LOE-BSMED procedure, $\kappa = i/3, i = 1, 2, 3$ for the 3 layers of follow up experiments, and different values of δ are considered.

The study shows that setting γ too low (equal to 1 as used for the SMED case) gives poor results. This is likely due to the repelling effect inherent in minimum energy designs. For the higher values of γ the results are better, yet the LOE-BSMED procedure performs better by all three measures. The superiority of the LOE-BSMED procedure is more readily seen when the number of points chosen in the follow-up experiments is larger. In other words, LOE-BSMED has a more clear advantage when the initial experiment consists of 5 points. In addition, LOE-BSMED has a very practical advantage. Choosing the value of δ is much more intuitive to choosing what value of γ to use. Moreover, the appropriate value of γ likely changes based on the unknown response function, yet δ is a more universal measure that can interpreted in the same way no matter the problem.

3.6.2 Application on Optimization Function

We illustrate the complete methodology using a modified six-hump camel back function known in optimization literature. The six-hump camel back function is given by

$$g(x_1, x_2) = (4 - 2.1x_1^2 + \frac{x_1^4}{3})x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2.$$

Table 6: Simulation Results (Average Over 2000 Simulations)–11 run initial design; 3 batches of 3 runs each; 20 runs total.

Repetitions	Method	Parameter Values	Max Probability	Mean Probability	# No Yield Points
r=5	BSMED	$\gamma = 1$	.581	.231	1.800
		$\gamma = 5$	.689	.248	1.595
		$\gamma = 10$	.682	.250	1.546
	LOE-BSMED	$\delta = .01$	.706	.259	1.344
		$\delta = .0001$	.721	.262	1.293
		$\delta = .000001$	.697	.261	1.281
r=10	BSMED	$\gamma = 1$	.591	.232	1.999
		$\gamma = 5$	.706	.253	1.580
		$\gamma = 10$	.694	.253	1.138
	LOE-BSMED	$\delta = .01$	.745	.269	1.184
		$\delta = .0001$	.758	.271	1.151
		$\delta = .000001$	.745	.271	1.112
r=25	BSMED	$\gamma = 1$	.586	.234	2.028
		$\gamma = 5$	.704	.255	1.623
		$\gamma = 10$	.702	.253	1.598
	LOE-BSMED	$\delta = .01$	.757	.275	1.126
		$\delta = .0001$	.771	.277	1.051
		$\delta = .000001$	.772	.275	1.083
r=100	BSMED	$\gamma = 1$	.597	.236	2.018
		$\gamma = 5$	.691	.252	1.732
		$\gamma = 10$	.675	.252	1.690
	LOE-BSMED	$\delta = .01$	.773	.280	1.079
		$\delta = .0001$	.776	.278	1.087
		$\delta = .000001$	.762	.278	1.099

Table 7: Simulation Results (Average Over 2000 Simulations)–8 run initial design; 3 batches of 4 runs each; 20 runs total.

Repetitions	Method	Parameter Values	Max Probability	Mean Probability	# No Yield Points
r=5	BSMED	$\gamma = 1$	.578	.221	2.71
		$\gamma = 5$	.730	.253	1.764
		$\gamma = 10$	.740	.255	1.709
	LOE-BSMED	$\delta = .01$	.724	.240	2.106
		$\delta = .0001$	.745	.251	1.794
		$\delta = .000001$	.748	.253	1.765
r=10	BSMED	$\gamma = 1$	.604	.232	1.999
		$\gamma = 5$	.790	.262	1.564
		$\gamma = 10$	.779	.262	1.504
	LOE-BSMED	$\delta = .01$	.784	.256	1.678
		$\delta = .0001$	.826	.264	1.480
		$\delta = .000001$	.804	.263	1.481
r=25	BSMED	$\gamma = 1$	.643	.229	2.490
		$\gamma = 5$	.822	.267	1.403
		$\gamma = 10$	.820	.267	1.381
	LOE-BSMED	$\delta = .01$	.845	.267	1.404
		$\delta = .0001$	.863	.272	1.257
		$\delta = .000001$	.865	.273	1.216
r=100	BSMED	$\gamma = 1$	.618	.231	2.451
		$\gamma = 5$	.853	.271	1.370
		$\gamma = 10$	.855	.273	1.273
	LOE-BSMED	$\delta = .01$	.879	.271	1.314
		$\delta = .0001$	.887	.277	1.139
		$\delta = .000001$	.886	.276	1.159

Table 8: Simulation Results (Average Over 2000 Simulations)–5 run initial design; 3 batches of 5 runs each; 20 runs total.

Repetitions	Method	Parameter Values	Max Probability	Mean Probability	# No Yield Points
r=5	BSMED	$\gamma = 1$	.533	.197	4.069
		$\gamma = 5$	.704	.226	2.900
		$\gamma = 10$	.693	.227	2.87
	LOE-BSMED	$\delta = .01$	.704	.228	2.791
		$\delta = .0001$	.750	.235	2.568
		$\delta = .000001$	.766	.236	2.484
r=10	BSMED	$\gamma = 1$	.533	.200	3.990
		$\gamma = 5$	.736	.231	2.716
		$\gamma = 10$	.759	.235	2.639
	LOE-BSMED	$\delta = .01$	.787	.242	2.281
		$\delta = .0001$	.809	.244	2.208
		$\delta = .000001$	.833	.247	2.129
r=25	BSMED	$\gamma = 1$	.558	.206	3.752
		$\gamma = 5$	.804	.238	2.524
		$\gamma = 10$	.822	.240	2.491
	LOE-BSMED	$\delta = .01$	.869	.251	2.011
		$\delta = .0001$	.870	.250	2.000
		$\delta = .000001$	.874	.251	1.947
r=100	BSMED	$\gamma = 1$	.568	.209	3.63
		$\gamma = 5$	.850	.239	2.468
		$\gamma = 10$	.830	.240	2.393
	LOE-BSMED	$\delta = .01$	.877	.255	1.892
		$\delta = .0001$	.879	.253	1.921
		$\delta = .000001$	.884	.253	1.847

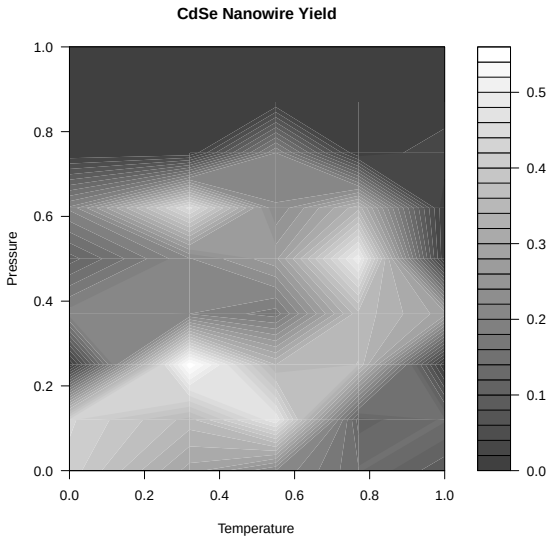

Figure 18: This figure shows contour plot of the CdSe nanowire yield originally studied in Dasgupta [17].

The function $g(x_1, x_2)$ is modified to give a unique global maximum, as opposed to two global minimum, by transforming $g(x_1, x_2)$ to

$$g^*(x_1, x_2) = 15 - (4 - 2.1x_1^2 + \frac{x_1^4}{3})x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2 + .5x_2.$$

Furthermore, the function is rescaled to the unit grid for analysis by substituting $x_1^* = 4x_1 - 2$ and $x_2^* = 2x_2 - 1$ in for x_1 and x_2, respectively. The contour of $g(x_1, x_2)$ is shown in Figure 22.

