



Analysis, Modeling, and Optimal Experimental Design under Uncertainty: From Carbon Nano-Structures to 3D Printing

Citation

Sosina, Sobambo. 2016. Analysis, Modeling, and Optimal Experimental Design under Uncertainty: From Carbon Nano-Structures to 3D Printing. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493439>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

Analysis, Modeling, and Optimal Experimental Design under Uncertainty: From Carbon Nano-Structures to 3D Printing

A DISSERTATION PRESENTED
BY
SOBAMBO OLUMUYIWA SOSINA
TO
THE DEPARTMENT OF STATISTICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
STATISTICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
FEBRUARY 2016

©2016 – SOBAMBO OLUMUYIWA SOSINA
ALL RIGHTS RESERVED.

Dissertation Advisors: Professor Tirthankar Dasgupta
Professor Joseph K. Blitzstein

Sobambo Olumuyiwa Sosina

Analysis, Modeling, and Optimal Experimental Design under Uncertainty: From Carbon Nano-Structures to 3D Printing

ABSTRACT

In this thesis, we develop approaches for carrying out inference and model-based experimental design, under both internal and external sources of uncertainty.

Specifically, in Chapter 1, we develop a stochastic growth model for the carbon-based super material, Graphene, and propose approaches for relating controllable experimental factors to the underlying growth mechanism.

In Chapter 2 we develop a unified framework for carrying out response surface optimization when the input factors are noisy, and in Chapter 3, we explore the problem of designing optimal experiments, under the extra uncertainty generated by noisy inputs. Internal noise, a term used to describe the phenomenon of noisy inputs, is found to adversely affect optimization and model-based optimal designs. We show that accounting for this internal noise during the design and modeling stages significantly improve inference. In particular, we develop a modified optimality criterion for generating optimal experimental data, and show improvements in subsequent inference based on that data.

In Chapter 4, a missing data perspective is used to improve inference on deformations along the profile of 3D printed products. We show that these deformations depend on missing angles, which can be used to infer global and local deformation patterns. We use the inferred deformation model to design compensation plans for minimizing deformations on future printed objects.

Contents

0	INTRODUCTION	I
1	A STOCHASTIC GRAPHENE GROWTH KINETICS MODEL	3
1.1	Brief overview of graphene growth, the WH-model and its limitations	5
1.2	A generalized stochastic graphene growth kinetics model	8
1.3	Observational level model for experimental data	21
1.4	Estimation and Results	27
1.5	Discussions and future extensions	34
2	RESPONSE SURFACE OPTIMIZATION IN THE PRESENCE OF INTERNAL NOISE	37
2.1	Research Motivation: CNT Alignment	40
2.2	How Severe is the Internal Noise? A Unit-free Measure	42
2.3	A Bayesian framework for Response Surface Optimization in the presence of internal noise	45
2.4	Simulation Studies for Assessing Sensitivity of Objective Function to Noise	51
2.5	Performance Loss when Internal Noise is Ignored	57
2.6	Application to CNT Alignment	62
2.7	Concluding remarks	66
3	OPTIMAL EXPERIMENTAL DESIGN IN THE PRESENCE OF INTERNAL NOISE	68
3.1	Identification of Design Optimality Criterion	70
3.2	Computational Strategies for Finding Optimal Designs under the Linear Model	73
3.3	Simulations to Assess Loss Accrued by Ignoring Internal Noise	75
3.4	Application to CNT alignment	82
3.5	Concluding Remarks	87
4	A MISSING DATA PERSPECTIVE FOR MODELING DEFORMATIONS IN 3D PRINTED PRODUCTS	88
4.1	Deformation as a Missing Data Problem	90
4.2	Models	94
4.3	Estimation and Results	99
4.4	Compensation Plan Design	106
4.5	Concluding Remarks	109

APPENDIX A APPENDIX FOR CHAPTER 1	110
A.0 Glossary of notations	110
A.1 Proof of Theorem 1.2.1	112
A.2 Diagnostic plots	114
APPENDIX B APPENDIX FOR CHAPTER 2	115
B.1 Signal to Noise Ratio	115
APPENDIX C APPENDIX FOR CHAPTER 4	117
C.1 MCMC Diagnostic Plots	117
REFERENCES	124

FOR MY FAMILY, BOTH PRESENT AND FUTURE.

Acknowledgments

WORDS CANNOT SUFFICIENTLY EXPRESS my deep gratitude to Professor Tirthankar Dasgupta and Professor Joseph K. Blitzstein, my dissertation advisors, for their patience, untiring guidance, and enthusiastic encouragement throughout my doctoral program. You have both been fundamental to my growth as a scientist, and I am deeply grateful for your constant belief in me. I am also very grateful to Professor Qiang Huang, for his advice, and the opportunities for working on the very interesting and meaningful research projects presented in this dissertation.

I would like to extend my sincere thanks to the faculty of the Statistics department, particularly Professors Carl Morris, Xiao-li Meng, Dave Harrington and Michael Parzen. Thank you for providing such wonderful and inspiring learning opportunities. Learning and teaching statistics has been exciting because of you. I would also like to thank the staff of the statistics department (Betsey Cogswell, Maureen Stanton, James Matejek, Madeleine Straubel, Alice Moses, and Ali Ba) for always assisting with the resources required in running the doctoral program. I am indebted to you all.

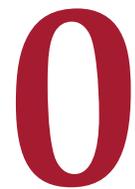
My completion of this doctoral program could not have been accomplished without the support of my colleagues and friends, especially Alex Franks, Lazhi Wang, David Jones, Xufei Wang, Yang Chen, Alex D'amour, Jiannan Liu, Dan Cervone, Janos Kramer, Arman Sabbaghi, Sam Wong, Viviana Garcia-Horton, Nikola Andrik, Xiao Tong, Zach Branson, and Stephane Shao. All experiences are richer with friends- Thank you for being such great ones.

I wish to thank my family for their support and encouragement throughout my study. Mom and

Dad, thank you for everything! Kay, Dolly-P, and Funmi, you have been my muse these past years.

Thank you!

Finally, I would like to thank my wonderful fiancée for supporting and encouraging me throughout my doctoral program. Tori, my deepest thanks.



Introduction

Ever since Sir R.A. Fisher's foundational contributions during the early twentieth century, modern statistics has been driven by the challenges of inference in the presence of uncertainty. Some of these challenges involve appropriate data sourcing, modeling, the quantification of uncertainty, and the combination of evidence from multiple sources, and have spawned several subfields in response, including experimental design, causal inference and Bayesian statistics.

The linking of statistical methods with theories underlying physical mechanisms, which is a critical step in inference, has been further strengthened in the last few decades by the advent of computing, especially through Bayesian parametric methodologies. This linking process, particularly as it exists at the interface of engineering theory and application, provides a major part of the inspiration and motivation for this thesis. Application of statistical theory to engineering problems is a rela-

tively nascent endeavor (relative to applications in life sciences and social sciences) and some of the challenges this thesis attempts to address include the correct quantification of uncertainty, model development relating underlying physics to data, and the generation of better data using inferred models.

The observational scale within engineering problems typically ranges from the micro to the macro level, and this thesis matches that by exploring uncertainty within Carbon nanostructures as well as 3D printed products. The first three chapters in particular focus on Graphene and Carbon Nanotubes (CNTs), two of the most promising and versatile nanomaterials relevant to industrial engineering. The fourth chapter, on the other hand, focuses on 3D printing, a highly promising manufacturing technique for building physical products layer-by-layer. In each chapter, these applications form the backbone through which methodology is developed to quantify uncertainty, create physics-based inferential models, and design new experiments.

Specifically, in Chapter 1, a stochastic growth model is developed for Graphene, with links proposed for relating controllable experimental factors to the underlying growth mechanism. Chapter 2 develops a unified approach to response surface optimization when the input factors are noisy, and studies the effect, on optimization, of ignoring available information on the noise. In Chapter 3, the problem of designing optimal experiments, under the extra uncertainty generated by noisy inputs, is explored. A modified optimality criterion for generating optimal experimental data is developed, and is shown to improve subsequent inference based on that data. In Chapter 4, a missing data perspective is used to improve inference on deformations along the profile of 3D printed products, and to design compensation plans for minimizing deformations on future printed objects.

*Growth is never by mere chance; it is the result of forces
working together.*

James Cash Penney

1

A Stochastic Graphene Growth Kinetics

Model

GRAPHENE IS THE NAME given to a flat monolayer of carbon atoms tightly packed into a two-dimensional (2D) honeycomb lattice, and is a basic building block for graphitic materials of all other dimensionalities¹⁶. In the realm of materials science and condensed matter physics, graphene represents an exciting new class of materials that continually offers new inroads into low-dimensional physics¹⁶, and a broad spectrum of applications including optoelectronics, chemical and biosensing^{26,52,53}. The unique room-temperature properties of graphene, such as the quantum hall effect⁵¹

and high electron mobility⁴, make its high yield production particularly desirable.

Although several methods currently exist for synthesizing graphene, recent interest has been focused on synthesizing graphene through the chemical decomposition of hydrocarbons on metals due to relatively low production costs. Specifically chemical vapor deposition of hydrocarbons on copper has proved particularly promising in producing high quality large-area monolayer graphene³⁰ and is thus our focus in this paper.

The synthesis of large-area high-quality monolayer graphene requires some understanding of its growth mechanism and the impact of underlying process factors (e.g., temperature and pressure), since faster growth kinetics will lead to larger-area graphene. The mechanisms underlying graphene growth and their relationship to the process conditions have recently received much interest^{3,8,25,65,59}, albeit through deterministic models. Recently, Wu & Huang⁶⁴ proposed a novel approach to deterministically model the growth kinetics of the graphene flakes, using the well-studied confined exponential model. Whereas the work of Wu & Huang⁶⁴ can be considered pioneering, their model, henceforth referred to as the WH-model, neither permits quantification of uncertainty associated with the growth kinetics nor relates the kinetics to the process conditions. In this article, we first develop a stochastic growth model which shares some mathematical properties with the WH-model, link it to the growth kinetics and then propose inferential models for learning about the kinetics from experimental data. Furthermore we propose links between controllable experimental factors and the kinetics, and consequently provide a framework for the statistical design and analysis of future experiments.

The paper is organized as follows: in Section 1.1, we briefly explain the graphene growth experiments, give a brief overview of the WH-model, argue that some of the assumptions are overly restrictive, and indicate how they can be relaxed. In Section 1.2, we discuss details of the proposed model and show how various model assumptions lead to derivation of the growth kinetics function. Section 1.3 details how an inferential model, which accounts for the sampling design, can be developed

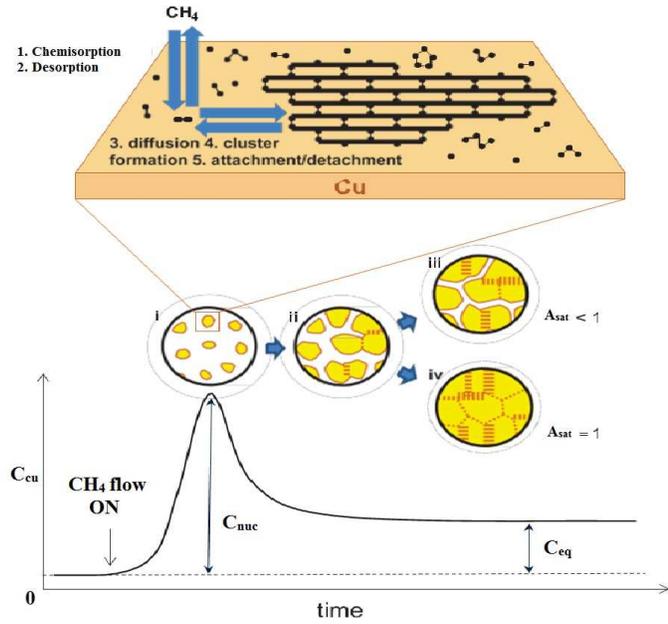


Figure 1.1: Overall illustration of the nucleation and growth mechanism of graphene on copper²⁴. Graphene growth proceeds from stage (i) to stage (ii), at which point it either continues to stage (iii) or stage (iv).

for learning about the parameters of the kinetics. In Section 1.4 we discuss how the parameters of the model can be estimated via a Bayesian approach and the uncertainty associated with estimation assessed. Section 1.4 also details sensitivity analyses based on the sampling design. In Section 1.5, we propose several model extensions and make concluding remarks.

1.1 BRIEF OVERVIEW OF GRAPHENE GROWTH, THE WH-MODEL AND ITS LIMITATIONS

Kim et al.²⁴ clearly explain the processes involved in the formation of graphene on copper. A typical experimental run consists of passing methane over a copper foil (or tray), at some pre-determined value for the process conditions (for example, Kim et al.²⁴ investigated the growth of graphene on a copper surface from 720 °C to 1050 °C).

As stated by Kim et al.²⁴ and shown in Figure 1.1, the breakdown of methane on the copper surface increases the concentration (C_{cu}) of the active carbon atoms (or adatoms), until it reaches a

critical supersaturation level ($C_{nuc} \geq 3.82 \times 10^{15} cm^{-2}$), where formation of stable graphene nuclei (islands) takes place (Figure 1.1 stage (i)). This formation is also referred to as *nucleation*. Kim et al.²⁴ note that impurities, surface roughness, grain boundary grooves, and stepped terraces on the copper foil often play important roles in graphene nucleation. In particular, they demonstrate that sites exhibiting those properties are energetically favorable sites for nucleation and that nucleation occurs almost exclusively in those regions.

As the nucleation and growth of the graphene islands deplete the adsorbed carbon atoms surrounding them, the C_{cu} is quickly reduced to a level where the nucleation rate is negligible. The growth of the nuclei continues until the supersaturated amount of surface carbon atoms above the equilibrium level C_{eq} is consumed and an equilibrium between graphene, surface carbon, and methane is reached (Figure 1.1 stage (ii)). This equilibrium is dependent on the twin balancing acts of adsorption/chemisorption and desorption, as well as attachment and detachment of carbon atoms to each other. Depending on the available carbon, graphene nuclei either coalesce to form eventually a saturated continuous film (Figure 1.1 stage (iii)) or stop growing to reach a final incomplete coverage (Figure 1.1 stage (iv)), where A_{sat} denotes the proportion of saturation on the copper surface.

The WH-model seeks to uncover the kinetics driving the growth of graphene during its production process. Two fundamental assumptions of the model are: (a) single growth mechanism: individual graphene islands under the same process conditions are driven by the same kinetics, and (b) self-similarity: the growth kinetics and island shapes have separable effects on the area growth velocity of each graphene island. Based on the above assumptions, the following multiplicative model was proposed for the angular dependent area growth velocity, $v_i(\vartheta, t)$, of the i th graphene island at angle ϑ and time t :

$$v_i(\vartheta, t) = g_i(\vartheta)q(t), \quad (1.1)$$

where $g_i(\vartheta)$ describes the shape of the i th graphene island, and $q(t)$ is the *deterministic* kinetics driving the growth at time t . It is important to note at this point that we describe the shape, $g_i(\vartheta)$, in polar coordinates as is natural when dealing with 2D closed profiles. Another way to interpret the model is to imagine that each island starts with shape $g_i(\vartheta)$ and is scaled up by the kinetic function⁶⁴.

In general, it is very difficult to observe or measure $v_i(\vartheta, t)$ and thus this model is instead linked to the more measurable (and thus more easily modeled) total area covered on the copper foil/tray at time t , $S(t)$. In Figure 1.1 for instance, $S(t)$ represents the sum total of the shaded areas within each circle in stages (i)-(iv). Denoting the constant number of graphene islands on the tray at any time t by N , Wu & Huang⁶⁴ derived the following model for the growth kinetics:

$$q(t) = \frac{\frac{d}{dt}S(t)}{2\sqrt{NS(t)}}, \quad (1.2)$$

based on the assumption that $g_i(\vartheta)$ follows the model:

$$g_i(\vartheta) = g(\vartheta) + \varepsilon_i, \quad \text{where } \varepsilon_i \sim \mathcal{N}(0, \sigma^2).$$

The above model appears to have several limitations. First, it does not allow for modeling and evaluation of uncertainty, because (1.2) contradicts the assumption that $q(t)$ is deterministic. This is due to the fact that all the $g_i(\vartheta)$, and consequently $S(t)$, are random. Second, the assumption of N remaining constant over time is clearly unrealistic and overly restrictive, especially for the first few instants of time. Third, the model does not account for the time-varying random behavior (attachment/detachment) of carbon atoms along the profile of each graphene island. Fourth, it does not suggest a way to link different process conditions to the growth kinetics. Finally, it does not allow us to understand what type of experimental data are needed to estimate the model parameters, and

how to analyze such data with the objective of estimating model parameters and making predictions from the fitted model.

Some of the models proposed in this paper retain the fundamental structure (1.1) of the WH-model, but makes several generalizations. First, the component shape $g_i(\vartheta)$ is assumed to be random, but restricted to have area, A_{nuc} . As Vlassiok et al.⁶⁰ note, a simple estimate for the critical size of surviving carbon cluster is about 140 - 320 atoms. Combining this with the lower bound for C_{nuc} (described above) gives the estimate $A_{nuc} \approx 6.02 \times 10^{-14} \text{cm}^2$. We note here that $A_{nuc} = 1$ was proposed by Wu & Huang⁶⁴ and this is appropriate only if the goal is to compare the kinetics for different copper trays. In our proposed model, the kinetics are reformulated in terms of $E(S(t))$ (where $E(X)$ stands for expectation of a random variable X), through acknowledgment of the effect of random carbon atom attachment and detachment along the profile of each island. Another major change is the use of the random process $N(t)$ instead of a constant (N) number of graphene islands to make the model much more realistic. Next, we consider the possibility of having multiple trays of graphene islands, each tray representing a single experimental run at a specific value of the process conditions. Such a representation facilitates the task of linking the growth kinetics with the process conditions. Finally we present a parametric approach for estimating $q(t)$, along with a procedure for quantifying the uncertainty in the estimation as well. In particular, we focus on Bayesian methods as they provide a natural framework for specifying the interdependence of the various components of the parametric model.