For illustration, suppose an experimenter is to conduct experiments on a system with true response surface given by $g^*(x_1^*, x_2^*)$ with an associated system standard deviation of $\sigma = .5$. The end goal of the experimenter is to find the optimal system input to create a product at a target value of $T = 16.5$ units with a tolerance of ± 1 unit, i.e., the goal is to find the input (x_1, x_2) that maximizes the probability that the product will have a size between 15.5 and 17.5 units. The true optimal input of the system is $(x_1^*, x_2^*) = (.48, .87)$, and the associated probability of producing the product within the tolerance limits is 0.9545.

Prior to experimentation, expert engineering knowledge provided the following analytical engineering model for the upper-left quadrant of the design region:

$$f_{eng}(x_1^*, x_2^*) = g^*(x_1^*, x_2^*) - x_2^* + 4(x_1^*)^6.$$

In addition, for the upper-right quadrant of the design space, a 16 run full factorial design was used to study a computer model given by

$$f_{comp}(x_1^*, x_2^*) = g^*(x_1^*, x_2^*) - 5.$$

As a result, a universal kriging model is used to analyze the computer experiment. Denote this metamodel by f_{meta}. Note that although the computer model only needs a minor adjustment by adding a constant, the kriging model, f_{meta} was built only for the upper-right quadrant. Outside this quadrant, f_{meta} is far from accurate, even after adjustment. Lastly, in the lower two quadrants, no prior knowledge is available. Thus, a purely data driven statistical model will be relied upon for these quadrants. Figure 19 summarizes the knowledge before experimentation.

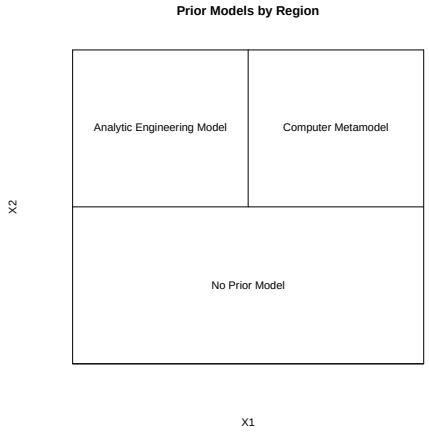

Figure 19: The partition of prior knowledge on the complete design space is shown.

A 12 point space-filling design is used as the initial experimental design, and two batches of follow-up experiments are conducted, each with 4 points. The designs are selected from a design space that has been gridded into 0.05×0.05 size squares.

The final design is shown in Figure 23. At each design point, three repetitions were simulated from $g^*(x_1^*, x_2^*) + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$. After each batch of experimentation (including the initial experiment), f_{eng} and f_{meta} are locally calibrated using functional and constant adjustment models, respectively.

To locally adjust f_{eng}, expert opinion is used to assign weights to the observed data. The intended use of f_{eng} was for the upper-left quadrant, so the observed data in the upper-left quadrant were given (unnormalized) weights of 1; data with inputs within a distance .05 of the upper-left quadrant were assigned a weight of .75; data with inputs within a distance of (.05, .1) of the upper-left quadrant were assigned a (unnormalized) weight of .5; and all other data points were assigned a weight of 0. Since only data in or near the upper-left quadrant were used for calibration, only the x_2^* term was detected for calibration. The term $4(x_1^*)^6$ doesn't significantly affect f_{eng} in and near the upper-left quadrant, so it was not detected for local calibration. As a result the locally calibrated version of f_{eng}, denoted by f_{eng}^F is only accurate for $x_1^* \leq 0.5$.

To locally adjust f_{meta}, expert opinion is used to assign weights to the observed data. The intended use of f_{meta} was for the upper-right quadrant, so the observed data in the upper-right quadrant are given (unnormalized) weights of 1; data with inputs within a distance .05 of the upper-right quadrant are assigned a (unnormalized) weight of .75; data with inputs within a distance of (.05, .1) of the upper-right quadrant were assigned a (unnormalized) weight of .5; and all other data points were assigned a weight of 0. The constant adjustment model was sufficient for detecting the constant difference between g^* and f_{meta}, so the calibrated version of f_{meta}, denoted by f_{meta}^C, accurately represents the overall trend in the upper-right quadrant. Outside of this quadrant, however, it remains inaccurate.

The response in the complete design space (all four quadrants) is modeled using local polynomial regression. Denote this model by f_{loess}. The model f_{loess} fills in

the inadequacies of the prior conjectured models, even after their local calibration. With all three models in hand, the local model weights (local posterior probabilities), denoted by w's, are calculated as in Section 3.3.2, and a unified model is given by

$$\hat{f}(\boldsymbol{x}) = w_{eng}(\boldsymbol{x}) f^F_{eng}(\boldsymbol{x}) + w_{meta}(\boldsymbol{x}) f^C_{meta}(\boldsymbol{x}) + w_{loess}(\boldsymbol{x}) f_{loess}(\boldsymbol{x}).$$

Figure 20 shows each models's weight in the entire design space after the initial experiment. In the two left quadrants, f^F_{eng} is given the majority of the weight; in the upper-right quadrant, f^C_{meta} is given the majority of the weight; and for the two right quadrants, f_{loess} is given a significant weight, especially in the lower-right quadrant. It is seen that f_{loess} "fills in" the gaps left by engineering knowledge.

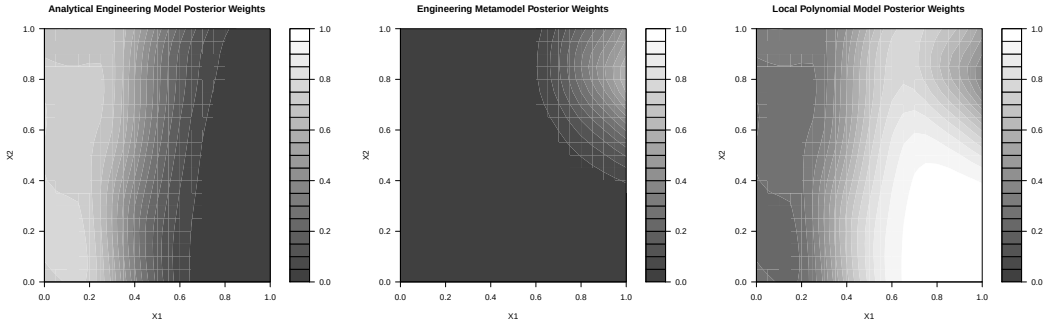

Figure 20: Contours of the estimated local mixture model weights after the initial experiment are shown.

The unified model, $\hat{f}(\boldsymbol{x})$, is used to estimate the the objective function, namely, $p(\boldsymbol{x}) = P(T - d \leq \hat{f}(\boldsymbol{x}) \leq T + d)$, in turn deciding: the assignment of energies, the choice of γ, and the follow-up design region. For this example, $\delta = 0.01$ and κ is set (in accordance with the first automated procedure) as 0.5 and 1 for the first and second follow up designs, respectively. The design regions for the two follow-up batch experiments are shown in Figure 21 along with the estimated objective functions. For the first follow-up design, the region is non-convex and larger than the region for the final follow-up design, which is focused around the region containing the maximum of the objective function.

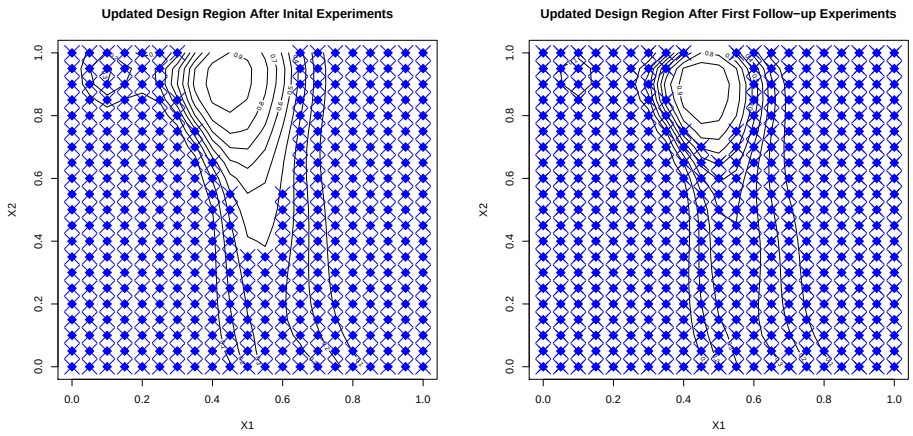

Figure 21: Adaptive design regions and estimated contours of the objective function for the first and second follow-up experiments are shown on the left and right, respectively.