1.2 A GENERALIZED STOCHASTIC GRAPHENE GROWTH KINETICS MODEL

Let $N(t)$ denote the random number of islands on the tray at time t . Denoting by $v_i(\vartheta, t)$, the angular velocity of the i th island at time t and angle ϑ , and $\tilde{v}_i(\vartheta, t)$ its expectation over the random

micro-scale attachment/detachment of carbon atoms, we can generalize (1.1) to

$$\tilde{v}_i(\mathcal{D}, t) = E[v_i(\mathcal{D}, t)] = g_i(\mathcal{D})q(t), \quad \{i = 1, \dots, N(t)\}, \quad (1.3)$$

where $g_i(\mathcal{D})$ denotes the fixed area random shape of island i and $q(t)$ denotes the kinetics driving the growth time t . Our objective is to estimate $q(t)$ as a function of time t , and to eventually compare $q(t)$ for multiple process conditions, since each kinetics curve reflects the effects of the process conditions on the growth rate.

Let $t_{oi} \geq 0$ denote the random time the nucleation process begins for the i th island and let t_{oi}^* (where $t_{oi}^* > t_{oi}$) denote the random time of nucleation for that island. Note that $t_{oi}^* - t_{oi} \approx 0$ due to the relatively small value of \mathcal{A}_{nuc} , and for mathematical simplicity we will assume $t_{oi}^* - t_{oi}$ to be infinitesimally small.

Now, let $S_i(t_{oi}, t)$ denote the area covered by the i th island by time t , and let $\tilde{S}_i(t_{oi}, t)$ denote the equivalent area covered when the random micro-scale behavior of the carbon atoms is averaged out. Two hypothetical curves $S_i(t_{oi}, t)$, one with $t_{oi} = 0$ i.e., instant nucleation (broken-line curve) and the other with an arbitrary $t_{oi} > 0$ (solid-line curve) are shown in Figure 1.2.

We mention at this point (and show in Appendix A.1) that $S_i(t_{oi}, t)$ only depends on the random shape, $g_i(\mathcal{D})$, through its non-random area, \mathcal{A}_{nuc} . Therefore, any randomness in $S_i(t_{oi}, t)$ is completely determined by the distribution of the nucleation times, t_{oi} , as well as the random attachment/detachment of carbon atoms to the edges of each island. Consequently, any randomness in $\tilde{S}_i(t_{oi}, t)$ is completely explained by the distribution of the nucleation times, t_{oi} .

The total area $S.(t)$ covered by the graphene islands on the tray by time t can be expressed as

$$S.(t) = \sum_{i=1}^{N(t)} S_i(t_{oi}, t) = \sum_{i=1}^{N(\infty)} S_i(t_{oi}, t) \cdot \mathbf{1}\{t_{oi}^* \leq t\} = \sum_{i=1}^{N(\infty)} S_i(t_{oi}, t) \cdot \mathbf{1}\{t_{oi} < t\}, \quad (1.4)$$

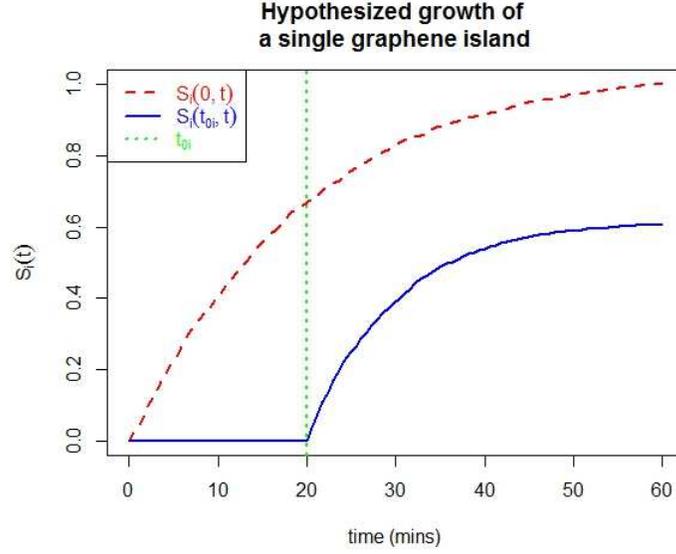


Figure 1.2: Hypothesized growth curves for a single island under assumptions of instant ($S_i(0, t)$), and random delayed ($S_i(t_{oi}, t)$) nucleation.

and equivalently, under the process where the random behavior of the carbon atoms is averaged out, the total area $\tilde{S}(t)$ covered by the graphene islands on the tray by time t can be expressed as

$$\tilde{S}(t) = \sum_{i=1}^{N(t)} \tilde{S}_i(t_{oi}, t) = \sum_{i=1}^{N(\infty)} \tilde{S}_i(t_{oi}, t) \cdot \mathbb{1}\{t_{oi}^* \leq t\} = \sum_{i=1}^{N(\infty)} \tilde{S}_i(t_{oi}, t) \cdot \mathbb{1}\{t_{oi} < t\}. \quad (1.5)$$

We will now make use of a simple example to show the effects of the nucleation times and random attachment/detachment of carbon atoms on a hypothesized tray.

Example: Individual island growth

Suppose, for a hypothetical growth process for graphene, we have that $q(t) = 100 \exp(-t)$. Further suppose that we observe three different graphene islands, each with a distinct shape ($g_i : i = 1, 2, 3$) and nucleation time ($t_{oi} : i = 1, 2, 3$). We let g_1, g_2 and g_3 represent specific circular, three-lobed, and four-lobed shapes respectively, with A_{nuc} set to 10^{-5} cm^2 . Specifically, we first select coefficients $\beta_{i,k}$ ($i = 1, 2, 3; k = 0, 1, \dots, 8$) for unnormalized shapes h_1, h_2 and h_3 in polar

coordinates such that

$$\log(b_i(\vartheta)) = \beta_{i,0} + \sum_{m=1}^4 \beta_{i,2m-1} \cos(m\vartheta) + \beta_{i,2m} \sin(m\vartheta),$$

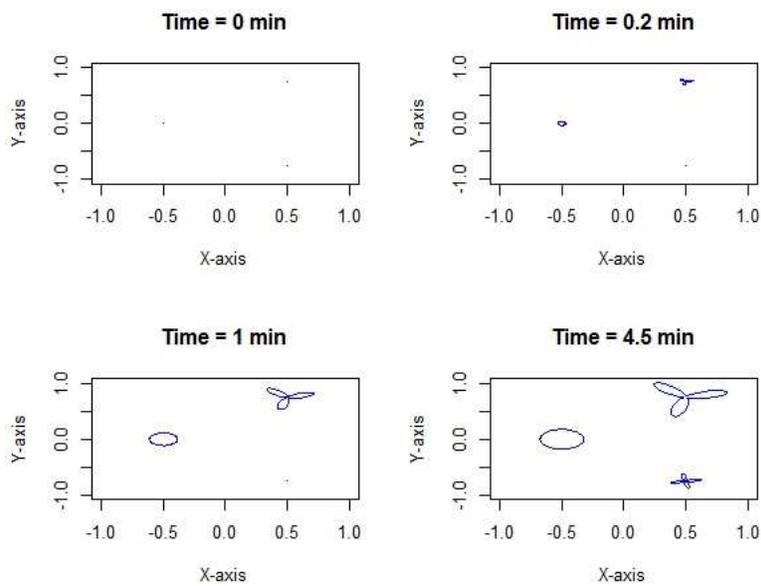
and then we normalize appropriately by defining g_1, g_2 and g_3 through

$$g_i(\vartheta) = \sqrt{A_{nuc}} \cdot \frac{b_i(\vartheta)}{\sqrt{\frac{1}{2} \int_0^{2\pi} b_i^2(\vartheta) d\vartheta}}.$$

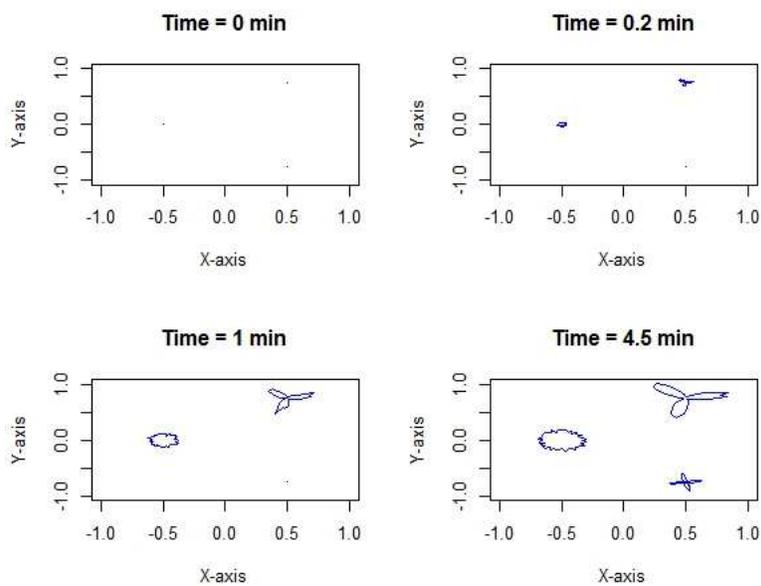
For the circular shape, g_1 , we set $\beta_{1,0} = 1$, and set $\beta_{1,k} = 0$ for all $k \neq 0$, so that $g_1(\vartheta)$ is constant, as expected of circles. For the three-lobed shape, g_2 , we set $\beta_{2,k} = 1$ for $k \in \{0, 5, 6\}$, and set $\beta_{2,k} = 0$ otherwise. Finally, for the four-lobed shape, g_3 , we set $\beta_{3,k} = 1$ for $k \in \{0, 7, 8\}$, and set $\beta_{3,k} = 0$ otherwise.

Next, we set nucleation times, $t_{o1} = 0$ mins, $t_{o2} = 0$ mins and $t_{o3} = 1$ min, and select observation times $t = 0, 0.2, 1, \text{ or } 4.5$ mins. In order to represent the effect of random carbon attachment and detachment from the profile edges, we add some noise to (1.3). Specifically, we set $v_i(\vartheta, t) = \tilde{v}_i(\vartheta, t) \cdot e_i(\vartheta, t)$, where $e_i(\vartheta, t)$ is a log-normal random variable with parameters $\mu = -0.005$ and $\sigma = 0.1$, i.e., $e_i(\vartheta, t) \sim \mathcal{LN}(-0.005, 0.1^2)$. Note that $E(e_i(\vartheta, t)) = 1$ always. Fig 1.3a shows the underlying growth process on the tray, when the random carbon attachment is averaged out, while Fig 1.3b shows the corresponding growth process as observed.

From Fig 1.3a we can see that average growth along any direction is proportional to the initial shape at nucleation, and that the total area covered on the tray is highly dependent on the nucleation time for each graphene island. Taking the circular island as an example under this average growth process, the radius and area at time $t = t_{o1}$ is given by 0 cm and 0 cm² respectively. At time $t = t_{o1}^* = t_{o1} + \varepsilon$ (for some $\varepsilon > 0$, an infinitesimally small number by assumption), the radius and area are $\sqrt{A_{nuc}/\pi}$ cm and A_{nuc} cm² respectively. More generally, at some general time t mins (where



(a) The growth of three graphene islands averaged over random carbon attachment.



(b) The actual growth of three graphene islands. The rough profiles show the effect of random attachment of carbon atoms.

Figure 1.3

$t > t_{oi}^*$), the radius and area become $\sqrt{A_{muc}/\pi} \cdot 100(1 - \exp(-t))$ and $A_{muc} \cdot [100(1 - \exp(-t))]^2$ respectively. Note that the assumption of infinitesimally small $t_{oi}^* - t_{oi}$ implies that the interval (t_{oi}, t_{oi}^*) will contain a general observation time t with zero probability, and allows us to use the same distribution for t_{oi} and t_{oi}^* . Furthermore, according to (1.5), the total area covered on the tray, by all three graphene islands in our example, at time $t = 4.5$ mins is therefore $\tilde{S}(4.5) = \tilde{S}_1(0, 4.5) + \tilde{S}_2(0, 4.5) + \tilde{S}_3(1, 4.5)$, where $\tilde{S}_1(0, 4.5)$, $\tilde{S}_2(0, 4.5)$ and $\tilde{S}_3(1, 4.5)$ are the areas of the three shapes observed in the bottom right panel of Fig 1.3a.

Fig 1.3b shows the lack of smoothness along the profile of each island, resulting from the random behavior of carbon atoms along the edge of each island. Since we will henceforth be concerned only with aggregated measures, such as the total area covered on each tray, such detailed modeling of the observed carbon atom behavior along the edge of each island will be unnecessary, and we will only need to model the noisy behavior at the aggregated level.

We now state a result (proof in Appendix A.1) which provides us with a closed-form expression for the kinetics, $q(t)$.

Theorem 1.2.1 *Assuming the multiplicative model (1.3), the kinetics $q(t)$ driving the growth on a specific tray has the following form:*

$$q(t) = \frac{1}{\sqrt{A_{muc}}} \left[\frac{\xi'(t) - \hat{\sigma}'(t)}{2\sqrt{\xi(t) - \hat{\sigma}(t)}} + \phi'(t) \right], \quad (1.6)$$

for $\xi(t) > \hat{\sigma}(t)$ at all times ($t > 0$), where

$$\begin{aligned} \xi(t) &= E(\tilde{S}_i(t_{oi}, t) | t_{oi} \leq t), \\ \phi(t) &= E\left(\sqrt{\tilde{S}_i(0, t_{oi})} | t_{oi} \leq t\right), \text{ and} \\ \hat{\sigma}(t) &= Var\left(\sqrt{\tilde{S}_i(0, t_{oi})} | t_{oi} \leq t\right). \end{aligned}$$

Remarks

1. Note that the area forfeited at any time t by an island with nucleation time t_{oi} is given by the difference $S_i(o, t) - S_i(t_{oi}, t)$. Thus, the quantity $\tilde{S}_i(o, t_{oi})$ under the square-root sign in the expressions for $\phi(t)$ and $\hat{\sigma}(t)$ can be interpreted as the forfeited area of the i th island at its nucleation time t_{oi} , when the random behavior of the edge carbon atoms is averaged out. This is because $S_i(t_{oi}, t_{oi}) = 0$ by definition.
2. The expectations in the expressions for $q(t)$ and $\phi(t)$, and the variance in the expression for $\hat{\sigma}(t)$ are with respect to the distribution of the nucleation times, which we determine in Section 1.2.2.
3. Under instant nucleation, $\tilde{S}_i(o, t)$ is the area covered by the i th island, when the random carbon atom behavior at the edge of the island is averaged out. In a departure from our example, $\tilde{S}_i(o, t)$ need not be implicitly defined through $\tilde{v}_i(\vartheta, t)$ and the distribution of carbon atom attachment/detachment; we can derive analytical expressions for $\tilde{S}_i(o, t)$ by first assuming suitable models for $E(S_i(o, t))$ and the nucleation time distribution $f(t_{oi})$, and then setting $\tilde{S}_i(o, t) = E(S_i(o, t))$. Note that in general we use $\tilde{S}_i(t_{oi}, t) = E(S_i(t_{oi}, t)|t_{oi})$, so that we can rewrite $\xi(t)$ from the theorem above as

$$\xi(t) = E(\tilde{S}_i(t_{oi}, t)|t_{oi} \leq t) = E(S_i(t_{oi}, t)|t_{oi} \leq t) .$$

4. From Theorem 1.2.1 and the previous three remarks, it follows that in order to estimate $q(t)$, we need first to decide on suitable models for $E(S_i(o, t))$, $f(t_{oi})$ (the nucleation time distribution), and $E(S_i(t_{oi}, t))$ (the expected graphene area for i th island at time t , with random nucleation time, $t_{oi} \sim f(t_{oi})$). These models are derived in Sections 1.2.1-1.2.3. While developing these models, we make use of the models already proposed by Wu & Huang⁶⁴ whenever

appropriate. Multiple model modifications which make uncertainty analyses possible and which reflect more realistic conditions have additionally been proposed.

1.2.1 A MODEL FOR $E(S_i(o, t))$

Let $S^\circ(t)$ denote the graphene area on the tray under the assumption of zero nucleation times for all islands on the tray ($t_{oi} = 0$ for all i). From (1.4), we can write by substituting $t_{oi} = 0$

$$S^\circ(t) = \sum_{i=1}^{N(\infty)} S_i(o, t) .$$

Applying the law of iterated expectations, we have that

$$E(S^\circ(t)) = E(N(\infty)) \cdot E(S_i(o, t)) ,$$

or

$$\tilde{S}_i(o, t) = E(S_i(o, t)) = \frac{E(S^\circ(t))}{E(N(\infty))} . \quad (1.7)$$

In developing a model for $E(S^\circ(t))$, we re-visit the model proposed by Wu & Huang⁶⁴, who noted that the area growth over the copper tray at time t (which we denote by $S^\circ(t)$) is well described by a confined exponential growth model, which is essentially the following first order differential equation (ODE)

$$\frac{d}{dt}\{S^\circ(t)\} = \alpha(S^* - S^\circ(t)) .$$

Solving the ODE under the assumption of deterministic $S^\circ(t)$ then leads to the following expression for the area

$$S^\circ(t) = S^*(1 - \exp(-\alpha t)) . \quad (1.8)$$

Here, $\alpha > 0$ is the parameter governing the expansion/growth on the tray, and S^* is the area of the

copper tray (the maximum achievable area).

By assuming a constant number of islands, Wu & Huang⁶⁴ implicitly assume zero nucleation times for all the islands. It is thus natural to develop a model for $E(S_i(o, t))$ from this confined exponential model. Furthermore, note that the limiting (at infinite time) area according to (1.8) is S^* , which indicates complete coverage of the tray. We relax this assumption and make an adjustment for incomplete coverage. Denoting the graphene area on the tray under the assumption of zero nucleation times for all islands on the tray as $S^\circ(t)$, now consider the following model for the expected area

$$E(S^\circ(t)) = E(S^\circ(\infty)) \cdot (1 - \exp(-\alpha t)) , \quad (1.9)$$

where $\alpha > 0$ is the parameter governing the expansion/growth on the tray, and $E(S^\circ(\infty))$ is the expected limiting area on that tray. By making the assumption that higher valued growth parameters lead to more tray coverage, we further propose the following

$$E(S^\circ(\infty)) = (1 - \exp(-\alpha\omega))S^* , \quad (1.10)$$

where $\omega > 0$ is a parameter which governs the relationship between α and coverage. Consequently from (1.9) and (1.10),

$$E(S^\circ(t)) = S^* (1 - \exp(-\alpha\omega))(1 - \exp(-\alpha t)) . \quad (1.11)$$

Substituting the expression for $E(S^\circ(t))$ from (1.11) into (1.7), the final model for $E(S_i(o, t))$ is thus given by

$$\tilde{S}_i(o, t) = E(S_i(o, t)) = \frac{(1 - \exp(-\alpha\omega))S^*}{E(N(\infty))} (1 - \exp(-\alpha t)) . \quad (1.12)$$

In the next section we develop a distribution for the nucleation times, $f(t_{oi})$, and subsequently obtain $E(N(\infty))$.

1.2.2 A MODEL FOR $f(t_{oi})$, THE NUCLEATION TIME DISTRIBUTION

One model for $f(t_{oi})$ stems from recognizing $N(t) = \sum_{i=1}^{\infty} \mathbf{1}_{\{t_{oi} \leq t\}}$ as a constrained birth process. By proposing the Poisson process $N(t) \sim PP(\lambda_j(t))$ (where $\lambda(t)$ is the underlying non-homogeneous intensity function for the tray), we have that:

$$f(t_{oi}) = \frac{\lambda(t_{oi})}{\int_0^{\infty} \lambda(s) ds}. \quad (1.13)$$

Any choice of the intensity function will need to take the following consideration into account. The appearance of a new graphene island depends on the proportion of area left untouched on the tray at time t . The less the available area, the lower the number of islands that can appear. We thus propose the following intensity function

$$\lambda(t) = \beta S^* \left(\frac{E(S^\circ(\infty)) - E(S^\circ(t))}{E(S^\circ(\infty))} \right),$$

where β is some positively valued parameter, S^* is the limiting area on the tray according to (1.8), and $E(S^\circ(t))$ is given by (1.9). After substitution of our model for $E(S^\circ(t))$ from (1.9), we have that

$$\lambda(t) = \beta S^* \exp(-\alpha t).$$

We can now obtain the nucleation time distribution as

$$f(t_{oi}) = \alpha \exp(-\alpha t_{oi}) \quad \text{which implies} \quad t_{oi} \sim \text{Exp}(\alpha), \quad (1.14)$$

where $\text{Exp}(a)$ denotes the exponential distribution with parameter a .

Subsequently, we also obtain that

$$E(N(t)) = \frac{\beta S^*}{\alpha} (1 - \exp(-\alpha t)) \quad \text{and} \quad E(N(\infty)) = \frac{\beta S^*}{\alpha}. \quad (1.15)$$

This inverse relation between the average number of islands at infinite time, and the growth parameter, matches the intuition that faster growth leaves less room for new islands to form, thus resulting in a lower number of islands at infinite time.

1.2.3 A MODEL FOR $E(S_i(t_{oi}, t))$

Based on the models derived in the last two sections, we can now develop a model for the expected graphene area for a single graphene island, with random nucleation time, $t_{oi} \sim \text{Expo}(\alpha)$. In order to achieve this, we first propose the following link between the expected graphene area on a tray assumed to have instantaneous nucleation ($E(S^\circ(t))$), and the expected graphene area on a tray where such an assumption is absent ($E(S(t))$)

$$E(S(t)) = E(S^\circ(t)) \cdot \text{Pr}(t_{oi} \leq t). \quad (1.16)$$

This formulation implies that on average, the area on a tray with random nucleation is determined by the area on the same tray under zero nucleation times, down-weighted by the proportion of islands which have nucleated. This allows us to recover $E(S^\circ(t))$ under the specific case of zero nucleation times, since $\text{Pr}(t_{oi} \leq t) = 1$ for all t in that case. Substituting the expression for $E(S^\circ(t))$ from (1.11) into (1.16), and noting that $\text{Pr}(t_{oi} \leq t) = 1 - \exp(-\alpha t)$, it follows that

$$E(S(t)) = S^* (1 - \exp(-\alpha \omega)) (1 - \exp(-\alpha t))^2. \quad (1.17)$$

By defining $S_i(t_{oi}, t) = 0$, whenever $t_{oi} > t$, and applying the law of iterated expectations on (1.4), we obtain

$$E(S_i(t)) = E(N(\infty)) \cdot E(S_i(t_{oi}, t)).$$

Substituting (1.15) and (1.17) into the above expression, the final model for $E(S_i(t_{oi}, t))$ is given by

$$E(S_i(t_{oi}, t)) = \frac{\alpha(1 - \exp(-\alpha\omega))}{\beta} (1 - \exp(-\alpha t))^2. \quad (1.18)$$

1.2.4 EXPRESSIONS FOR $\xi(t)$, $\phi(t)$ AND $\hat{\sigma}(t)$

Based on the models developed in Sections 1.2.1-1.2.3, we can now derive expressions for $\xi(t)$, $\phi(t)$ and $\hat{\sigma}(t)$. By the law of total expectation, we know that

$$E(S_i(t_{oi}, t)) = E(S_i(t_{oi}, t)|t_{oi} \leq t) \cdot Pr(t_{oi} \leq t) + E(S_i(t_{oi}, t)|t_{oi} > t) \cdot Pr(t_{oi} > t).$$

Since we defined $S_i(t_{oi}, t) = 0$, whenever $t_{oi} > t$, we have that $E(S_i(t_{oi}, t)|t_{oi} > t) = 0$, and thus,

$$\xi(t) = E(S_i(t_{oi}, t)|t_{oi} \leq t) = \frac{E(S_i(t_{oi}, t))}{Pr(t_{oi} \leq t)} = \frac{\alpha(1 - \exp(-\alpha\omega))}{\beta} (1 - \exp(-\alpha t)). \quad (1.19)$$

The last step follows from (1.18) and the fact that $Pr(t_{oi} \leq t) = 1 - \exp(-\alpha t)$. In order to obtain expressions for $\phi(t)$ and $\hat{\sigma}(t)$, we first recall from (1.12), after plugging in the model for $E(N(\infty))$, that

$$\tilde{S}_i(0, t) = E(S_i(0, t)) = \frac{\alpha(1 - \exp(-\alpha\omega))}{\beta} (1 - \exp(-\alpha t)).$$

Now,

$$E(\sqrt{\tilde{S}_i(0, t_{oi})|t_{oi} \leq t}) = \sqrt{\frac{\alpha(1 - \exp(-\alpha\omega))}{\beta}} \int_0^t \sqrt{(1 - \exp(-\alpha t_{oi}))} \frac{f(t_{oi})}{Pr(t_{oi} < t)} dt_{oi}$$

$$\begin{aligned}
&= \sqrt{\frac{\alpha(1 - \exp(-\alpha\omega))}{\beta}} \int_0^t \sqrt{(1 - \exp(-\alpha t_{oi}))} \frac{\alpha \exp(-\alpha t_{oi})}{(1 - \exp(-\alpha t))} dt_{oi} \\
&= \sqrt{\frac{\alpha(1 - \exp(-\alpha\omega))}{\beta}} \cdot \frac{2}{3} \sqrt{(1 - \exp(-\alpha t))} \text{ after some manipulation ,}
\end{aligned}$$

and

$$\begin{aligned}
E(\tilde{S}_i(o, t_{oi}) | t_{oi} \leq t) &= \frac{\alpha(1 - \exp(-\alpha\omega))}{\beta} \int_0^t (1 - \exp(-\alpha t_{oi})) \frac{f(t_{oi})}{Pr(t_{oi} < t)} dt_{oi} \\
&= \frac{\alpha(1 - \exp(-\alpha\omega))}{\beta} \int_0^t (1 - \exp(-\alpha t_{oi})) \frac{\alpha \exp(-\alpha t_{oi})}{(1 - \exp(-\alpha t))} dt_{oi} \\
&= \frac{\alpha(1 - \exp(-\alpha\omega))}{\beta} \cdot \frac{1}{2} (1 - \exp(-\alpha t)) \text{ after some manipulation .}
\end{aligned}$$

We can now write

$$\phi(t) = \frac{2}{3} \sqrt{\frac{\alpha(1 - \exp(-\alpha\omega))}{\beta}} \sqrt{(1 - \exp(-\alpha t))}, \quad (1.20)$$

and

$$\varpi(t) = \frac{1}{2} \frac{\alpha(1 - \exp(-\alpha\omega))}{\beta} (1 - \exp(-\alpha t)) - \phi^2(t) = \frac{\alpha(1 - \exp(-\alpha\omega))}{18\beta} (1 - \exp(-\alpha t)). \quad (1.21)$$

1.2.5 MODIFIED EXPRESSION FOR THE KINETICS

We are now in a position to obtain a closed-form expression of equation (1.6) for the kinetics $q(t)$ governing the graphene growth on the tray under the models for the shape, expected area and distribution of nucleation times developed in Sections 1.2.1-1.2.3. Based on the expressions in (1.6), (1.19),

(1.20) and (1.21), we obtain the following form for $q(t)$:

$$q(t) = \frac{\left(\frac{1}{3} + \sqrt{\frac{17}{72}}\right) \sqrt{\alpha^3(1 - \exp(-\alpha\omega))} \exp(-\alpha t)}{\sqrt{A_{nuc}\beta(1 - \exp(-\alpha t))}}. \quad (1.22)$$

The role each term plays in the kinetics can now be clearly seen. As A_{nuc} becomes smaller one should expect the kinetics driving the growth process to a fixed maximum area to increase. An increase in β should naturally lead to a decrease in the kinetics driving the growth of each graphene island when all else is fixed. This is due to a decrease in the available space which each island has to grow, as a result of higher expected number of islands. An increase in ω , which signifies a higher overall limiting area, should naturally lead to an increase in the kinetics as well. Increasing α will generally lead to an increase in the kinetics, especially in the early stages of growth.

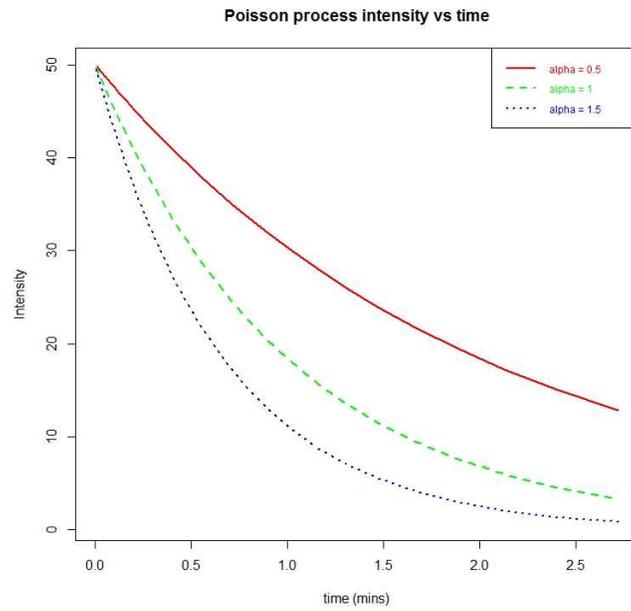
A simple example

We now illustrate the effect of α on the proposed kinetics model with a simple example. Assume we have three unique process conditions corresponding to $\alpha_1 = 0.5$, $\alpha_2 = 1$ and $\alpha_3 = 1.5$. Suppose $S^* = 1 \text{ cm}^2$, $A_{nuc} = 6.02 \times 10^{-14} \text{ cm}^2$, $\omega = 4$ and $\beta = 50$. The plots of the intensity functions and kinetics versus time are shown for the three different values of α_j (Figure 1.4).

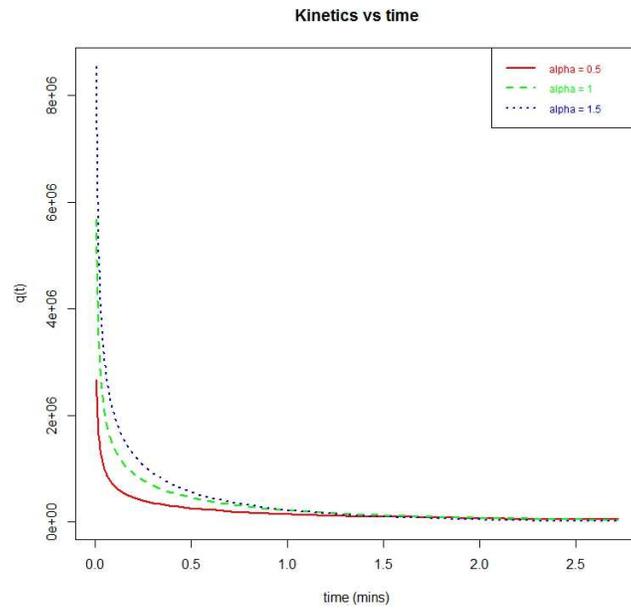
In Figure 1.4a, the intensity starts at β and rapidly declines with time for each value of α . Observe that higher values of α decay much faster, leading to a lower final nucleation density overall. In Figure 1.4b, note that higher values of α lead to higher growth kinetics especially in the beginning, when the copper trays are yet to be covered.

1.3 OBSERVATIONAL LEVEL MODEL FOR EXPERIMENTAL DATA

In the previous sections, we determined an expression for the kinetics driving graphene growth under several parametric model assumptions. The next goal is to learn about those parameters from



(a) Intensity vs time



(b) Kinetics vs time

Figure 1.4: Intensity and Kinetic plots based on a simple example.

experimental data, and thus we will need to propose a final observational level model for carrying out inference on the parameters. If we can observe $S_i(t)$ at different time points $t = t_1, \dots, t_{n_T}$, then the inference on the parameters is somewhat straightforward, and in Section 1.5, we propose a method to estimate the kinetics under this scenario. However, in many situations, the data collection mechanism prohibits observation of the whole tray, and there is therefore a need to account for the extra noise generated at the observational level. In order to motivate the observational model, we first briefly describe the experimental data.

1.3.1 MULTI-RESOLUTION DATA

Graphene grown on copper substrate was monitored at a single process condition combination (temperature = 1050°C, methane concentration = 60 ppm). At four time points during the growth, 5-8 non-overlapping images are taken at random locations on the copper tray. The resolutions of the images differ, thus ensuring that the sizes of the images, at some baseline magnification, are different. Figure 1.5 shows a sample of the data, where each time point is represented.

Measuring the area of graphene in each image requires the use of image processing software and we made use of ImageJ's Fiji software^{49,50} in the processing and Figure 1.6 shows the sample from Figure 1.5 after processing.

1.3.2 A NOISY OBSERVATION MODEL

From the description of the data and Figure 1.5, it is evident that due to the merging of islands, measurement of $N(t)$ will be difficult, if not impossible, particularly at later stages of the growth. Thus the parameter β will not be estimable from the data. What that means is that we will know the kinetics, $q(t)$ only up to some proportionality constant. Since our eventual goal is to compare the kinetics under different process conditions, we do not lose anything if we make ratio-based comparisons.

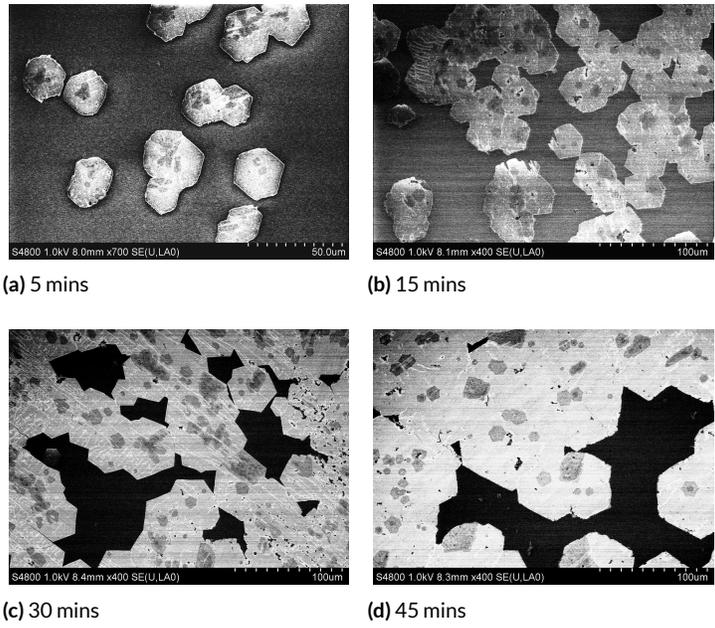


Figure 1.5: A sample of the multi-resolution data.