The complete modeling and design processes iterate as in Figure 17 until the maximum allowable runs are used up–for this case, 20 runs. Figure 22 shows the true contour of $g^*(x_1^*, x_2^*)$ next to its estimated contour based on the experimental runs. To save resources, our design plan focuses on only placing design points in regions that potentially contain the optimal value. As a result, the estimated contour is very accurate near the optimal region (around the upper-left quadrant) and only a rough approximation elsewhere. The results of the design process did not suffer from this fact; the estimated optimal input is $(x_1^*, x_2^*) = (.45, .9)$ with an estimated probability of success of 0.9997. Recall, the true optimal input is $(x_1^*, x_2^*) = (.48, .87)$, and the associated success probability is 0.9545. The complete design is shown in Figure 23.

3.7 Conclusion

In this work, we provided a complete methodology for finding the optimal operating condition for high cost, low resource experiments motivated by the LOE approach in Kim [32]. Two novel approaches in local modeling are provided: a very general local calibration/adjustment approach and a local model averaging approach. Equipped

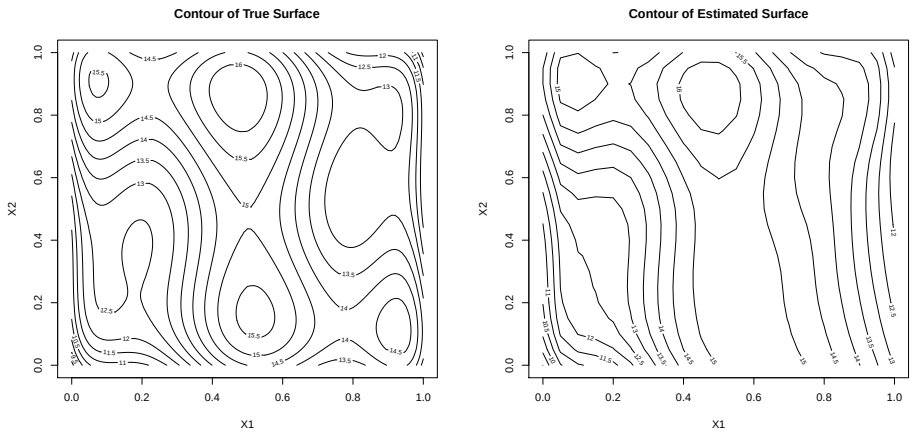

Figure 22: Contours of the true (*left*) and estimated surfaces (*right*) are shown.

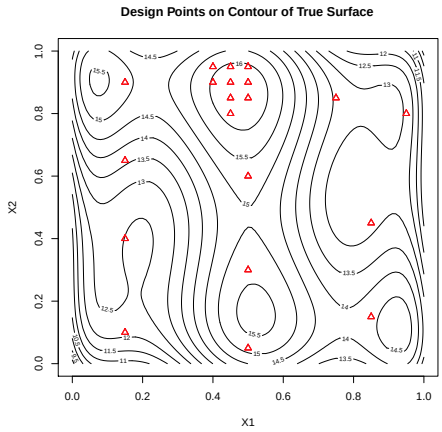

Figure 23: The final experimental design is shown overlaying the true contour.

with a powerful modeling approach, a model-based batch sequential design procedure, referred to as LOE-BSMED, is introduced to balance space-filling properties with the search for the optimal operating condition while adapting the design region. Performance of the LOE-BSMED procedure is studied on a CdSe nanowire growth dataset provided in Dasgupta [17] and a modified six-camel hump optimization function.

We presented new advancements in statistical modeling and experimental design in conjunction to address experimental designs for highly complicated physical systems. It is worth noting that each of these methodologies can stand-alone or in pairs

to address modeling and/or experimental design problems. For example, the design methodology showed good performance without the use of engineering knowledge. In addition, conditional on competing models, the local model averaging methodology can be applied to draw from the strengths of each model for better process characterization or prediction. Similarly, engineering models can be calibrated and adjusted based on the "quality"–in this case, location–of the data. The local calibration/adjustment procedure is very general and non-restrictive, as it is likely to have the ability to be incorporated in many existing and future calibration procedures.

3.8 Future Work

Future work on the topic of *Layers-of-Experiments* (LOE) could be the development of a unified Bayesian framework. As multiple sources of information are combined and models are calibrated, adjusted, and updated over time, a Bayesian framework seems like a natural extension. For example, information from previous rounds of experimentation can be used to set prior distributions for the current round of experimentation. Additionally, Bayesian methods are typically flexible in dealing with sequentially observed data. We discuss some relevant literature and their potential uses and extensions for building a Bayesian LOE framework. In particular, discussions on modeling, design criteria, and subsequent design region choice are provided.

Many advances in modeling data come from multiple sources have focused on using a Bayesian methodology. For instance, in the seminal work of Kennedy and O'Hagan [30], a hierarchical Bayesian framework was proposed for calibrating and adjusting engineering/computer models to physical data. Denoting z_i as the observed data, $\eta(\boldsymbol{x}_i, \boldsymbol{\theta})$ as the engineering/computer model, and $\delta(\boldsymbol{x}_i)$ as the model inadequacy function, they propose the following calibration model:

$$z_i = \rho\eta(\boldsymbol{x}_i, \boldsymbol{\theta}) + \delta(\boldsymbol{x}_i) + e_i,$$

where ρ is an unknown regression parameter and e_i is the observation error for the

i-th observation. Both $\eta(\boldsymbol{x}_i, \boldsymbol{\theta})$ and $\delta(\boldsymbol{x}_i)$ are assumed to be independent Gaussian processes (GP). The model parameters are estimated through a hierarchical Bayesian framework. The flexibility in the GP model could be useful for the local approaches used here, as it should be able to, in some sense, update locally. Many other Bayesian calibration/adjustment methods based off of the work of Kennedy and O'Hagan have been proposed. See Section 3.2.2 for additional references. A further advancement that a Bayesian approach could allow is the incorporation of models of differing accuracies as done in Qian and Wu [42]. Based off the work of Kennedy and O'Hagan, they propose combining *high-accuracy* and *low-accuracy* by using the following adjustment model:

$$y_h(\boldsymbol{x}_i) = \rho(\boldsymbol{x}_i) y_l(\boldsymbol{x}_i) + \delta(\boldsymbol{x}_i) + \epsilon(\boldsymbol{x}_i),$$

where y_h and y_l are the low and high accuracy observations, respectively, ϵ is the observation error, and ρ and δ are GP models. A Bayesian hierarchical framework is used for estimation. Again, the flexibility of the GP models to adapt locally is an advantage in LOE framework, where follow-up experimentation is only on a subregion of the complete design space. Lastly, a different approach to the Bayesian modeling of data from multiple sources is given in Reese *et al.* [44], who propose setting prior distributions based on expert opinion and computer models. In particular, posterior distributions of the parameter values in the expert opinion model are used as the prior distributions for the parameters in the computer model. Similarly, the posterior distributions of the computer model parameters are used as the prior distributions for the physical model parameters. This approach could potentially be modified to fit nicely into the LOE approach. For example, posterior distributions for model parameters obtained after the initial experiment could be used as the prior distributions for the first set of follow-up experimentation and so on. A downfall, however, is the modeling approach in Reese *et al.* assumes the same multiple regression model for all sources of data. If considering the LOE approach as provided in Kim [32],

an advantage is that the regression models can change over time so that non-linear response surfaces can be reasonably approximated by linear models on subregions of the complete design space. Possibly a more flexible modeling procedure than linear regression could be developed and incorporated into this procedure. As a final note on modeling, many of the methodologies above and referenced in Section 3.2.2 are motivated by computer experiments, and as a result, modifications may be needed to better deal with physical experimentation.