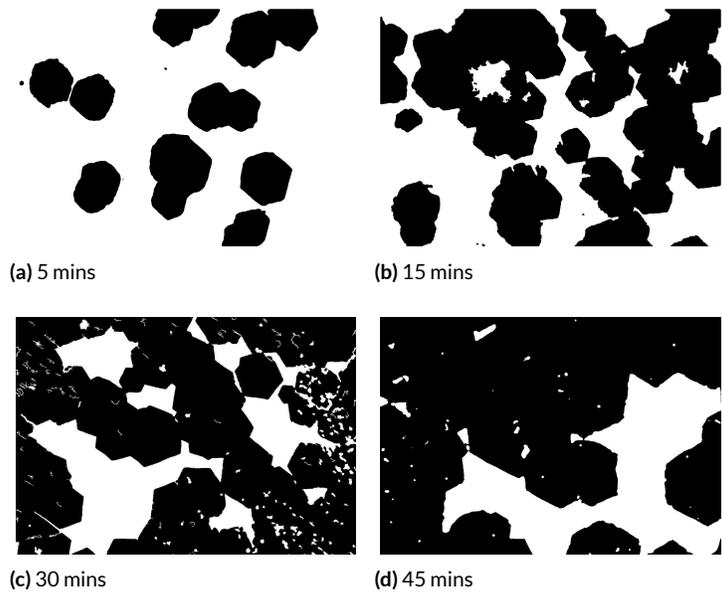


Figure 1.6: A sample of the processed multi-resolution data.

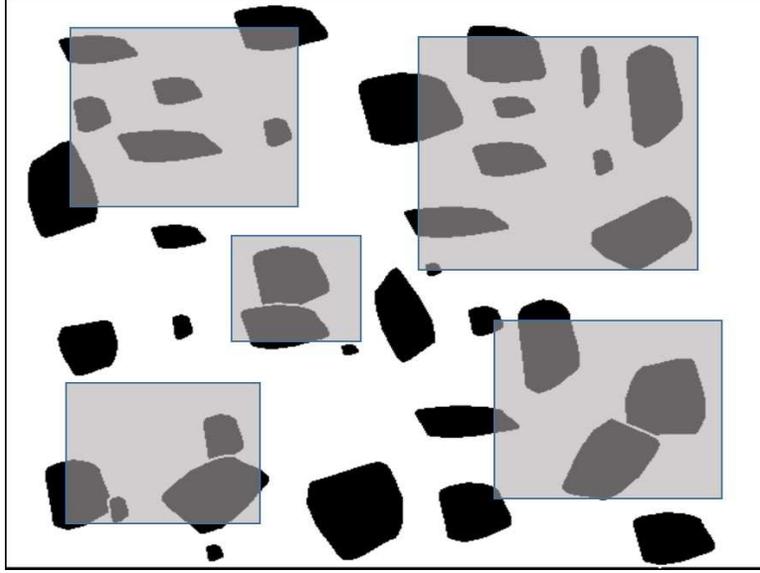


Figure 1.7: A hypothetical copper tray at some time t . The grey rectangles represent sampled portions (images) of the tray and the black objects represent graphene islands. Smaller rectangles correspond to higher resolutions/magnifications.

In order to develop a model which incorporates information from multiple resolutions, consider the hypothetical state of a given tray, along with 5 images at different resolutions depicted in Figure 1.7.

Let us denote the measured graphene area covered in l th image on the tray at time t by $u_{l,t}$, and the area of that l th image by $U_{l,t}$. In Figure 1.7, the area of a grey rectangle is $U_{l,t}$, and the sum total of the black area within that rectangle is $u_{l,t}$. Also denote the total area of all images on the tray at time t as U_t^{obs} , and the corresponding non-observed area as U_t^{mis} . Further denote the total area covered by graphene in all images on the tray at time t as $S_t^{obs}(t)$, the unknown graphene area in the unobserved portion as $S_t^{mis}(t)$, and the number of images on the tray at time t as m_t . Recall that $S(t)$ is the total area covered by graphene on the tray at time t , and S^* is the area of the copper tray. Then,

$$U_t^{obs} = \sum_{l=1}^{m_t} U_{l,t}, \quad U_t^{mis} = S^* - U_t^{obs}, \quad S_t^{obs}(t) = \sum_{l=1}^{m_t} u_{l,t}, \quad \text{and} \quad S_t^{mis}(t) = S(t) - S_t^{obs}(t).$$

Note that $\{u_{1,t}, u_{2,t}, \dots, u_{m_t,t}, S^{mis}(t)\}$, which we now denote by \mathbf{u}_t , defines a random partition of $S(t)$, with $S^{mis}(t)$ always missing/unobserved. On the other hand, $\{U_{1,t}, U_{2,t}, \dots, U_{m_t,t}, U_t^{mis}\}$, which we denote by \mathbf{U}_t , defines a known partition of S^* .

We can now build a model for the observations by making the following considerations:

1. Each $u_{l,t}/U_{l,t}$ should be an unbiased estimator of $E(S(t))/S^*$, based on the random location of the images on the tray.
2. As $U_{l,t} \rightarrow \infty$, the variance of $u_{l,t}/U_{l,t}$ should approach 0. On the other hand, as $U_{l,t} \rightarrow 0$, $u_{l,t}/U_{l,t}$ should increasingly exhibit bi-modality at 0 and 1, due to the clustering behavior of the graphene islands.
3. The variance of $u_{l,t}/U_{l,t}$ should approach zero when $E(S(t)) \rightarrow S^*$ and when $E(S(t)) \rightarrow 0$. These are the boundary conditions of the growth process.
4. Due to the clustering behavior of graphene islands, the correlation between any $u_{l,t}/U_{l,t}$ and $u_{k,t}/U_{k,t}$ should be positive and decay with the distance between the images. This distance can be defined in terms of the minimum distance between boundary points of the two images. In the absence of spatial information, this final consideration can be ignored based on a simplifying assumption of zero correlation, at the risk of underestimating the uncertainty in the model.

Based on considerations (1) - (3), we propose the following:

$$\frac{u_{l,t}}{U_{l,t}} \sim \text{Beta} \left(\nu U_{l,t} \frac{E(S(t))}{S^*}, \nu U_{l,t} \left(1 - \frac{E(S(t))}{S^*} \right) \right), \quad (1.23)$$

where ν is an unknown parameter which scales the influence of each $U_{l,t}$ on the variance, so that any rescaling of all $U_{l,t}$'s will have no effect, since $1/\nu$ will be rescaled as well. Relatively high values of ν ,

for fixed $U_{l,t}$, result in $u_{l,t}$ very close to its expected value for all $\{l, t\}$, and is appropriate for situations where clustering behavior is not expected. Relatively low values, on the other hand, represent situations where clustering is the norm, as in the case of graphene growth. The parameter ν thus provides some useful information about clustering behavior exhibited in the data.

By plugging in the model for $E(S_l(t))$ from (1.9), (1.10), (1.14) and (1.16), we can now write the full hierarchical data generating model for the observed area as

$$\begin{aligned} \frac{u_{l,t}}{U_{l,t}} &\sim \text{Beta}(\nu U_{l,t}(1 - \exp(-\alpha\omega))(1 - \exp(-\alpha t))^2, \nu U_{l,t}(1 - (1 - \exp(-\alpha\omega))(1 - \exp(-\alpha t))^2)), \\ (\nu, \omega, \alpha) &\sim \Pi(\nu, \omega, \alpha), \end{aligned}$$

where $\Pi(\cdot)$ represents some specified distribution.

1.4 ESTIMATION AND RESULTS

1.4.1 BAYESIAN ESTIMATION

We denote the data by \mathcal{D} and the parameters by φ , noting that $\mathcal{D} \equiv (\{u_{l,t}\}, \{U_{l,t}\})$ and $\varphi \equiv (\nu, \omega, \{\alpha\})$. Our goal is to draw posterior samples for $q(t)$ through posterior samples from $\Pi(\varphi|\mathcal{D})$. After obtaining posterior samples for $q(t)$ through posterior samples from $\Pi(\varphi|\mathcal{D})$, we can estimate $q(t)$ by its posterior mean $\hat{q}(t)$ and obtain the posterior intervals as a measure of the uncertainty of estimation (see Gelman et al.¹⁷ for more on Bayesian inference).

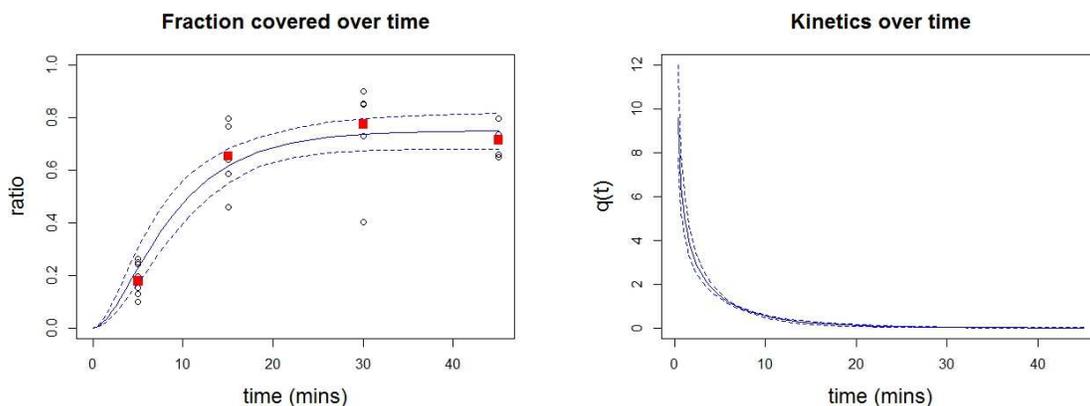
We first assume a flat hyper-prior distribution for the log-transformed parameters so that the posterior estimates correspond with the maximum likelihood estimates. i.e.,

$$\Pi(\log(\nu), \log(\omega), \log(\alpha)) \propto \mathbf{1}.$$

Table 1.1 shows the posterior estimates for the parameters, and Figure 1.8 shows the fitted area-

parameters	estimates	lower	upper
ν	94.61	50.56	165.15
ω	8.82	5.87	12.90
α	0.16	0.13	0.21

Table 1.1: Posterior statistics for the kinetics parameters



(a) Proportion covered over time. The circles are $u_{l,t}/U_{l,t}$, the red squares are $S_t^{obs}(t)/U_t^{obs}$, the solid line is the estimated $E(S(t))/S^*$, and the broken lines are 95% posterior intervals.

(b) The estimated kinetics along with 95% posterior intervals (assuming $\beta = 1$).

Figure 1.8: Intensity and Kinetic plots based on the Graphene data.

coverage and kinetics (where for the sake of plotting the shape, we have assumed $\beta = 1$). The posterior samples are obtained using the *MCMCpack* package³⁵ in R⁴³, and the diagnostic plots are shown in Appendix A.2.

1.4.2 SENSITIVITY TO SAMPLING DESIGN

Based on the estimated observation model, we are also interested in determining the sensitivity of the estimation procedure to the choice of sampling design. Specifically, we would like to know if and how the number of images, the similarity of resolutions across images, the number of time points chosen, and the percentage of missing data (unobserved tray portion) affect inference. This

information will be used in designing optimum sampling plans for new experiments.

In determining the effect of the above sampling factors, we turn to simulations. In the simulations, we assume the data generating model is the same model given in (1.23) and is estimated using the Bayesian methodology as described in Section 1.4.1. We also assume the following sampling design features:

1. The number of time points is n_T , and $n_T \in 4, 5, \dots, 40$. Note that we select equi-spaced time points on the log scale, so as to finely capture the initial moments of the growth process.
2. The fraction of missing data, U_t^{mis}/S^* , is ξ_{mis} , and $\xi_{mis} \in [0.1, 0.99]$.
3. The number of images at every time point throughout a single growth process is n_p , and $n_p \in 4, 5, \dots, 20$.
4. The image sizes follow a Dirichlet distribution, i.e.,

$$(U_{1,t}, U_{2,t}, \dots, U_{n_p,t}) \sim \text{Dirichlet}(\gamma \cdot \mathbf{1}_{n_p}),$$

where $\mathbf{1}_{n_p}$ denotes a vector of 1's, of length n_p . Note that increasing γ increases the similarity of resolutions (i.e. the balance of sizes across images at time t). We assume that $\log(\gamma) \in [0.5, 2.5]$.

A full factorial design for the four factors listed above is computationally expensive due to the curse of dimensionality. For instance, a full factorial design for four factors at ten levels each requires 10^4 runs. In order to reduce the number of runs, while maintaining space filling properties for each factor, we make use of approximate Latin Hypercube Sampling (LHS) designs^{36,37}, with the approximation due to integer rounding for the discrete factors. Specifically, we generate a 600-run LHS design for the four factors above, using the *SLHD* package¹ in R. The metric we use to determine

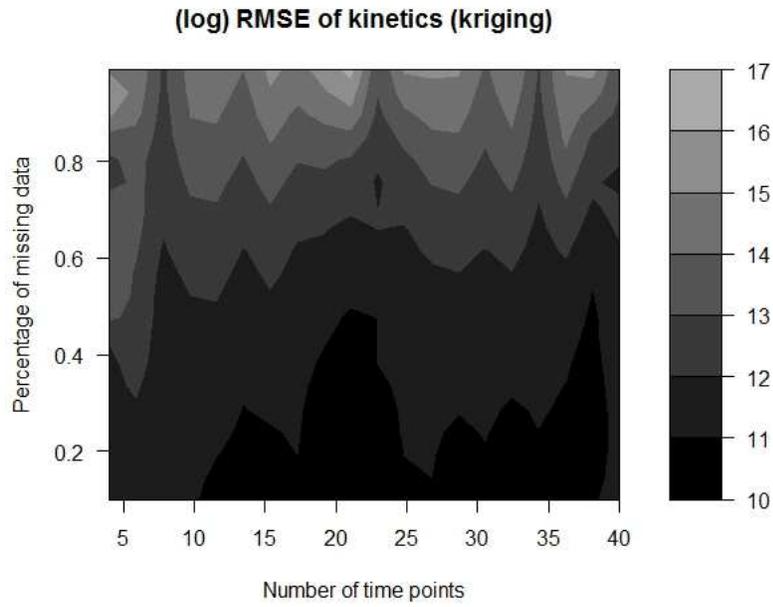
sensitivity at each run is the posterior RMSE (root mean square error) defined as

$$RMSE = \sqrt{\frac{1}{n_T \mathcal{M}} \sum_{k=1}^{n_T} \sum_{m=1}^{\mathcal{M}} (\hat{q}_m(t_k) - q(t_k))^2},$$

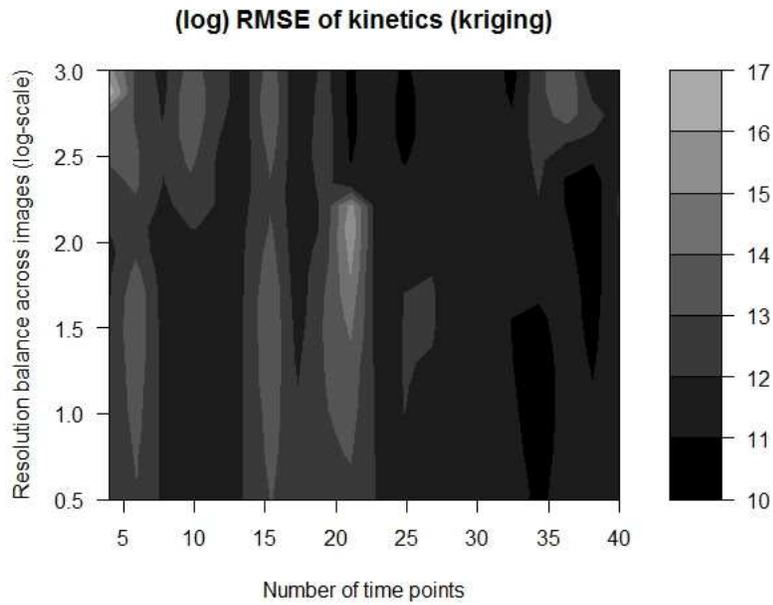
where $\mathcal{M}(= 5)$ denotes the number of replications at each factor combination. Since LHS designs generate irregularly spaced domain points, some smoothing is needed to interpolate unsampled domain points on a grid. In order to produce all two factor interactions, we apply a kriging interpolator on the calculated posterior RMSE for each two factor combination using the *geoR* package⁴⁶ in R, using the default Matérn covariance function. All the two factor interaction plots are shown in Figures 1.9, 1.10, and 1.11.

REMARKS

1. Figures 1.9a, 1.9b and 1.10a, which depict the two-factor joint effect of the number of time points ($n_T = 4, \dots, 40$) with the other factors, suggest that a higher number of time points will in general lead to better inference.
2. Figures 1.9a, 1.10b and 1.11a, which depict the two-factor joint effect of the fraction of missing data ($\xi_{mis} \in [0.1, 0.99]$) with the other factors, suggest that a lower fraction of missing data will also lead to better inference. In fact, judging from the plots, the fraction of missing data seems to be the most relevant factor in how good the inference procedure will be, and its effect changes depending on the values of the other factors.
3. Figures 1.10a, 1.11a and 1.11b show the two-factor joint effect of the number of images ($n_p \in 4, 5, \dots, 20$) with the other factors. The plots, especially Figure 1.11a, suggest that a lower number of images generally leads to slightly improved inference, with the effect of the number of images being somewhat weak in comparison to the effect of the other factors.

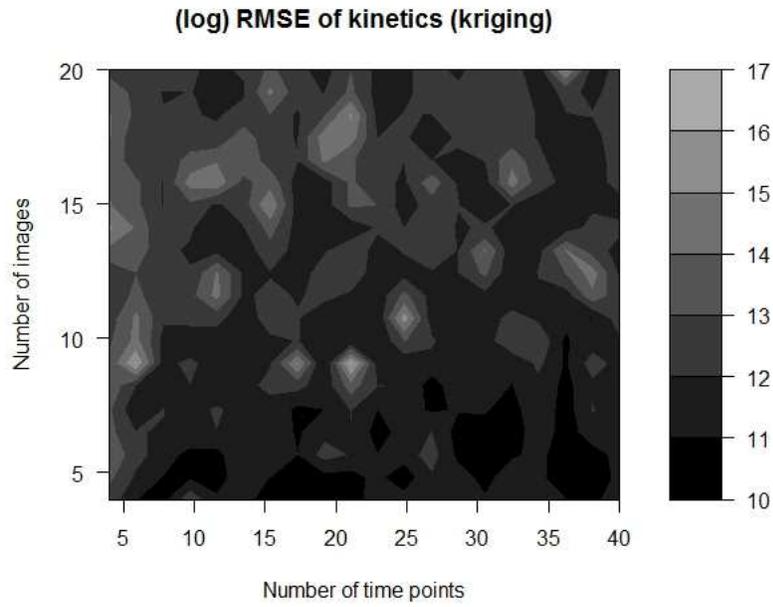


(a) Joint effect of n_T and ϱ_{mis} .

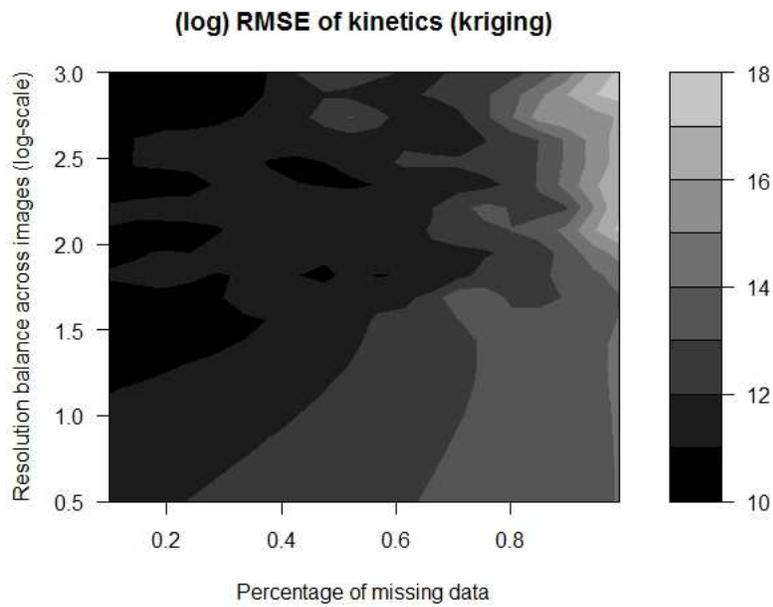


(b) Joint effect of n_T and γ .

Figure 1.9: Two factor interactions. Darker shades represent smaller RMSEs (i.e. lower estimation uncertainty), and are desirable.

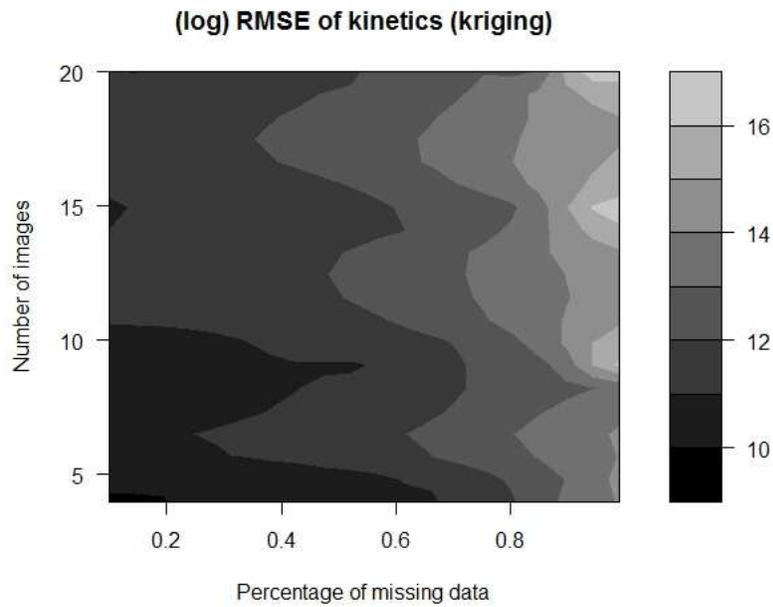


(a) Joint effect of n_T and n_p .

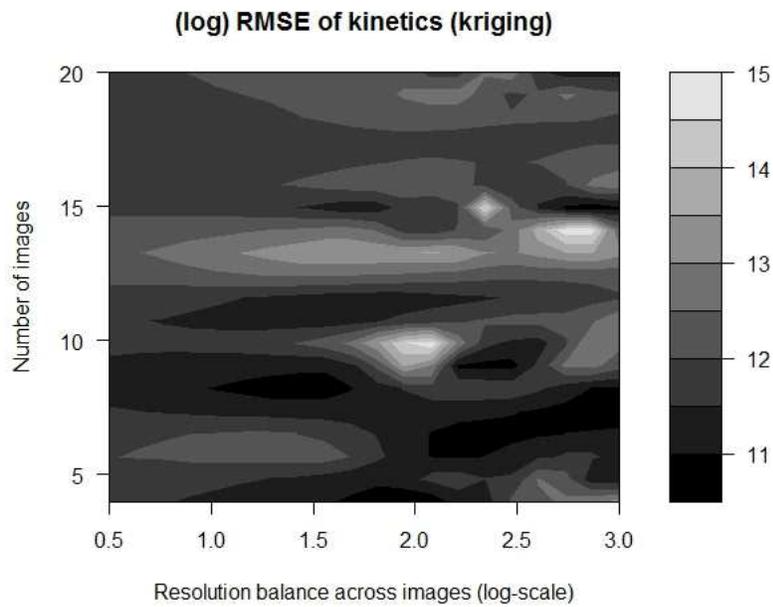


(b) Joint effect of γ and ϵ_{mis} .

Figure 1.10: Two factor interactions. Darker shades represent smaller RMSEs (i.e. lower estimation uncertainty), and are desirable.



(a) Joint effect of n_p and e_{mis} .



(b) Joint effect of γ and n_p .

Figure 1.11: Two factor interactions. Darker shades represent smaller RMSEs (i.e. lower estimation uncertainty), and are desirable.

4. Figures 1.9b, 1.10b and 1.11b show the two-factor joint effect of the resolution balance ($\log(\gamma) \in [0.5, 2.5]$) with the other factors. Figure 1.9b does not yield much information, but the other two plots (especially Figure 1.10b) imply that inference improves with more balanced resolutions across the images.

Judging from the plots in Figures 1.9, 1.10, and 1.11, the fraction of missing data seems to be the most important factor, followed by the number of time points, the number of images and the resolution balance across the images. Combining this with the four interpretations above, we thus recommend observing the whole tray at as many time points as possible. If observation of the entire tray is not feasible, then the observable region should be partitioned into as few images as possible, while those images should be at approximately the same resolution.

1.5 DISCUSSIONS AND FUTURE EXTENSIONS

In this paper, we have developed a stochastic framework for understanding the kinetics driving graphene growth by proposing models for the observed shape, time-varying area of each individual island, as well as the nucleation time distribution, while ensuring each component model reflects practical considerations and expectations. The resulting inference and sensitivity analyses show the stability of the model, and suggest some new sampling designs for graphene growth experiments.

It is worthwhile to point out that there are a lot of unsettled issues such as the co-existence of multiple growth mechanisms and multi-layer growth etc., but we believe our work provides a starting point to address these issues. Specifically, the supersaturation assumption, considered debatable by some experts, has actually been relaxed in our work (Eqs. 10 and 11), in anticipation of the growth conditions affecting the final graphene coverage. Further, we have used the single growth mechanism only as a description of the average nucleation behavior – the actual nucleation behavior at any site on the copper tray is assumed to be random (equation 4), thus ensuring a simple class

of multiple growth mechanisms for each graphene island. More sophisticated classes of multiple growth mechanisms, based on more detailed information of the copper foil's roughness, are certainly possible.

The graphene process can simultaneously have multi-layer growth. The proposed modeling framework handles the growth of the first layer (right on top of the copper foil) and describes how the first layer covers the foil over time. Since only the data of the first layer are collected for modeling, the secondary and tertiary layers do not affect the modeling of the first layer. Thus, one limitation of our final model is that it can only handle monolayer (two-dimensional) graphene growth. Expanding the model to describe multi-layer (three-dimensional) growth is an important research direction that we plan to explore.

We now propose a couple of model extensions/modifications for richer experimental datasets.

1.5.1 OBSERVATIONAL MODEL FOR ENTIRE TRAY

If instead of $u_{l,t}$, we observe $S_l(t)$ as per our recommendation in Section 1.4.2, then we can update the observational level model based on the following considerations

1. $S_l(t)$ should be non-decreasing over time.
2. The variance of $S_l(t)$ should approach zero when $E(S_l(t)) \rightarrow S^*$ and when $E(S_l(t)) \rightarrow 0$.

These are the boundary conditions of the growth process.

3. The models proposed in Sections 1.2.1-1.2.3 remain valid.

We can thus propose

$$\begin{aligned} \frac{S_l(t)}{S^*} | S_l(\infty) &\sim \frac{S_l(\infty)}{S^*} \cdot DP(\nu, (1 - \exp(-\alpha t))^2) , \\ \frac{S_l(\infty)}{S^*} &\sim \text{Beta}(\nu(1 - \exp(-\alpha\omega)), \nu \exp(-\alpha\omega)) , \end{aligned}$$

where $DP(\nu, H(\cdot))$ denotes the Dirichlet process distribution with concentration parameter ν and base function $H(\cdot)$. This observational level model captures the random behavior of carbon atom attachment to the edges of graphene islands, as well as ensuring that temporal dependence is accounted for. It is thus recommended for inference when the whole tray is observed.

1.5.2 GROWTH AS A FUNCTION OF PROCESS CONDITIONS

We can expand the component models to include covariate information, representing the different process conditions (e.g., growth temperature, methane flow rate, and ambient pressure). We can do this by first considering graphene growth on multiple trays, and introduce an extra index, $j \in \{1, \dots, J\}$, representing each tray. We then explicitly connect each tray specific α with \mathbf{X}_j , the vector of covariates describing the process condition for tray j . i.e., we use $\alpha_j = \alpha(\mathbf{X}_j)$, some positively valued function of the covariates \mathbf{X}_j for tray j . We can then write $q_j(t)$ as $q(t|\mathbf{X}_j)$ by simply replacing α_j with $\alpha(\mathbf{X}_j)$, thus explicitly showing the dependence of the kinetics on the process covariates \mathbf{X}_j .

Since $\alpha_j > 0$, one natural way of modeling $\alpha(\mathbf{X}_j)$ is to use the log-linear model,

$$\log(\alpha(\mathbf{X}_j)) = \sum_{k=1}^K a_k \cdot b_k(\mathbf{X}_j) .$$

where the basis functions $b_k(\cdot)$ can be determined through some exploratory data analysis.

It is our expectation that future experimental data will lead to improvements on the component models, especially the log-linear model for $\alpha(\cdot)$. Some future research directions include designing and evaluating experiments to obtain graphene data, along with process condition information.

*A noise annoys an oyster, But a noisy noise annoys an
oyster more.*

R.P. Weston and Bert Lee

2

Response Surface Optimization in the Presence of Internal Noise

IN MOST EXPERIMENTS that are conducted by engineers or physical scientists to study input-output relationships and to identify the optimal value(s) of the input variable(s) that optimize the response or the output variable, it is common to have some level of *internal noise*, typically associated with noisy input measurements. A well-studied field of statistics (more specifically, experimental design) is robust parameter design^{54,63}, where the goal is to choose the optimal levels of process inputs that make the response least sensitive to noise. Most of the work done by statisticians on robust design in

the 1980s and 1990s dealt with *external noise* and were motivated by manufacturing (e.g., automobile) applications. The role of internal noise in the optimization process was mostly limited to the so-called “tolerance design,” where the idea was to exploit a possible non-linear relationship between the response and the input factor with internal noise to determine the optimum setting of the input factor as well as a tolerance around that set value^{41,29}.

The fact that experiments in nanotechnology involve input factors that are prone to internal noise was noted and discussed by Dasgupta et al.ⁱⁱ, who treated the model-fitting and optimization under internal noise as two separate problems. We will now illustrate such an approach through the following toy example. Assume that the experimenter is interested in maximizing the following response surface with a quadratic mean function

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon, \quad (2.1)$$

where X is a single input variable (or control factor as in robust design literature) like temperature of a furnace, and ε represents the additive effect of the external noise that can neither be controlled nor measured during experimentation. Now suppose the input X (e.g., actual temperature) is subject to internal noise i.e., X is itself a random function of a true input Z (e.g., the set temperature). Then we can write, for example,

$$X = g(Z) + v, \quad (2.2)$$

where $g(\cdot)$ is a deterministic, invertible function, and v denotes the additive internal noise. Then we can rewrite (2.1) as

$$Y = \beta_0 + \beta_1 g(Z) + \beta_2 g^2(Z) + \beta_1 v + 2\beta_2 v g(Z) + \beta_2 v^2 + \varepsilon. \quad (2.3)$$

The optimization problem therefore boils down to finding a Z^* that maximizes $E(Y)$ while keep-

ing $\text{var}(Y)$ within acceptable limits. Such an optimization problem can also be formulated in terms of a suitable performance measure, like the signal-to-noise ratios proposed by Taguchi⁵⁴ or other measures by Leon et al.²⁸.

Data from experiments involving internal noise are typically of the form $(Y_i, Z_i)_{i=1, \dots, n}$, permitting experimenters to fit models connecting Y directly to Z , i.e., (2.3). Information on the internal noise model as (2.2) are often obtained from surrogate experiments or observational studies that yield pairwise data $(X_j, Z_j)_{j=1, \dots, m}$. The approach adopted by Dasgupta et al.¹¹ consisted of the following two steps: (i) fitting the response surface model (2.3) using the data on (Y, Z) , and (ii) subsequently incorporating the surrogate information about (2.2) obtained from the data on (X, Z) for optimization of (2.3) using Monte-Carlo simulations. Clearly, in this approach, the fitted response model does not take into account the impact of internal noise. Thus the response surface modeling and its optimization are treated as two separate problems. It is also obvious from the form of (2.3), that even in a relatively simple scenario, the model linking Y to Z may not be straightforward to estimate. It is therefore natural to explore a unified approach for modeling and optimization of the response function under internal noise. Such an approach that permits integration of the available information on internal noise into the modeling and optimization of the response function has not been studied so far.

It is also worthwhile to note that, with some additional effort (e.g., extra cost and time), it is sometimes possible to collect data in the form of triplets (Z_i, X_i, Y_i) from the same experiment, as in Remillard et al.⁴⁵. Intuitively it seems that such data should be more helpful in accounting for internal noise in the optimization problem, but quantification of such benefits is not possible without a unified framework as described earlier.

In this paper, we propose a unified Bayesian approach to response surface optimization in the presence of internal noise. Our key ideas and contributions include: (a) Finding a suitable, unit-free, interpretable performance measure to quantify the strength of internal noise, (b) postulating a hi-

erarchical model that captures the mutual relationship among Y , X and Z , (c) Identifying suitable objective functions or performance measures consistent with the overall goal of optimizing the response function, (d) Proposing methods for estimating the performance measures from available experimental data using posterior estimates of model parameters, (e) Comparing different performance measures using Monte Carlo simulations and (f) Studying the effect of ignoring available information on internal noise while solving the optimization problem.

In order to do this, we first describe a motivating example in Section 2.1 from a Carbon nanotube (CNT) alignment experiment. In Section 2.2, we propose a unit-free summary measure of internal noise, that can be interpreted as the “strength” of the internal noise, and helps design simulation studies by varying the level of internal noise. In Section 2.3 we propose a Bayesian framework for performing response surface optimization in the presence of internal noise. In particular, we define and evaluate appropriate objective functions to be optimized, and discuss their estimation. In Section 2.4, we perform extensive simulations to compare the proposed objective functions. Section 2.5 explores the effect of ignoring internal noise on the correctness of identification of the optimal solution. We apply the results of our analyses to the experimental data in Section 2.6, and some concluding remarks are presented in Section 2.7.

2.1 RESEARCH MOTIVATION: CNT ALIGNMENT

Scientific interest in carbon nanotubes (CNTs) has risen over the past several decades, primarily due to their versatility as a result of properties such as high electrical and thermal conductivity, high mechanical strength, and optical anisotropy. For example, CNTs play a primary role in flexible electronics⁴⁰, energy storage devices¹², optical displays^{27,32}, chemical sensors¹⁵, and water treatment technologies^{10,58}. Due to the extreme aspect ratio of CNTs, transferring these unique properties from the nano-scale to industrial-scale products requires their orientation in a single direction. Remillard et al.⁴⁵ utilize an external electric field, one of the most promising techniques for industrial-scale

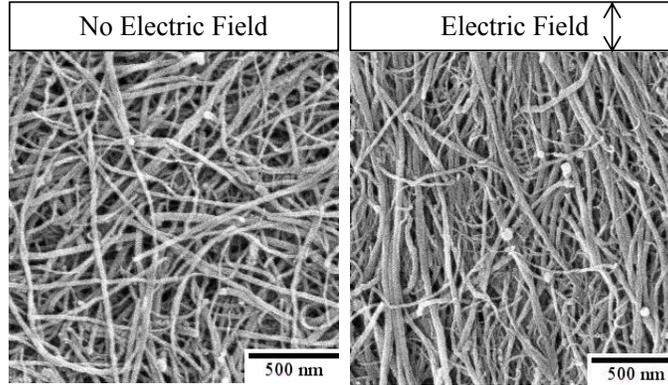


Figure 2.1: Examples of randomly oriented CNT and CNT aligned in the presence on an electric field.

alignment, to orient CNTs, and their work provides a basis for investigating the role of internal noise.