A good review of Bayesian experimental design procedures is provided in Chaloner and Verdinelli [12]. The general framework is as follows. Let ξ be a design to be chosen from the design space $\mathcal{X}$, $\boldsymbol{y}$ be the observed sample from $\mathcal{Y}$, d be a decision taken from a set $\mathcal{D}$, and θ be an unknown parameter in Θ. A utility function $U(d, \theta, \xi, \boldsymbol{y})$ is defined to incorporate the objective of the design. Then, the optimal design is chosen to maximize the expected utility; the optimal design is the design ξ^* that maximizes

$$U(\xi^*) = \int_{\mathcal{Y}} \max_{d \in \mathcal{D}} \int_{\Theta} U(d, \theta, \xi, \boldsymbol{y}) p(\theta \mid \boldsymbol{y}, \xi) p(\boldsymbol{y} \mid \xi) d\theta d\boldsymbol{y}.$$

In the LOE framework in Kim [32], a D-optimal design was combined with the space filling designs. The Bayesian counterpart to D-optimal designs is related to the Shannon information. In particular, the design that maximizes the expected gain in Shannon information is given by the utility function

$$U(\xi) = \int \log \left\{ p(\theta \mid \boldsymbol{y}, \xi) \right\} p(\boldsymbol{y}, \theta \mid \xi) d\theta d\boldsymbol{y}.$$

Using Gaussian linear regression models (as in Kim), the utility reduces to

$$U(\xi) = -\frac{k}{2} \log(2\pi) - \frac{k}{2} + \frac{1}{2} \log \det \left\{ \sigma^{-2} (nM(\xi) + R \right\},$$

where k is the number of unknown parameters (σ^2 is considered known), $nM(\xi) = X^T X$, and R is a known $k \times k$ matrix such that the variance-covariance matrix is $\sigma^2 R^{-1}$. This utility function is known as the Bayesian D-optimality criteria. As in Kim, a Bayesian D-optimality criteria could be combined with a space-filling design

criteria by restricting the distance between any two points. Dealing with non-linear models, such as some engineering models, can complicate the analysis. See Chaloner and Verdinelli [12] for more details. Verdinelli and Kadane [47] considered another utility function to balance the Shannon information gain and the goal of maximizing the total output, two objectives that are useful for the LOE framework discussed here and in Kim. Their utility function is defined as

$$U(\xi) = \int \left[\rho \boldsymbol{y}^T \mathbf{1} + \beta \log p(\theta \mid \boldsymbol{y}, \xi)\right] p(\boldsymbol{y}, \theta \mid \xi) d\boldsymbol{y} d\theta,$$

where ρ and β are chosen to balance the two objectives. In the LOE framework, ρ and β can be adaptive parameters. This utility function could serve as a starting point for proposing new, adaptive experimental designs for the Bayesian LOE framework.

Updating the design region over time can be addressed in a Bayesian approach or an alternative approach as discussed in this work, in Kim [32], or Mandal, Ranjan, and Wu [38]. These procedures include setting a threshold based on quantiles, using a numerical procedure to build a region around the "best" observed point, and finding level sets, respectively. However, these approaches are more numerical in nature. To conclude a Bayesian LOE framework, a Bayesian decision theoretic framework could provide a nice addition. Potentially, a framework for deciding subsequent design regions could be developed by incorporating a formal treatment of the risk in selecting among differing follow-up design regions. A comprehensive introduction to Bayesian decision theory can be found in Berger [7].

APPENDIX A

SUPPLEMENTAL MATERIAL FOR INITIAL EXPERIMENTAL DESIGN

A.1 Engineering Model Data

Let $f(\mathbf{X};\beta)$ be an engineering model describing the system of interest. From a statistical point of view, f only provides part of the information about the system; the mean of the model is specified with no mention of variability around the mean, rendering statistical inference, such as creating a prediction interval, impossible. Therefore, conjectured data will be provided in the same vector form as the expert opinion data was provided. For example, based on the work of Reese *et al* ([44]), an expert will specify $(\sigma^2_{e_i}, \xi_i, q_{\xi_i})$, or equivalently, $(\log(\sigma^2_{e_i}), \xi_i, \log(q_{\xi_i}))$,where $\sigma^2_{e_i}$ is the expected mean of the variance model at design point $\boldsymbol{x}_i$ and ξ_i and q_{ξ_i} are such that $P(\sigma^2_{e_i} \leq \xi_i) = q_{\xi_i}$. Then a log-variance model is:

$$\begin{aligned}
\log \boldsymbol{\sigma}^2_e &\sim N(\mathbf{X}_{\sigma_e}\boldsymbol{\beta}_{\sigma_e} + \boldsymbol{\delta}_{\sigma_e}, \sigma^2_{\sigma_e}\boldsymbol{\Sigma}_{\sigma_e}) \\
\boldsymbol{\beta}_{\sigma_e} \mid \sigma^2_{\sigma_e} &\sim N(\boldsymbol{\mu}_{\sigma_e}, \sigma^2_{\sigma_e}\boldsymbol{C}_{\sigma_e}) \\
\sigma^2_{\sigma_e} &\sim IG(\alpha_{\sigma_e}, \gamma_{\sigma_e}) \\
\boldsymbol{\delta}_{\sigma_e} &\sim N(\theta_{\delta_{\sigma_e}}, \xi^2_{\delta_{\sigma_e}}\boldsymbol{I}),
\end{aligned}$$

where $\boldsymbol{\delta}_{\sigma_e}$ is an unknown bias term that arises do to the lack of knowledge of the true variance model. The prior distributions for the hyperparameters, as well as the form of the covariance matrices $\boldsymbol{\Sigma}_{\sigma_e}$ and $\boldsymbol{C}_{\sigma_e}$, can be set in accord with a researchers beliefs.

After the variance model is specified, an approximate model for the engineering model can be defined as: $Y_{eng}(\boldsymbol{x}_i) \overset{ind}{\sim} N(f(\boldsymbol{x}_i;\boldsymbol{\beta}), \sigma^2_e(\boldsymbol{x}_i))$.

A.2 Expert Opinion and Details for Modifying an Expert Opinion Method for Modeling Past Experimental Data Point by Point.

There are many methods for modeling and eliciting expert opinion data ([22]). A recent method which interested readers can refer to in order to incorporate expert opinion models is provided in Reese *et al.* ([44]). In this section, however, we provide details on how the method for modeling expert opinion data given in Reese *et al.* can be modified to model past experimental data when each data point is conjectured individually instead of by a complete model shift. When conjecturing data using past experimental data as a reference point, the true mean response is not conjectured, but instead, the expected difference between the past and future responses is conjectured. Thus, we need to take into account that the lack of knowledge is modeled through $\boldsymbol{\lambda}$ now instead of $\boldsymbol{Y}$. Another distinction between modeling expert opinion in the style of Reese *et al.* and past experimental data is that the true design point is subject to bias, adding an additional source of variability which needs to be considered in the modeling process.

Suppose there are n_p responses from past experiments. For each response, we will have a data vector of the form $(\lambda_i, \xi_i, q_{\xi_i}, \boldsymbol{x}_i, m_{\lambda,i})$, where λ_i is the conjectured difference between the response of the new and past experiments, $\boldsymbol{x}_i$ is the conjectured design point in the new experimental region, $m_{\lambda,i}$ is a relative worth (see Reese *et al.* ([44])), and ξ_i and q_{ξ_i} are a coverage probability and a quantile associated with that probability, respectively, such that at the i-th experimental point, it is believed that $P(\lambda_i \leq q_{\xi_i}) = \xi_i$. We use the following hierarchical model:

$$\begin{aligned}
\boldsymbol{Y}_p &\sim N(\boldsymbol{X}_p\boldsymbol{\beta} + \boldsymbol{\delta}_p, \sigma^2\boldsymbol{\Sigma}_p) \\
\boldsymbol{\beta} \mid \sigma^2 &\sim N(\boldsymbol{\mu}_p, \boldsymbol{\sigma}^2\mathbf{C}_p) \\
\sigma^2 &\sim IG(\alpha_p, \gamma_p) \\
\boldsymbol{X}_{p,i} &\overset{independent}{\sim} N(\boldsymbol{\mu}_{\boldsymbol{X}_i}, \sigma^2_{\boldsymbol{X}}\Sigma_{\boldsymbol{X}_i}),
\end{aligned}$$

where $\boldsymbol{\Sigma}_p = \text{diag}\,[1/k_i]_{i=1}^{n_p}$, $\boldsymbol{C}_p = \text{diag}\,[1/c_i]_{i=1}^{\dim(\boldsymbol{\beta})}$ and $\boldsymbol{\Sigma}_{\boldsymbol{X}_i} = [1/\chi_i]$. Let $\boldsymbol{\eta}_p$ be the set of all hyperparameters that will be included as hyperparameters. By use of conjugate priors, it follows that

$$\pi(\boldsymbol{\beta}|\sigma^2, \boldsymbol{\eta}_p, \boldsymbol{y}_p, \boldsymbol{X}_p) \;\sim\; N\left((\boldsymbol{X}_p'\boldsymbol{\Sigma}_p^{-1}\boldsymbol{X}_p + \boldsymbol{C}_p^{-1})^{-1}\boldsymbol{z}_p, \sigma^2(\boldsymbol{X}_p'\boldsymbol{\Sigma}_p^{-1}\boldsymbol{X}_p + \boldsymbol{C}_p^{-1})^{-1}\right),$$

and

$$\begin{aligned}\pi(\sigma^2 \mid \boldsymbol{\eta}_p, \boldsymbol{y}_p, \boldsymbol{X}_p) \;&\sim\; IG\Big(\alpha_p + \frac{n_p}{2}, \gamma_p + .5\big[(\boldsymbol{y}_p - \boldsymbol{\delta}_p)'\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{y}_p - \boldsymbol{\delta}_p)\\ &+\; \boldsymbol{\mu}_p'\boldsymbol{C}_p^{-1}\boldsymbol{\mu}_p - \boldsymbol{z}'(\boldsymbol{X}_p'\boldsymbol{\Sigma}_p^{-1}\boldsymbol{X}_p + \boldsymbol{C}_p^{-1})^{-1}\boldsymbol{z}_p\big]\Big),\end{aligned}$$

where $\boldsymbol{z}_p = \boldsymbol{X}_p'\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\lambda} - (\boldsymbol{\delta}_p - \tilde{\boldsymbol{y}})) + \boldsymbol{C}_p^{-1}\boldsymbol{\mu}_p$.

To incorporate the elicited vector of values in the modeling procedure, suppose we have $m_{\lambda,i}$ hypothetical repetitions of λ_i were elicited at each design point with the mean and variance equal to the elicited values. Then

$$\begin{aligned}&\big(\boldsymbol{X}_p'\boldsymbol{\Sigma}_p^{-1}(\tilde{\boldsymbol{y}} - (\boldsymbol{\delta}_p - \boldsymbol{\lambda}))\big)_j\\ =\;& k_{p_1}x_{1,j}\Big(\sum_{n=1}^{m_{\lambda,i}}(\lambda_{j,n} - (\delta_{p,1} - \tilde{y}_1))\Big) + \cdots + k_{p_{n_p}}x_{n_p,j}\Big(\sum_{n=1}^{m_{\lambda,n_p}}(\lambda_{j,n} - (\delta_{p,n_p} - \tilde{y}_{n_p}))\Big)\\ =\;& k_{p_1}m_{\lambda,1}x_{1,j}(\lambda_1 - (\delta_{p,1} - \tilde{y}_1)) + \cdots + k_{p_{n_p}}m_{\lambda,n_p}x_{n_p,j}(\lambda_{n_p} - (\delta_{p,n_p} - \tilde{y}_{n_p})).\end{aligned}$$

Let $s_{\lambda,i}^2 = (\lambda_i - q_{\xi_{\lambda,i}})^2/Z_{\xi_{\lambda,i}}^2$ be the elicited variance approximation for λ_i at design point $\boldsymbol{x}_i$. Then

$$\begin{aligned}(\boldsymbol{y}_p - \boldsymbol{\delta}_p)'\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{y}_p - \boldsymbol{\delta}_p) \;&=\; (\boldsymbol{\lambda} - (\boldsymbol{\delta}_p - \tilde{\boldsymbol{y}}))'\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{\lambda} - (\boldsymbol{\delta}_p - \tilde{\boldsymbol{y}}))\\ &=\; k_{p_1}\Big(\sum_{n=1}^{m_{\lambda,1}}(\lambda_{j,n} - (\delta_{p,1} - \tilde{y}_1))^2\Big) + \cdots\\ &+\; k_{p_{n_p}}\Big(\sum_{n=1}^{m_{\lambda,n_p}}(\lambda_{j,n} - (\delta_{p,n_p} - \tilde{y}_{n_p}))^2\Big)\\ &=\; \sum_{i=1}^{n_p}k_{p_i}m_{\lambda,i}\big(s_{\lambda_i}^2 + (\lambda_i - (\delta_{p,i} - \tilde{y}_i))^2\big),\end{aligned}$$

which follows from the identity, $Var(\Lambda) = E[\Lambda^2] - E[\Lambda]^2$.

Lastly, a similar argument yields

$$\left(\boldsymbol{X}_p'\boldsymbol{\Sigma}_p^{-1}\boldsymbol{X}_p\right)_{ij} = \sum_{n=1}^{n_p} k_{p_n} m_{\lambda,n} x_{n,i} x_{n,j}.$$

With these equivalences, the conditional distributions can now be expressed in terms of the elicited values of the mean difference in the past and future data. If desired, more prior distributions can be placed on the hyperparameters as done in Reese *et al.* ([44]). Parameter estimation and prediction can be done by Gibbs sampling.

A.2.1 Proof of Proposition 2.1.

(P1) If $\tilde{p}(\mathbf{x_i}) = m$, then $\operatorname{sgn}(m - \tilde{p}(\mathbf{x_i})) = 0$, implying $f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i)) = \tilde{p}(\mathbf{x_i})$. Since $\tilde{p}(\mathbf{x}_i)$ is a rescaled probability, still between 0 and 1, $0 \leq f \leq 1$.

If $\tilde{p}(\mathbf{x_i}) > m$, then $\operatorname{sgn}(m - \tilde{p}(\mathbf{x_i})) = -1$ and $\left|\frac{\tilde{p}(\mathbf{x_i})}{m} - 1\right| = \left(\frac{\tilde{p}(\mathbf{x_i})}{m} - 1\right)$. Therefore,

$$\begin{aligned}
\operatorname{sgn}(m - \tilde{p}(\mathbf{x_i}))(m - m\tilde{u}(\mathbf{x}_i))\left|\frac{\tilde{p}(\mathbf{x_i})}{m} - 1\right| &= -(m - m\tilde{u}(\mathbf{x}_i))\left(\frac{\tilde{p}(\mathbf{x_i})}{m} - 1\right) \\
&= -(\tilde{p}(\mathbf{x_i}) - m)(1 - \tilde{u}(\mathbf{x}_i)) \\
&\geq -(\tilde{p}(\mathbf{x_i}) - m) \\
&= (m - \tilde{p}(\mathbf{x_i})).
\end{aligned}$$

Therefore, $f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i)) = \tilde{p}(\mathbf{x_i}) + \operatorname{sgn}(m - \tilde{p}(\mathbf{x_i}))(m - m\tilde{u}(\mathbf{x}_i))\left|\frac{\tilde{p}(\mathbf{x_i})}{m} - 1\right| \geq m \geq 0$. Lastly, since $(m - m\tilde{u}(\mathbf{x}_i))\left|\frac{\tilde{p}(\mathbf{x_i})}{m} - 1\right| \geq 0$, $f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i)) \leq \tilde{p}(\mathbf{x_i}) \leq 1$. Thus, $0 \leq f \leq 1$. By an analogous argument, if $\tilde{p}(\mathbf{x_i}) < m$, then $0 \leq f \leq 1$.