In their paper, Remillard et al. ⁴⁵ conduct an experiment to determine the optimal level of electric field strength that maximizes alignment of CNTs. Figure 2.1 shows an SEM image of multi-walled carbon nanotubes aligned using a combination of these factors compared to a randomly oriented sample ⁴⁵.

Experiments were also conducted by simultaneously varying two other factors in addition to electrical field strength (X): CNT mass (\mathcal{M}), and volume of suspension fluid (measured by its inverse I). Initial exploration of the factor space led to determination of a response model of the form

$$R = \psi(X, \mathcal{M}, I) + \varepsilon,$$

where R is a response that measures alignment, $\psi(\cdot)$ is a function of the input variables, and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ denotes the observational noise. The expected response function was optimized to determine the desired levels of \mathcal{M} , I and X . However, an important aspect that was not considered in this rudimentary optimization approach was the presence of internal noise making X a random variable. In this experiment, the electrical field was generated amplifying a voltage input to some desired tar-

get strength and applying it to the CNT samples suspended in fluid. This electric field amplification process is prone to random fluctuations around the desired value (due to the machine), and thus the amplified electric field, as an input to the alignment process, is inherently noisy. Figure 2.2 shows the relationship between original input voltages (Z) and their corresponding electric field strength (X) from data generated during the experiment, as well as from supplementary (pre-experiment and post-experiment) data. Therefore, the goal of the experiment is to determine the optimal value of Z that maximizes some desired objective function associated with achieving a desired level of alignment.

From the figure, it appears that X possess a mean-dependent variance, leading to a heteroscedastic and asymmetric marginal distribution for R . Since X can only be controlled through its noisy link with Z , the question that naturally comes to mind is, should such heteroscedastic and asymmetric internal noise taken into account while determining the optimal solution? If so, how?

2.2 HOW SEVERE IS THE INTERNAL NOISE? A UNIT-FREE MEASURE

In order to understand the magnitude of internal noise before any experiments are planned, a unit-free and response-independent measure for determining the internal noise severity is needed. We now define one such measure, based only on the available (X, Z) pairs. This means that it is possible to determine, based on supplementary data only, if it will be necessary to account for internal noise in the main experiment, or if it is relatively safe to ignore.

Intuitively, any good unit-free measure for determining the level of internal noise should compare some function of dispersion before and after the internal noise occurs (i.e., dispersion of X under the assumption of zero-noise, relative to the true dispersion of X). The dispersion of X under the assumption of zero noise is directly related to the dispersion of Z , which is itself governed by the experimental design, i.e., the experimenter's choice of the input points Z . Such a design can be visualized a probability density function $f(z)$ that represents the uncertainty associated with the

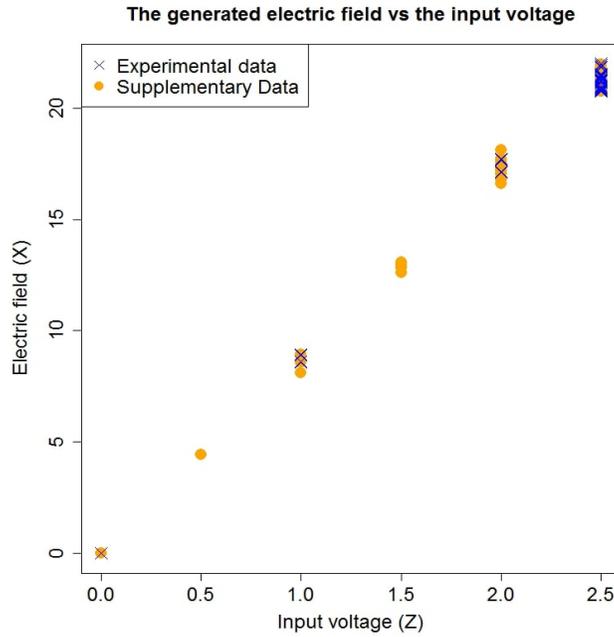


Figure 2.2: The original (generated) voltage (in volts) versus the amplified voltage (in volts).

selection of z from Ω_Z (the domain of Z 's). Suppose the experimenter is considering choosing the input points in the closed interval $[a, b]$. Consider the following two examples: (i) three levels of Z are chosen at a, c, b , where $a < c < b$, (ii) Z is chosen uniformly in the interval $[a, b]$. In the former case, we can write

$$f_1(z) = \begin{cases} 1/3, & z = a, c, b \\ 0 & \text{otherwise.} \end{cases}$$

In the latter case,

$$f_2(z) = \frac{1}{b-a} \mathbb{I}(a \leq z \leq b),$$

where $\mathbb{I}(\cdot)$ denotes the indicator function.

Let $\sigma_f^2(X, \tau, \gamma)$ denote the variance of X , where $f(\cdot)$ represents the experimental design, γ a parameter (scalar or vector) associated with $E(X|Z)$ in model (2.2), and τ a variance parameter, typi-

cally associated with the variance of the internal noise v in (2.2), that measures the degree of internal noise. Thus,

$$\sigma_f^2(X, \tau, \gamma) = E_f \text{var}(X|Z) + \text{var}_f E(X|Z), \quad (2.4)$$

Then a reasonable unit-free measure of internal noise is:

$$S_f(\gamma, \tau) = 1 - \frac{\sigma_f(X, 0, \gamma)}{\sigma_f(X, \tau, \gamma)}, \quad (2.5)$$

where $\sigma_f^2(X, \tau, \gamma)$ is given by (2.4). Note that $S_f(\gamma, \tau)$, as defined by (2.5), is always bounded between 0 and 1, with higher values representing higher internal noise levels. To illustrate how this measure can be computed analytically, assume that in model (2.2), $g(z) = \gamma Z$ and $\text{var}(v) = \tau^2$, and let the pdf associated with the experimental design be $f_z(z) = \frac{1}{b-a} \mathbb{I}(a \leq z \leq b)$ as in the uniform case introduced earlier. Then, from (2.4), we have

$$\sigma_f^2(X, \tau, \gamma) = E_f(\tau^2) + \text{var}_f(\gamma Z) = \tau^2 + \gamma^2 \text{var}_f(Z) = \tau^2 + \frac{\gamma^2(b-a)^2}{12}.$$

Substitution of the above in (2.5) yields:

$$S_f(\gamma, \tau) = 1 - 1/\sqrt{1 + 12 \frac{\tau^2}{\gamma^2(b-a)^2}}.$$

In the case of heteroscedastic internal noise where we have a mean-dependent internal variance, $\text{Var}(X|Z) = \tau^2 Z$, we have that $E(\text{Var}(X|Z)) = (\tau^2/2)(a+b)$, and thus

$$S_f(\gamma, \tau) = 1 - 1/\sqrt{1 + 6 \frac{\tau^2(a+b)}{\gamma^2(b-a)^2}}.$$

In addition to being unit-free, $S_f(\gamma, \tau)$ enables direct comparisons of noise magnitude across different internal noise distributions. We make use of this property when comparing internal noise

magnitudes from our simulation studies in Section 2.4, and our CNT alignment model in Section 2.6.

Since γ and τ will have to be estimated in practice, we utilize a Bayesian framework to estimate $S(\gamma, \tau)$ by the posterior mean, $\hat{S} = E(S(\gamma, \tau) | \mathbf{X}^{obs}, \mathbf{Z}^{obs})$, with the estimation uncertainty described through posterior intervals. In the next Section we describe this Bayesian framework, and investigate the problem of optimizing a response surface in the presence of internal noise.

2.3 A BAYESIAN FRAMEWORK FOR RESPONSE SURFACE OPTIMIZATION IN THE PRESENCE OF INTERNAL NOISE

We now define a general relationship between a response Y , a single noisy factor X , and the actual input Z , which yields X as an intermediate value, using the following hierarchical model

$$\begin{aligned} Y|X, \vartheta &\sim \mathcal{F}(Y|X, \vartheta) \\ X|Z, \varphi &\sim \mathcal{G}(X|Z, \varphi) \\ (\vartheta, \varphi) &\sim \pi(\vartheta, \varphi), \end{aligned} \tag{2.6}$$

where the distributions $\mathcal{F}(\cdot)$ and $\mathcal{G}(\cdot)$ are parametrized by ϑ and φ respectively, and $\pi(\cdot)$ denotes a prior distribution for the parameters. Note that we use the term distribution loosely here, but more specifically refer to $\mathcal{F}(Y|X, \vartheta)$ and $\mathcal{G}(X|Z, \varphi)$ as density functions later on. We restrict the distribution \mathcal{G} such that $E(X|Z, \varphi)$, denoted by $g(Z)$ is an invertible function of Z .

Without loss of generality, let us assume henceforth that the main objective is the maximization of the response (the minimization problem is similar). In the absence of internal noise, $\mathcal{G}(\cdot)$ is degenerate and X is equivalent to Z . Then the maximization problem can be stated as

$$X^* = \arg \max_X E(Y|X, \mathbf{Y}^{obs}, \mathbf{X}^{obs}),$$

where $(\mathbf{Y}^{obs}, \mathbf{X}^{obs})$ denote the data observed from the experiment. The optimal Z , say Z^* , can be obtained by substituting X^* into the deterministic relation between Z and X .

Note that $E(Y|X, \mathbf{Y}^{obs}, \mathbf{X}^{obs})$ refers to the posterior predictive mean of Y for a given value of X , observed responses \mathbf{Y}^{obs} and corresponding observed inputs \mathbf{X}^{obs} , and is given by:

$$E(Y|X, \mathbf{Y}^{obs}, \mathbf{X}^{obs}) = \int E(Y|X, \vartheta) \pi(\vartheta|\mathbf{Y}^{obs}, \mathbf{X}^{obs}) d\vartheta,$$

where $\pi(\vartheta|\mathbf{Y}^{obs}, \mathbf{X}^{obs})$ denotes the posterior distribution of ϑ . To extend this optimization to the case where internal noise is present, we need to define an appropriate objective function $\eta(Z)$, and maximize such a function with respect to Z to obtain Z^* . There are many ways to do this and we explore the most promising ones next.

2.3.1 OBJECTIVE FUNCTIONS

One common objective function that can be borrowed from robust design literature is the signal-to-noise (S/N) ratio (i.e., the inverse coefficient of variation). Applying this to (2.6), we have the objective function

$$\eta_1(Z) = \frac{E(Y|Z, \mathbf{Y}^{obs}, \mathbf{X}^{obs}, \mathbf{Z}^{obs})}{\sqrt{Var(Y|Z, \mathbf{Y}^{obs}, \mathbf{X}^{obs}, \mathbf{Z}^{obs})}}. \quad (2.7)$$

Here, $(\mathbf{Y}^{obs}, \mathbf{X}^{obs}, \mathbf{Z}^{obs})$ denotes the experimental data, and will be henceforth denoted by \mathcal{D} . It is possible that the data are generated as triplets (Z, X, Y) from the same experiment, as reported in⁴⁵ or as two separate pairs (Y, Z) and (X, Z) from two different experiments/sources, as inⁱⁱ. The numerator and denominator of (2.7) are respectively the expectation and variance of the posterior predictive distribution of Y given the observed data \mathcal{D} . Thus, η_1 can be written more explicitly as:

$$\eta_1(Z) = \frac{\int E(Y|Z, \vartheta, \varphi) \pi(\vartheta, \varphi|\mathcal{D}) d\vartheta d\varphi}{\sqrt{\int Var(Y|Z, \vartheta, \varphi) \pi(\vartheta, \varphi|\mathcal{D}) d\vartheta d\varphi}}, \quad (2.8)$$

where $\pi(\vartheta, \varphi|\mathcal{D})$ is the posterior distribution of (ϑ, φ) given the data.

Note that this objective function is only appropriate when the expected posterior predictive response, $E(Y|Z, \mathbf{Y}^{\text{obs}}, \mathbf{X}^{\text{obs}}, \mathbf{Z}^{\text{obs}})$, is always non-negative over the domain of interest. The objective function, $\eta_1(Z)$, has the advantage of often having a closed form expression, and easy to calculate (e.g., when the conditional distributions in (2.6) are normal, it has closed-form expressions). However, several researchers have investigated the role of η_1 , defined in (2.7), as an objective function in robust design and have found it to be appropriate only under specific (often strong) assumptions. See, for example, Leon et al.²⁸, Box⁵, Nair³⁹, Bérubé & Wu². Our preliminary investigation (discussed later) shows that it can be highly sensitive to noise. Its dependence only on the first two moments of the conditional distribution is also a major drawback. Therefore, we explore alternative objective functions that are (i) meaningful from the practical standpoint, (ii) are based on higher-order moments of the conditional distribution of Y given (Z, \mathcal{D}) , and (iii) are less sensitive to the nature and magnitude of noise compared to the S/N ratio. One alternative objective function, which satisfies these three criteria, is

$$\eta_2(Z) = Pr(Y > y^* | Z, \mathcal{D}) = \int Pr(Y > y^* | Z, \vartheta, \varphi) \pi(\vartheta, \varphi | \mathcal{D}) d\vartheta d\varphi. \quad (2.9)$$

The criterion η_2 is the marginal probability of the response exceeding some predefined threshold, y^* . Intuitively, maximizing this marginal tail probability for some well chosen y^* should lead to larger response values overall. Another objective function, which is even better aligned with our objective of maximizing the response, is

$$\eta_3(Z) = E(I_{\{Y > y^*\}} Y | Z, \mathcal{D}) = \int E(I_{\{Y > y^*\}} Y | Z, \vartheta, \varphi) \pi(\vartheta, \varphi | \mathcal{D}) d\vartheta d\varphi. \quad (2.10)$$

which can be interpreted as the marginal expectation of the response which exceeds the pre-specified

threshold y^* . Although many good choices for y^* exist, one particularly appealing choice, due to our objective of maximizing the response, is to set

$$y^* = \max(E(Y|Z, \mathcal{D})) .$$

This choice of y^* should intuitively discriminate well among competing optimal points.

Although η_2 and η_3 enjoy several advantages over the S/N ratio, including lower sensitivity to internal noise, they pose some serious computational challenges, as we show and address in the next subsection.

2.3.2 COMPUTATION OF OBJECTIVE FUNCTIONS

The computation of the three objective functions given by (2.8), (2.9) and (2.10) involve computation of terms of the following form:

$$\int E[H(Y)|Z, \vartheta, \varphi] \pi(\vartheta, \varphi|\mathcal{D}) d\vartheta d\varphi, \quad (2.11)$$

where,

$$H(Y) = \begin{cases} Y & \text{corresponding to the numerator of (2.8)} \\ I_{\{Y>y^*\}} & \text{corresponding to (2.9)} \\ I_{\{Y>y^*\}}Y & \text{corresponding to (2.10)} \end{cases}$$

The computational challenges in (2.11) arise from:

1. The need to “integrate out” the parameters ϑ and φ , i.e., compute the average of $E[H(Y)|Z, \vartheta, \varphi]$ with respect to the posterior distribution $\pi(\vartheta, \varphi|\mathcal{D}) d\vartheta d\varphi$.
2. The evaluation of $E[H(Y)|Z, \vartheta, \varphi]$.

To address the first challenge, we by-pass the integration by multiplying the posterior density,

$\pi(\vartheta, \varphi|\mathcal{D})$, with zero everywhere, except at the mode. This essentially plugs in the posterior modes, denoted by $\hat{\vartheta}$ and $\hat{\varphi}$, and reduces (2.11) to $E\left(H(Y)|Z, \hat{\vartheta}, \hat{\varphi}\right)$. The effect of such approximations is studied through carefully designed simulation studies in Section 2.4.

To address the second computational challenge, first note that evaluation of the marginal expectation, $E[H(Y)|Z, \vartheta, \varphi]$ is often possible in closed-form when $H(Y) = Y$. However, When either $H(Y) = I_{\{Y>y^*\}}$ or $H(Y) = I_{\{Y>y^*\}}Y$, the calculations involve integrals which in general have to be numerically evaluated. To see this, we denote the indicator function for some event \mathcal{A} as $I_{\mathcal{A}}$, and obtain

$$\begin{aligned} E(H(Y)|Z, \vartheta, \varphi) &= E[E[H(Y)|X, Z, \vartheta, \varphi] |Z, \varphi] \\ &= \int_{-\infty}^{\infty} E[H(Y)|X, \vartheta] \cdot \mathcal{G}(x|Z, \varphi) dx . \end{aligned} \quad (2.12)$$

In general, the integral in (2.12) cannot be obtained in closed form expression, and thus must be evaluated numerically. This is computationally expensive for carrying out optimization, and we can simplify the calculations by instead considering Monte-Carlo approximations.

First note from the last line in (2.12), that $E[H(Y)|X, \vartheta]$, is the same for all Z , while the density function, $\mathcal{G}(X|Z, \varphi)$, changes with respect to Z . Thus the naive approach of first sampling a pre-determined number of input points, say N , from $\mathcal{G}(X|Z, \varphi)$, and then calculating the approximate quantity

$$E(H(Y)|Z, \vartheta, \varphi) \approx \frac{1}{N} \sum_{i=1}^N E(H(Y)|X = x_i, \vartheta) ,$$

every time Z changes, is computationally inefficient due to the unavoidably large number of times $\mathcal{G}(X|Z, \varphi)$ will need to be sampled from. A more efficient approach, inspired by importance sampling ideas³¹, involves a batch-sequential algorithm:

- i. Sample a single representative set of N X 's, denoted by \mathcal{X}_N , from some distribution indepen-

dent of Z , denoted by $\mathcal{H}(X|\varphi)$.