(P2) Suppose $\tilde{p}(\mathbf{x_i}) > m$. Then $\operatorname{sgn}(m - \tilde{p}(\mathbf{x_i})) = -1$. Moreover, $\left|\frac{\tilde{p}(\mathbf{x_i})}{m} - 1\right| = \frac{\tilde{p}(\mathbf{x_i})}{m} - 1$. In this case, we assume we have sufficient trust in the model, i.e., $\tilde{u}(\boldsymbol{x}_i) > \frac{1}{2}$,

implying $-(m - m\tilde{u}(\boldsymbol{x}_i)) > -\left(\frac{m}{2}\right)$. Therefore,

$$\begin{aligned} f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i)) &= \tilde{p}(\mathbf{x_i}) - (m - m\tilde{u}(\boldsymbol{x}_i)) \left| \frac{\tilde{p}(\mathbf{x_i})}{m} - 1 \right| \\ &> \tilde{p}(\mathbf{x_i}) - \frac{m}{2} \left(\frac{\tilde{p}(\mathbf{x_i})}{m} - 1 \right) \\ &= \frac{\tilde{p}(\mathbf{x_i})}{2} + \frac{m}{2} \\ &> m. \end{aligned}$$

If $\tilde{p}(\mathbf{x_i}) < m$ and in the cases with insufficient trust, the proofs follow in a similar manner.

(P3) If $\tilde{p}(\mathbf{x_i}) > m$. Then, $f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i))$ is a increasing function of $\tilde{u}(\mathbf{x}_i)$ with a minimum value of m and a maximum value of $\tilde{p}(\mathbf{x_i})$ occurring when $\tilde{u}(\mathbf{x}_i) = 0$ and $\tilde{u}(\mathbf{x}_i) = 1$, respectively. If $\tilde{p}(\mathbf{x_i}) < m$, then $f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i))$ is a decreasing function of $\tilde{u}(\mathbf{x}_i)$ with a maximum value of m and a minimum value of $\tilde{p}(\mathbf{x_i})$ occurring when $\tilde{u}(\mathbf{x}_i) = 0$ and $\tilde{u}(\mathbf{x}_i) = 1$, respectively. Lastly, if $\tilde{p}(\mathbf{x_i}) = m$, then $f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i)) = m$ whenever $\tilde{u}(\mathbf{x}_i) = 0$. Thus, as $\tilde{u}(\mathbf{x}_i)$ increases to 1, $f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i))$ reflects the information of the prior model, and when $\tilde{u}(\mathbf{x}_i)$ decreases to 0, $f(p(\mathbf{x}_i), \mathbf{u}(\mathbf{x}_i))$ is normalized to m. □

APPENDIX B

SUPPLEMENTAL MATERIAL FOR MODEL ROBUST PROCESS OPTIMIZATION

B.1 Proof of Proposition 3.1.

For a fixed design point, $\tilde{\boldsymbol{x}}$, consider the mixture model

$$f(\boldsymbol{y} \mid \Theta) \;=\; \sum_{j=1}^{m} w_j f_j(\boldsymbol{y} \mid \theta_j),$$

where the weight vector $\boldsymbol{w} = \{w_j\}_{j=1}^{m}$ implicitly depends on $\tilde{\boldsymbol{x}}$. Let $\boldsymbol{k} = \{k_i\}_{i=1}^{n}$ denote the vector of kernel weights at design point $\tilde{\boldsymbol{x}}$. The local log-likelihood can be written as

$$\begin{aligned}
\log \mathcal{L}(\Theta \mid \boldsymbol{y}) &= C + \sum_{i=1}^{n} k_i \log f(y_i \mid \Theta) \\
&= C + \sum_{i=1}^{n} k_i \log \sum_{j=1}^{m} w_j f_j(y_i \mid \theta_j),
\end{aligned}$$

where C is the log of the normalizing constant for the local log-likelihood. Let z_i be a latent (unobserved) variable which denotes the model which generated observation i. For example, if observation i resulted from the k-th model, $k = 1, \ldots, m$, then $z_i = k$. Let $\boldsymbol{z} = \{z_j\}_{j=1}^{m}$. Conditional on $\boldsymbol{z}$, the local log-likelihood becomes

$$\begin{aligned}
\log \mathcal{L}(\Theta \mid \boldsymbol{y}, \boldsymbol{z}) &= C + \sum_{i=1}^{n} k_i \log f(y_i \mid \Theta, z_i) f(z_i \mid \Theta) \\
&= C + \sum_{i=1}^{n} k_i \log w_{z_i} f_{z_i}(y_i \mid \Theta).
\end{aligned}$$

Moreover, by Bayes's rule, the distribution of $\boldsymbol{z}$ conditional on the responses, $\boldsymbol{y}$, and the vector of parameters, Θ, can be expressed as

$$\begin{aligned} f(\boldsymbol{z} \mid \boldsymbol{y}, \Theta) &= \prod_{i=1}^{n} f(z_i \mid y_i, \theta_{z_i}) \\ &= \prod_{i=1}^{n} \frac{f(z_i|\theta_{z_i}) f(y_i \mid z_i, \theta_{z_i})}{f(y_i \mid \theta_{z_i})} \\ &= \prod_{i=1}^{n} \frac{w_{z_i} f_{z_i}(y_i \mid \theta_{z_i})}{\sum_{j=1}^{M} w_j f_j(y_i \mid \theta_j)} \end{aligned}$$

Let $\Theta^{(t)}$ be the estimates of the parameter Θ after the t-th iteration of the algorithm and $\mathcal{Z}$ be the sample space of $\boldsymbol{z}$. We then have

$$\begin{aligned} Q(\Theta, \Theta^{(t-1)}) &\equiv E_{\boldsymbol{z}} \left[\log f(\boldsymbol{y}, \boldsymbol{z} \mid \Theta) \mid \boldsymbol{y}, \Theta^{(t-1)} \right] \\ &= \sum_{\boldsymbol{z} \in \mathcal{Z}} \log(\mathcal{L}(\Theta \mid \boldsymbol{y}, \boldsymbol{z})) f(\boldsymbol{z} \mid \boldsymbol{y}, \Theta^{(t-1)}) \\ &= C + \sum_{\boldsymbol{z} \in \mathcal{Z}} \sum_{i=1}^{n} k_i \log(w_{z_i} f_{z_i}(y_i \mid \Theta)) \prod_{j=1}^{n} f(z_j \mid y_j, \Theta^{(t-1)}) \\ &= C + \sum_{z_1=1}^{m} \cdots \sum_{z_n=1}^{m} \sum_{i=1}^{n} k_i \log(w_{z_i} f_{z_i}(y_i \mid \Theta)) \prod_{j=1}^{n} f(z_j \mid y_j, \Theta^{(t-1)}). \end{aligned}$$

Let $\delta_{l,z_i} = 1$ if $z_i = l$ and 0 otherwise, where $l = 1, \ldots, m$, and let $f(l \mid y_j, \Theta^{(t-1)}) = P(z_i = l \mid y_i, \Theta^{(t-1)})$. Then

$$\begin{aligned} Q(\Theta, \Theta^{(t-1)}) &= C + \sum_{z_1=1}^{m} \cdots \sum_{z_n=1}^{m} \sum_{i=1}^{n} \sum_{l=1}^{m} \delta_{l,z_i} k_i \log(w_l f_l(y_i \mid \Theta)) \prod_{j=1}^{n} f(z_j \mid y_j, \Theta^{(t-1)}) \\ &= C + \sum_{l=1}^{m} \sum_{i=1}^{n} k_i \log(w_l f_l(y_i \mid \Theta)) \sum_{z_1=1}^{m} \cdots \sum_{z_n=1}^{m} \delta_{l,z_i} \prod_{j=1}^{n} f(z_j \mid y_j, \Theta^{(t-1)}) \end{aligned}$$