2. Calculate the approximate quantity

$$E(H(Y)|Z, \mathfrak{g}, \varphi) \approx \frac{\sum_{x \in \mathcal{X}_N} E(H(Y)|X = x, \mathfrak{g}) \cdot \omega(x|Z, \varphi)}{\sum_{x \in \mathcal{X}_N} \omega(x|Z, \varphi)}.$$

where

$$\omega(x|Z, \varphi) = \frac{\mathcal{G}(x|Z, \varphi)}{\mathcal{H}(x|\varphi)}$$

In order to ensure \mathcal{X}_N is truly representative of all possible X 's over the domain of Z 's, denoted by Ω_Z , we can define $\mathcal{H}(X|\varphi)$ as the mixture

$$\mathcal{H}(X|\varphi) = \int_{\Omega_Z} \mathcal{G}(X|Z = z, \varphi) \cdot f(z) dz, \quad (2.13)$$

where $f(z)$ is a probability density function (doubling as a weight function) representing prior belief of observing z from Ω_Z . Note that a close-form expression for $\mathcal{H}(X|\varphi)$ in (2.13) is not always available and therefore a further approximation is needed. Denoting a set of N Z 's sampled from $f(z)$ by \mathcal{Z}_N , we can write

$$\mathcal{H}(X|\varphi) \approx \frac{1}{N} \sum_{z \in \mathcal{Z}_N} \mathcal{G}(X|Z = z, \varphi).$$

This means any member of \mathcal{X}_N can be generated by first sampling Z from Ω_Z using $f(z)$, and then sampling a single X from $\mathcal{G}(X|Z, \varphi)$. Ideally, $f(z)$ will be based on the specific experimental design used to generate the input locations Z within the domain, Ω_Z . If instead, the design is not known or specified, but all points within a finite size Ω_Z (i.e., $|\Omega_Z| < \infty$) are equally permissible in

the design, then we can set $f(z)$ to be the uniform density over Ω_Z , i.e.,

$$f(z) = \frac{1}{|\Omega_Z|}.$$

Since the paired representative set $\{\mathcal{Z}_N, \mathcal{X}_N\}$ only need to be sampled once for each estimated model, there is some leeway for choosing N as high as we want, and the law of large numbers ensures that as $N \rightarrow \infty$, the difference due to approximations go to zero.

2.4 SIMULATION STUDIES FOR ASSESSING SENSITIVITY OF OBJECTIVE FUNCTION TO NOISE

We now assess the performance of each objective function under different magnitudes of internal noise, using simulations. We consider two scenarios for the data generating mechanism: (i) homoscedastic internal noise, and (ii) a special case of heteroscedastic internal noise, using a mean dependent variance.

2.4.1 SIMULATION I: HOMOSCEDASTIC INTERNAL NOISE

Consider the following data generating model

$$\begin{aligned} Y|X &\sim \mathcal{N}(\mu_X, \sigma^2) \text{ where } \mu_X = \beta_0 + \beta_1 X + \beta_2 X^2 \\ X|Z &\sim \mathcal{N}(\gamma Z, \tau^2) \\ (\mathfrak{D}, \varphi) &\sim \pi(\mathfrak{D}, \varphi) \propto \mathbf{1}. \end{aligned} \tag{2.14}$$

where $\mathfrak{D} = (\beta_0, \beta_1, \beta_2, \sigma)$, with $\beta_2 < 0$ (so that the optimization corresponds to maximization), and $\varphi = (\gamma, \tau)$. In order to visualize the three objective functions, we first need to set up a data generation mechanism:

- i. Set Ω_Z to the interval $[0, 0.1]$.

2. Fix the following true values for the parameters

$$\beta_0 = 200, \beta_1 = 1, \beta_2 = -1, \sigma = 0.05, \gamma = 10, \text{ and } \tau \in \{0, 5\},$$

with the two different values of τ representing zero internal noise ($S(\gamma, \tau) = 0$), and high internal noise ($S(\gamma, \tau) \approx 0.94$) respectively.

3. Generate 50 triplets of (Z, X, Y) , with a further 50 tuples of (Z, X) representing supplementary data.

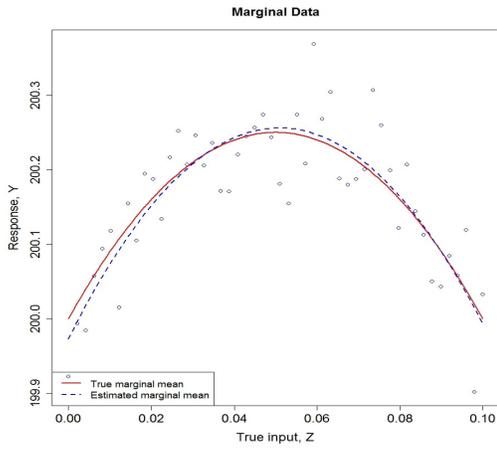
As mentioned earlier, the objective functions, η_2 and η_3 , do not have closed form expressions.

However, we can write the objective function, η_1 , in closed form (see Appendix B.1 for details) as

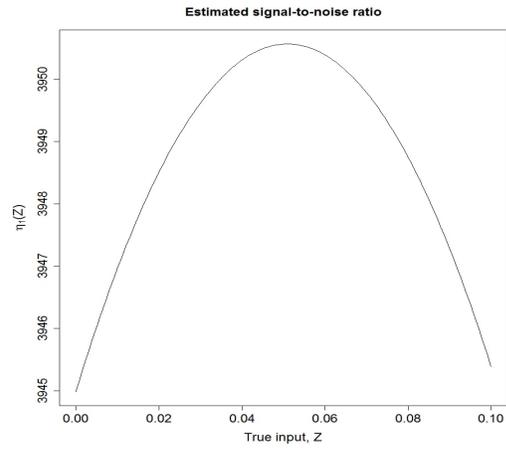
$$\eta_1(Z) = \left((\hat{\beta}_0 + \hat{\beta}_2 \hat{\tau}^2) + (\hat{\beta}_1 \hat{\gamma})Z + (\hat{\beta}_2 \hat{\gamma}^2)Z^2 \right) / \sqrt{\hat{\sigma}^2 + 2\hat{\beta}_2^2 \hat{\tau}^2 \left(\hat{\tau}^2 + 2 \left(\hat{\gamma}Z + \frac{\hat{\beta}_1}{2\hat{\beta}_2} \right)^2 \right)}. \quad (2.15)$$

The generated data, true model, estimated fit as well as the three objective functions are shown under zero-internal noise in Figure 2.3, and under high-internal noise in Figure 2.4.

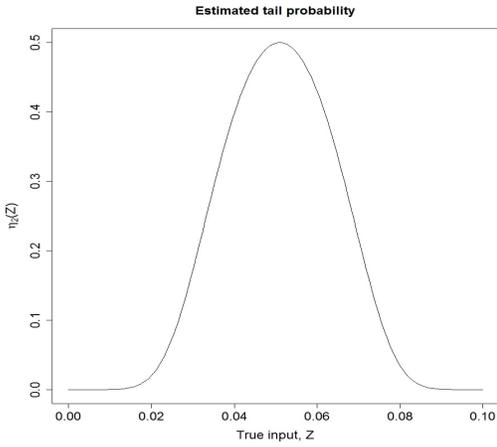
As expected, Figure 2.3a and Figure 2.4a show that high internal noise in the data generation mechanism leads to loss of the signal. i.e., the response simply ceases showing any relationship to the input. The rest of the figures show severe distortions of η_1 and η_2 as the internal noise is increased. $\eta_1(Z)$ and $\eta_2(Z)$ seem particularly sensitive to the underestimation of γ which occurs as the internal noise drowns out the internal signal. This explains their inclination towards the right boundary of Ω_Z . η_3 , on the other hand, appears to be the most reliable for finding the true optimum at $-\beta_1/2\gamma\beta_2 = 0.05$, when the internal noise level becomes high.



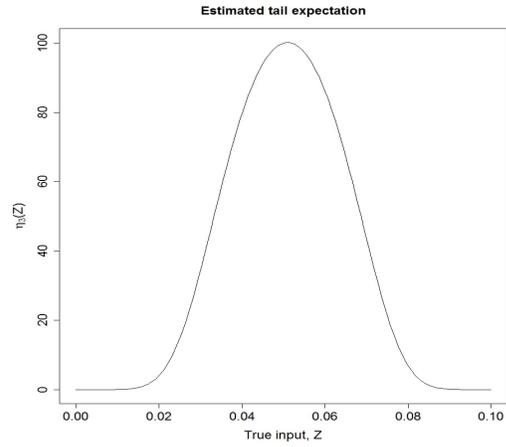
(a) The generated data, along with the true and estimated marginal response curves.



(b) The S/N ratio.

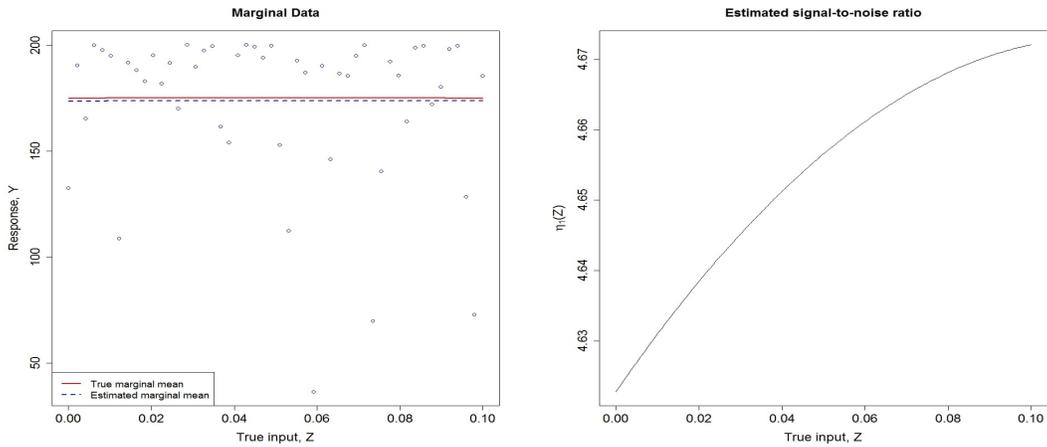


(c) The tail probability.



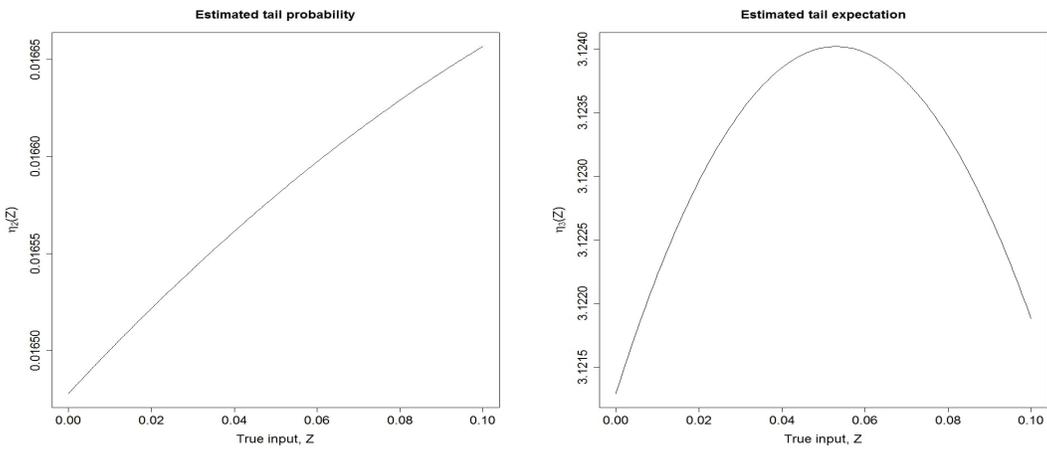
(d) The tail expectation.

Figure 2.3: Plots for zero internal noise ($\tau = 0$ and $S(\gamma, \tau) = 0$).



(a) The generated data, along with the true and estimated marginal response curves.

(b) The S/N ratio.



(c) The tail probability.

(d) The tail expectation.

Figure 2.4: Plots for high internal noise ($\tau = 5$ and $S(\gamma, \tau) \approx 0.94$).

2.4.2 SIMULATION 2: HETEROSCEDASTIC INTERNAL NOISE

For an example of the heteroscedastic case which is more representative of the CNT data explored in Section 2.1, consider the same data generating model as in (2.14) except for the constant internal

variance, which we now choose to be a linear function of the mean. i.e., we now use

$$X|Z \sim \mathcal{N}(\gamma Z, \tau^2 Z).$$

We also use the same data generation mechanism as for the homoscedastic case in Section 2.4.1, except that $\tau \in \{0, 2.25\}$, with the two different values of τ representing zero internal noise ($S(\gamma, \tau) = 0$), and medium internal noise ($S(\gamma, \tau) \approx 0.5$) respectively.

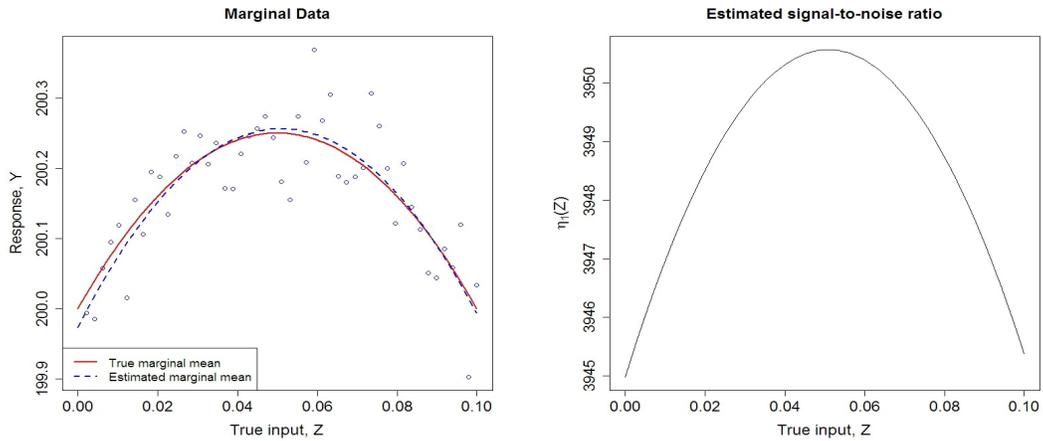
Replacing every instance of τ^2 in (2.15) with $\tau^2 Z$, gives the following closed form expression for the first objective function

$$\eta_1(Z) = \left(\hat{\beta}_0 + (\hat{\beta}_1 \hat{\gamma} + \hat{\beta}_2 \hat{\tau}^2) Z + (\hat{\beta}_2 \hat{\gamma}^2) Z^2 \right) / \sqrt{\hat{\sigma}^2 + 2 \hat{\beta}_2^2 \hat{\tau}^2 Z \left(\hat{\tau}^2 Z + 2 \left(\hat{\gamma} Z + \frac{\hat{\beta}_1}{2 \hat{\beta}_2} \right)^2 \right)}.$$

Independent of the underestimation of γ , the internal noise level now directly factors into the calculation of the true optimum, since the optimal point at $\tilde{z}^{(1)} = -(\hat{\beta}_1 \hat{\gamma} + \hat{\beta}_2 \hat{\tau}^2) / 2 \hat{\gamma}^2 \hat{\beta}_2$ represents a shrinkage towards zero, with the degree of shrinkage being directly proportional to the noise level. This shrinkage counterbalances and overwhelms the effect of the underestimation of γ , which invariably happens when the noise level is increased.

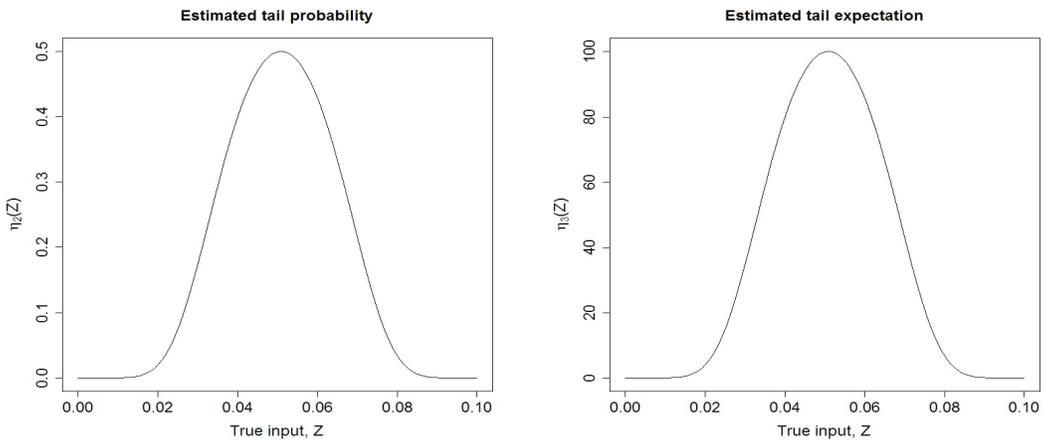
From Figure 2.5a and Figure 2.6a, we can see the effect of the internal heteroscedasticity on the response. i.e., the observed variance increases as Z increases, and the signal is weakened and distorted, but not completely lost. Figure 2.5b and Figure 2.6b show the somewhat acute sensitivity of the signal-to-noise ratio, $\eta_1(Z)$, to the noise level. Particularly, as τ is increased, $\eta_1(Z)$ quickly suggests an optimal point at $Z = 0$. The rest of the figures depict similar behavior for $\eta_2(Z)$ and $\eta_3(Z)$ when the noise level is increased, i.e., generally reduced maximum values, and a general shrinkage towards zero which is not as acute as $\eta_1(Z)$'s shrinkage. This makes $\eta_2(Z)$ and $\eta_3(Z)$ more reliable than the signal-to-noise ratio, $\eta_1(Z)$, in this scenario where the internal variance is a linear function of the in-

ternal mean function. In particular, the estimated tail expectation, $\eta_3(Z)$, has proven to be reliable in the homoscedastic case as well and will thus be the objective function of choice in the rest of this paper.