Notice that

$$\begin{aligned}
\sum_{z_1=1}^{m}\cdots\sum_{z_n=1}^{m} \quad \delta_{l,z_i} \quad & \prod_{i=1}^{n} f(z_i \mid y_i, \Theta^{(t-1)}) = \\
& = \left(\sum_{z_1=1}^{m}\cdots\sum_{z_{i-1}=1}^{m}\sum_{z_{i+1}=1}^{m}\cdots\sum_{z_n=1}^{m}\prod_{j=1,j\neq i}^{n} f(z_j \mid y_j, \Theta^{(t-1)})\right) \\
& \times \quad f(l \mid y_i, \Theta^{(t-1)}) \\
& = \prod_{j=1,j\neq i}^{n}\left(\sum_{z_j=1}^{m} f(z_j \mid y_j, \Theta^{(t-1)})\right) f(l \mid y_i, \Theta^{(t-1)}) \\
& = f(l \mid y_i, \Theta^{(t-1)}),
\end{aligned}$$

where we use the fact that $\sum_{z_j=1}^{m} f(z_j \mid y_j, \Theta^{(t-1)})$. Then, $Q(\Theta, \Theta^{(t-1)})$ can then be expressed as

$$\begin{aligned}
Q(\Theta, \Theta^{(t-1)}) &= C + \sum_{l=1}^{m}\sum_{i=1}^{n} k_i \log(w_l f_l(y_i \mid \Theta)) f(l \mid y_i, \Theta^{(t-1)}) \\
&= C + \sum_{l=1}^{m}\sum_{i=1}^{n} k_i \log(w_l) f(l \mid y_i, \Theta^{(t-1)}) \\
&+ \sum_{l=1}^{m}\sum_{i=1}^{n} k_i \log(f_l(y_i \mid \Theta)) f(l \mid y_i, \Theta^{(t-1)}).
\end{aligned}$$

At time t, the weights, w_l, $l = 1, \ldots, M$, are found as the arguments maximizing $Q(\Theta, \Theta^{(t-1)})$ subject to $\sum_{l=1}^{m} w_l = 1$. Employing Lagrange multipliers, this results in solving

$$\frac{d}{dw_l}\left[\sum_{l=1}^{m}\sum_{i=1}^{n} k_i \log(w_l) f(l \mid y_i, \Theta^{(t-1)}) + \lambda\left(\sum_{l=1}^{m} w_l - 1\right)\right] = 0.$$

Taking the partial derivative results in the finding

$$\sum_{i=1}^{n} \frac{k_i}{w_l} f(l \mid y_i, \Theta^{(t-1)}) = -\lambda,$$

implying thats

$$w_l = -\frac{1}{\lambda}\sum_{i=1}^{n} k_i f(l \mid y_i, \Theta^{(t-1)})$$

Summing up both sides over $l = 1, \ldots, m$, we get

$$1 = -\frac{1}{\lambda} \sum_{i=1}^{n} k_i \sum_{l=1}^{m} f(l \mid y_i, \Theta^{(t-1)}) = -\frac{1}{\lambda} \sum_{i=1}^{n} k_i,$$

giving

$$\lambda = -\sum_{i=1}^{n} k_i.$$

The estimated weights are therefore given by

$$\hat{w}_l = \frac{\sum_{i=1}^{n} k_i f(l \mid y_i, \Theta^{(t-1)})}{\sum_{i=1}^{n} k_i},$$

where

$$f(l \mid y_i, \Theta^{(t-1)}) = \frac{w_l f_l(y_i \mid \theta_l)}{\sum_{j=1}^{m} w_j f_j(y_i \mid \theta_j)}.$$

Lastly, since $Q(\Theta, \Theta^{(t-1)})$ is decomposed into two sums, the first of which is independent of Θ, the model parameters are estimated as

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta \setminus \boldsymbol{w}} \sum_{l=1}^{m} \sum_{i=1}^{n} k_i \log(f_l(y_i \mid \Theta)) f(l \mid y_i, \Theta^{(t-1)}).$$

□

REFERENCES

[1] Adiga, N., *Contributions to Variable Selection for Mean Modeling and Variance Modeling in Computer Experiments*. PhD thesis, Georgia Institute of Technology, 2012.

[2] Atkinson, A., "Dt-optimum designs for model discrimination and parameter estimation," *Journal for Statistical Planning and Inference*, vol. 138, pp. 56–64, 2008.

[3] Ba, S. and Joseph, V. R., "Multi-layer designs for computer experiments," *Journal of the American Statistical Association*, vol. 106, pp. 1139–1149, 2012.

[4] Barbe, P. and Bertail, P., *The Weighted Bootstrap*. Springer, 1995.

[5] Basumallick, A., Das, G., and Mukherjee, S., "Design of experiments for synthesizing in situ Ni-SiO2 and CO-SiO2 nanocomposites by non-isothermal reduction treatment," *Nanotechnology*, vol. 14, pp. 903–906, 2003.

[6] Bayarri, M., Berger, J., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C., and Tu, J., "A framework for validation of computer models," *Technometrics*, vol. 49, pp. 138–154, 2007.

[7] Berger, J., *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, Inc., 1985.

[8] Box, G. and Draper, N., *Response Surfaces, Mixtures, and Ridge Analyses: Empirical Model-building and Response Surfaces*. New York: Wiley, 2007.

[9] Casciato, M., Kim, S., Lu, J. C., Hess, D., and Grover, M., "Optimization of a carbon dioxide-assisted nanoparticle deposition process using sequential experimental design with adaptive design space," *Industrial & Engineering Chemistry Research*, vol. 51, pp. 4363–4370, 2012.

[10] Casciato, M., Kim, S., Lu, J. C., Hess, D., and Grover, M., "Optimization of carbon dioxide-assisted nanoparticle deposition process with uncertain design space," *Computer-Aided Chemical Engineering*, vol. In Press, 2012.

[11] Casciato, M., Levitin, G., Hess, D., and Grover, M., "Controlling the properties of silver nanoparticles deposited on surfaces using supercritical carbon dioxide for surface-enhanced raman spectroscopy," *Journal of Nanoparticle Research*, vol. 14, pp. 836–850, 2012.

[12] Chaloner, K. and Verdinelli, I., "Bayesian experimental design: A review," *Statistical Science*, vol. 10, pp. 273–304, 1995.

[13] Chang, C.-J., *Statistical and Engineering Methods for Model Enhancement*. PhD thesis, Georgia Institute of Technology, 2012.

[14] Chang, K., Hong, L., and Wan, H., "Stochastic trust region gradient-free method (strong)-a new response-surface-based algorithm in simulation optimization," in *Proceedings of the 2007 Winter Simulation Conference*.

[15] Chang, K. and Wan, H., "Stochastic trust region response surface convergent method for generally-distributed response surface," in *Proceedings of the 2009 Winter Simulation Conference*.

[16] Craig, P., Goldstein, M., Rougier, J., and Seheult, A., "Bayesian forecasting for complex systems using computer simulators," *Journal of the American Statistical Association*, vol. 96, pp. 717–729, 2001.

[17] Dasgupta, T., *Robust Parameter Design for Automatically Controlled Systems and Nanostructure Synthesis*. PhD thesis, Georgia Institute of Technology, 2007.

[18] Dasgupta, T., Ma, C., Joseph, V. R., Wang, Z. L., and Wu, C. F. J., "Statistical modeling and analysis for robust synthesis of nanostructures," *Journal of the American Statistical Association*, vol. 103, pp. 594–603, 2008.

[19] Fedorov, V., *Theory of Optimal Design*. Academic, 1972.

[20] Fedorov, V. and Hackl, P., *Model-oriented Design of Experiments*. Springer-Verlag, 1997.

[21] Franssila, S., *Introduction of Microfabrication*. Wiley, 2 ed., 2010.

[22] Garthwaite, P., Kadane, J., and O'Hagan, A., "Statistical methods for eliciting probability distributions," *Journal of the American Statistical Association*, vol. 100, pp. 680–700, 2005.

[23] Goel, T., Haftka, R., Shyy, W., and Watson, L., "Pitfalls of using a single criterion for selecting experimental designs," *International Journal for Numerical Methods in Engineering*, vol. 75, pp. 127–155, 2008.

[24] Hedges, L. and Olkin, I., *Statistical Methods for Meta Analysis*. New York: Wiley, 1987.

[25] Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C., "Bayesian model averaging: A tutorial," *Statistical Science*, vol. 14, pp. 382–417, 1999.

[26] Jones, D., Schonlau, M., and Welch, W., "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, pp. 455–492, 1998.

[27] Joseph, V. R., Dasgupta, T., and Wu, C. F. J., "Minimum energy designs: From nanostructure synthesis to sequential optimization." Working Paper, 2012.

[28] Joseph, V. R. and Hung, Y., "Orthogonal-maximin latin hypercube designs," *Statistica Sinica*, vol. 18, pp. 171–186, 2008.

[29] Joseph, V. R. and Melkote, S. N., "Statistical adjustments to engineering models," *Journal of Quality Technology*, vol. 41, pp. 362–375, 2009.

[30] Kennedy, M. and O'Hagan, A., "Bayesian calibration for computer models," *Journal of the Royal Statistical Society, Ser. B*, vol. 63, pp. 425–464, 2001.

[31] Kessels, R., Jones, B., Goos, P., and Vanderbrook, M., "An effiecient algorithm for constructing bayesian optimal choice designs." Research Report KBI 0616, Department of Decision Sciences and Information Management, Katholieke Universiteit Leuven, 2006.

[32] Kim, S., *Experimental Design Method for Nano-fabrication Processes*. PhD thesis, Georgia Institute of Technology, 2011.

[33] Kleiber, W., Raftery, A., Baars, J., Gneiting, T., Mass, C., and Grimit, E., "Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local bayesian model averaging," *Monthly Weather Review*, vol. 139, pp. 2630–2649, 2011.

[34] Lindgren, L.-E., Alberg, H., and Domkin, K., "Constitutive modeling and parameter optimization," in *7th International Conference on Computational Plasticity*.

[35] Loader, C., *Local Regression and Likelihood*. Springer, 1999.

[36] Loeppky, J., Sacks, J., and Welch, W., "Choosing the sample size of a computer experiment: A practical guide," *Technometrics*, vol. 51, pp. 366–376, 2009.

[37] Lu, J. C., Jeng, S., and Wang, K., "A review of statistical methods for quality improvement and control in nanotechnology," *Journal of Quality Technology*, vol. 41, pp. 148–164, 2009.

[38] Mandal, A., Ranjan, P., and Wu, C. F. J., "G-SELC: Optimization of sequential elimination of level combinations using genetic algorithms and gaussian processes," *Annals of Applied Statistics*, vol. 3, pp. 398–421, 2009.

[39] Mandal, A., Wu, C. F. J., and Johnson, K., "SELC: Sequential elimination of level combinations by means of modified genetic algorithms," *Technometrics*, vol. 48, pp. 273–283, 2006.

[40] Myers, R., Montgomery, D., Vining, G., and Robinson, T., *Generalized Linear Models with Applications in Engineering and the Sciences*. Wiley, 2 ed., 2010.

[41] Peña, D. and Redondas, D., "Bayesian curve estimation by model averaging," *Computational Statistics & and Data Analysis*, vol. 50, pp. 688–709, 2006.

[42] Qian, Z. and Wu, C. F. J., "Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments," *Technometrics*, vol. 50, pp. 192–204, 2005.

[43] Rajagopal, R. and del Castillo, E., "Model-robust process optimization using bayesian model averaging," *Technometrics*, vol. 47, pp. 152–163, 2005.

[44] Reese, C., Wilson, A., Hamada, M., Martz, H., and Ryan, K., "Integrated analysis of computer and physical experiments," *Technometrics*, vol. 46, pp. 153–164, 2004.

[45] Santner, T., Williams, B., and Notz, W., *The Design and Analysis of Computer Experiments*. Springer-Verlag, 2003.

[46] Vastola, J., Lu, J. C., Casciato, M., Hess, D., and Grover, M., "A framework for initial experimental design under competing prior knowledge." Working Paper, 2012.

[47] Verdinelli, I. and Kadane, J., "Bayesian designs for maximizing information and outcome," *Journal of the American Statistical Association*, vol. 87, pp. 510–515, 1992.

[48] Wang, S., Chen, W., and Tsui, K., "Bayesian validation of computer models," *Technometrics*, vol. 51, pp. 439–451, 2009.

[49] Wissmann, P. and Grover, M., "A new approach for batch process optimization using experimental design," *AIChE Journal*, vol. 55, pp. 342–353, 2009.

[50] Wissmann, P. and Grover, M., "Optimization of a chemical vapor deposition process using sequential experimental design," *Industrial & Engineering Chemistry Research*, vol. 49, pp. 5694–5701, 2010.

[51] Woo, H., *Multiscale fractality with application and statistical modeling and estimation for computer experiment of nano-particle fabrication*. PhD thesis, Georgia Institute of Technology, 2012.

[52] Wu, C. F. J. and Hamada, M., *Experiments: Planning, Analysis, and Optimization*. New York: Wiley, 2009.

[53] Wu, C. F. J., Mao, S., and Ma, F., "SEL: A search method based on orthogonal arrays," in *Statistical Design and Analysis of Industrial Experiments* (Ghosh, S., ed.), pp. 279–310, New York: Marcel Dekker Inc., 1990.

[54] Xiong, Y., Chen, W., Tsui, K.-L., and Apley, D., "A better understanding of model updating strategies in validating engineering models," *Computational Methods in Applied Mechanical Engineering*, vol. 198, pp. 1327–1337, 2009.

[55] Xu, S., Adiga, N., Ba, S., Dasgupta, T., Wu, C. F. J., and Wang, Z. L., "Optimizing and improving the growth quality of zno nanowire arrays guided by statistical design of experiments," *American Chemical Society, Nano*, vol. 3, pp. 1803–1812, 2009.

[56] Ye, X., Lin, Y., Wang, C., Engelhard, M., Wang, Y., and Wai, C., "Supercritical fluid synthesis and characterization of catalytic metal nanoparticles on carbon nantotubes," *Journal of Materials Chemistry*, vol. 14, pp. 908–913, 2004.

[57] Zeckhauser, R., "Combining overlapping information," *Journal of the American Statistical Association*, vol. 66, pp. 91–92, 1971.

[58] Zhao, H., Jin, R., Wu, S., and Shi, J., "PDE-constrained gaussian process model on material removal rate of wire saw slicing process," *Journal of Manufacturing Science and Engineering*, vol. 133, 2011.

[59] Zong, Y. and Watkins, J., "Deposition of copper by the H_2-assisted reduction of $Cu(tmod)_2$ in supercritical carbon dioxide: Kinetics and reaction mechanism," *Chemistry of Materials*, vol. 17, pp. 560–565, 2005.

VITA

Justin Vastola was born in Chicago, Illinois. In May 2007, he received a B.S. in Mathematics and a B.S. in Statistics from the University of Georgia, Athens, Georgia. Afterwards, he joined the School of Industrial and Systems Engineering at the Georgia Institute of Technology as a doctoral student in August of 2007. He was awarded an M.S. in Operations Research in 2010 from the Georgia Institute of Technology and spent the following two summers as a research fellow at The Wharton School at the University of Pennsylvania, Philadelphia, Pennsylvania. In December 2012, he completed his Ph.D. in Industrial Engineering, Statistics Specialization.