(a) The generated data, along with the true and estimated marginal response curves.

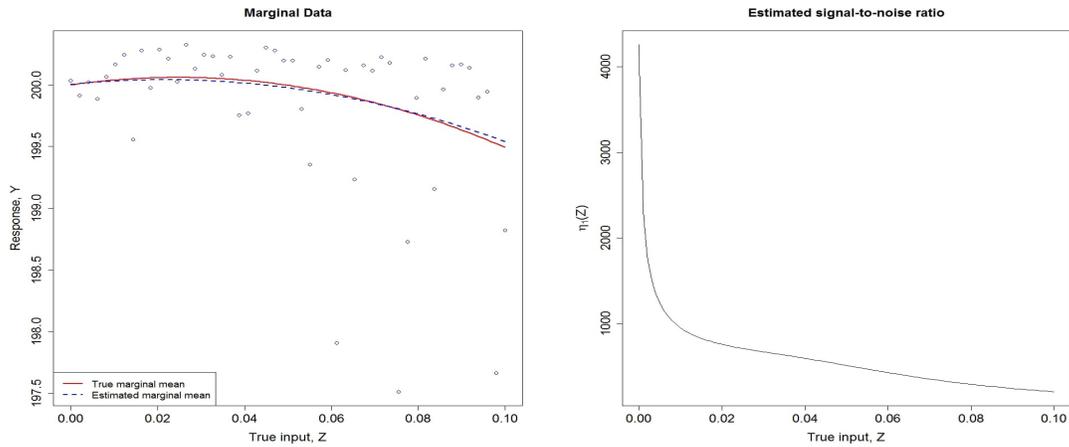
(b) The S/N ratio.



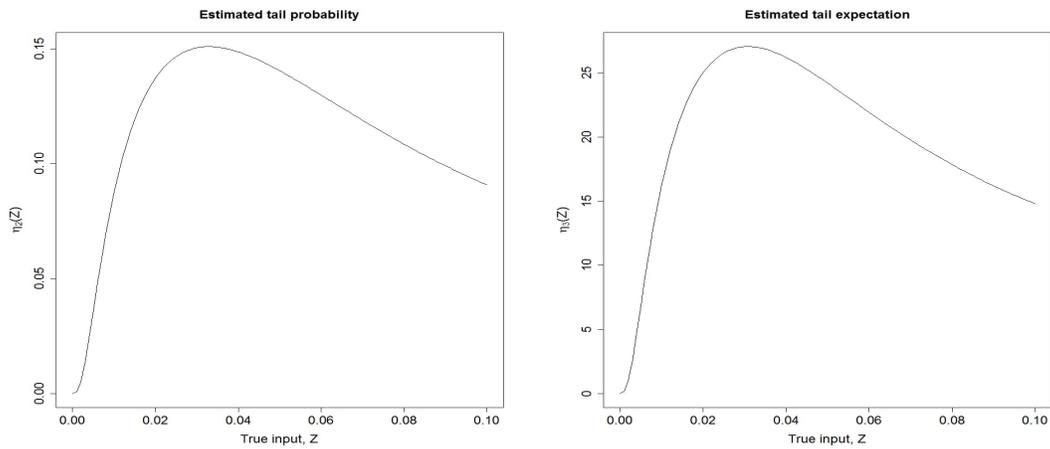
(c) The tail probability.

(d) The tail expectation.

Figure 2.5: Plots for zero internal noise ($\tau = 0$ and $S(\gamma, \tau) = 0$).



(a) The generated data, along with the true and estimated marginal response curves. (b) The S/N ratio.



(c) The tail probability. (d) The tail expectation.

Figure 2.6: Medium internal noise ($\tau = 2.25$ and $S(\gamma, \tau) \approx 0.5$).

2.5 PERFORMANCE LOSS WHEN INTERNAL NOISE IS IGNORED

In order to fully appreciate the gains made when internal noise is accounted for, we can explore a different scenario where the intermediate values X are neither recorded nor observed, and the analysis is based only on the recorded tuples (Z, Y) . This is a fairly common scenario, therefore it is im-

portant to understand the point at which the internal noise levels begin to distort (and degrade) the results obtained from such analyses. A modification of the general hierarchical model (2.6), for inferring the relationship between Z and Y when X is missing, is the following misspecified model

$$\begin{aligned} Y|Z, \vartheta &\sim \mathcal{F}(Y|Z, \vartheta) \\ \vartheta &\sim \pi(\vartheta). \end{aligned} \tag{2.16}$$

We now introduce an approach for assessment of the price that an experimenter has to pay to utilize the misspecified model (2.16) for optimization. Let $\tilde{Z}_{\text{True}}(\tau)$ denote the optimal Z under the true model (2.6) when the level of internal noise as defined earlier is τ . Similarly, let $\tilde{Z}_{\text{Mis}}(\tau)$ denote the optimal Z under the misspecified model (2.16). Denoting some chosen utility function by $\mathcal{J}(\cdot)$, a comparison of the functions $\mathcal{J}(\tilde{Z}_{\text{True}}(\tau))$ and $\mathcal{J}(\tilde{Z}_{\text{Mis}}(\tau))$ as functions of the internal noise level τ provides useful insight into the loss incurred by ignoring internal noise. Clearly, the utility functions should coincide for $\tau = 0$.

Possibilities for the utility function, $\mathcal{J}(Z)$, which are closely tied with our main goal of optimizing the response include the *true tail probability*, $P(Y > y^*|Z, \vartheta, \phi)$, and the *true tail expectation*, $E(I_{\{Y > y^*\}} Y|Z, \vartheta, \phi)$, where y^* is chosen to be $\max(E(Y|X, \vartheta))$. These utility functions, which depend directly on the true parameter values differ from the objective functions, $\eta_2(Z)$ and $\eta_3(Z)$, which depend instead on the observed data \mathcal{D} .

Another method for assessing the loss of performance is to introduce a loss function termed as the *true probability of domination* defined as:

$$\mathcal{K}(\tau) = P[Y(\tilde{z}_{\text{True}}(\tau)) > Y(\tilde{z}_{\text{Mis}}(\tau))], \tag{2.17}$$

where the probability is defined with respect to the true model (2.6).

2.5.1 STUDYING THE EFFECT OF IGNORING HOMOSCEDASTIC INTERNAL NOISE THROUGH SIMULATION

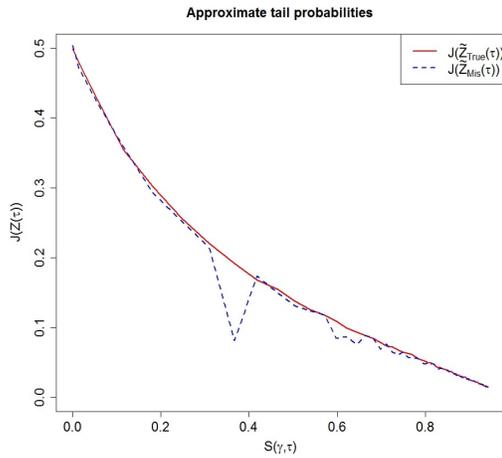
We now explore the loss of performance under the homoscedastic internal noise scenario. For the simulations in this section, we once again use the same simulation set-up as in Sections 2.4.1 and 2.4.2, except that we select τ from 100 equi-spaced points within the interval $[0, 5]$. For each of the 100 variations of the true underlying model, we generate data as in Section 2.4.1. Model estimation and objective function optimization is then carried out using both the misspecified model and the correctly specified model, to obtain the optimal points, $\tilde{z}_{\text{Mis}}(\tau)$ and $\tilde{z}_{\text{True}}(\tau)$ respectively.

For evaluating the loss of performance when using $\tilde{z}_{\text{Mis}}(\tau)$ instead $\tilde{z}_{\text{True}}(\tau)$, we select the utility function, $P(Y > y^* | Z, \vartheta, \phi)$ (the true tail probability), and the loss function, $\kappa\tau$ defined by (2.17). Note that these two functions cannot be written in closed form and have to be numerically approximated using the Monte-Carlo approximation:

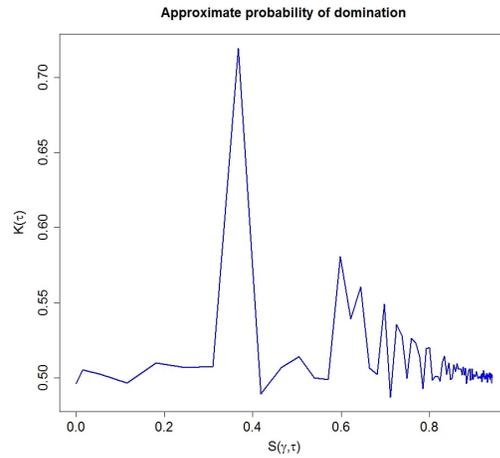
$$P(\mathcal{A}) = \frac{1}{n} \sum_{i=1}^n I_i(\mathcal{A}),$$

where \mathcal{A} denotes some event and $I_i(\mathcal{A})$ is the indicator function for the i th observation belonging to the event \mathcal{A} . The two measures in a single run are shown in Figure 2.7, while results averaged over 500 runs are shown in Figure 2.8. In these plots, the unit-free measure of internal noise $S(\gamma, \tau)$ is plotted on the horizontal axis.

The plots in Figure 2.8 seem to suggest that there is little to be gained on average, from explicitly accounting for homoscedastic internal noise. The plots in Figure 2.7, however, tell a somewhat different, if subtle story. The rough and spiky shapes are not simply as a result of the data sampling process, but reflect abrupt departures in the estimated relationship between Y and Z , from the true concave relationship. i.e., the spikes occur when the truly concave μ_Z is estimated to be non-concave, a situation which happens more often as the noise level increases, but is overtaken at higher



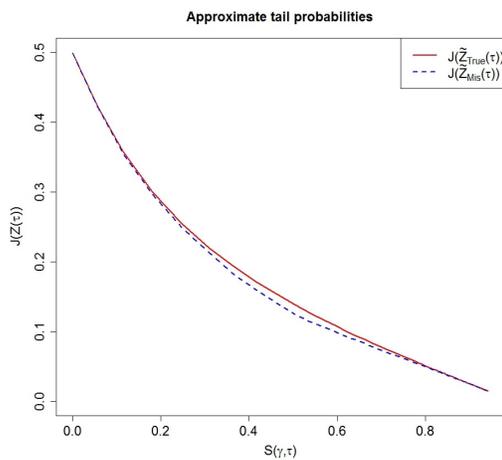
(a) The tail probability.



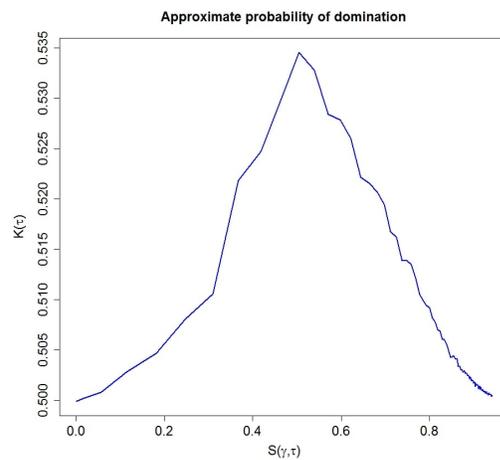
(b) The probability of domination.

Figure 2.7: Results for a single run.

noise levels by the general flattening of the signal. There is thus some benefit in accounting for the internal noise, especially if the shape of the true relationship between the input and response is not known prior to experimentation.



(a) The tail probability.

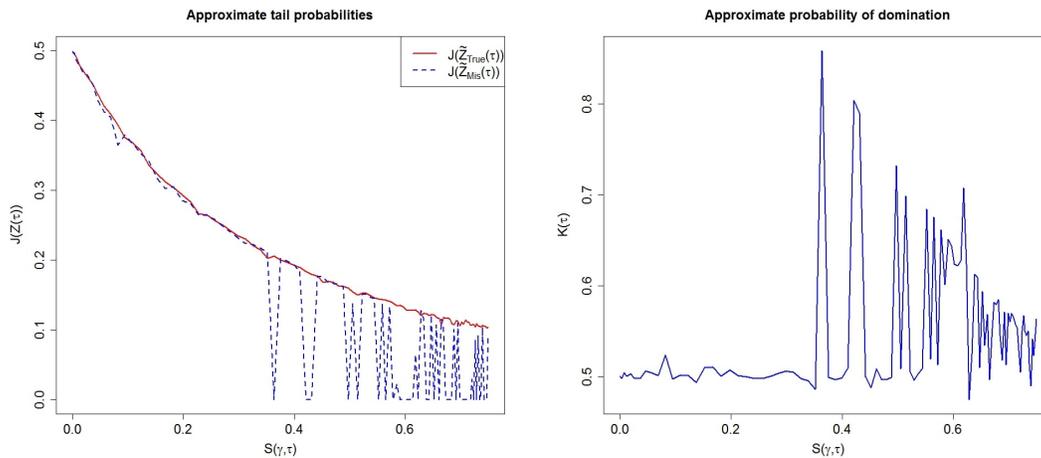


(b) The probability of domination.

Figure 2.8: Results averaged over 500 runs.

2.5.2 STUDYING THE EFFECT OF IGNORING HETEROSCEDASTIC INTERNAL NOISE THROUGH SIMULATION

In the heteroscedastic case, we can also visualize the performance measures by setting up the simulations as in the homoscedastic case, in the last subsection.



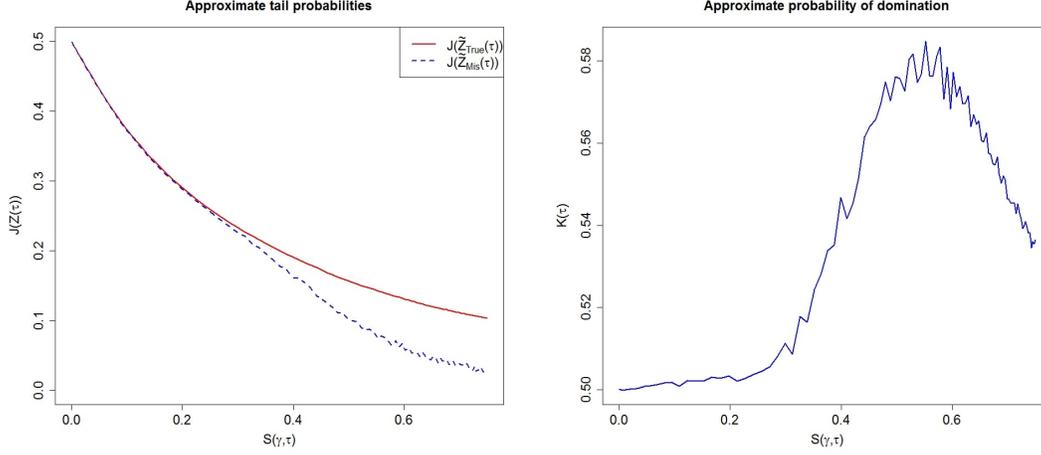
(a) The tail probability.

(b) The probability of domination.

Figure 2.9: Results for a single run.

We only need to change the distribution of X to $\mathcal{N}(X | \gamma Z, \tau^2 Z)$, and make the necessary adjustments to the performance measures. The two measures in a single run are shown in Figure 2.9, and averaged over 500 runs in Figure 2.10.

In Figure 2.9, the noise-dependent departures, in the estimated relationship between Y and Z , from the true concave relationship are more obvious now since the flattening of the signal is less severe in this heteroscedastic case. Figure 2.10 also shows the relatively sub par performance of misspecified model on average, when compared to the correctly specified one, as the noise level increases. Thus there is some clear benefit in accounting for the internal noise, particularly when the shape of the true relationship between the input and response is unknown prior to experimentation.



(a) The tail probability.

(b) The probability of domination.

Figure 2.10: Results averaged over 500 runs.

2.6 APPLICATION TO CNT ALIGNMENT

In order to apply the results of our simulations to the optimization of CNT alignment, we first need to propose a model for the internal noise and then update the objective function $\eta_3(Z)$ from (2.10), and the internal noise measure, $S(\gamma, \tau)$ accordingly. In the experiment described in Section 2.1, the following Bayesian hierarchical model was proposed to model the response R

$$\begin{aligned}
 R|X, I, M, \vartheta &\sim \mathcal{N}(\psi(X, I, M), \sigma^2) \\
 X|Z, \varphi &\sim \Gamma(\alpha Z, \gamma) \equiv \frac{1}{\gamma} \Gamma(\alpha Z) \\
 (\vartheta, \varphi) &\sim \Pi(\vartheta, \varphi) \propto \mathbf{1},
 \end{aligned} \tag{2.18}$$

where Γ denotes the gamma distribution, $\vartheta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \sigma)$, $\varphi = (\alpha, \gamma)$, and

$$\psi(X, I, M) = \beta_0 M^2 + (\beta_1 + \beta_2 MI + \beta_3 M^2 I^2)X + \beta_4 X^2,$$

following the results in Remillard et al. ⁴⁵. The gamma distribution model for the internal noise ensures that the generated electric field is non-negative, and has a linear mean and a mean-dependent variance in agreement with Figure 2.2. i.e.,

$$E(X|Z) = \frac{\alpha}{\gamma}Z \text{ and } Var(X|Z) = \frac{\alpha}{\gamma^2}Z.$$

The noise measure, $S(\gamma, \tau)$, now has the form

$$S(\alpha, \gamma) = 1 - 1/\sqrt{1 + 6\frac{(a+b)}{\alpha(b-a)^2}}.$$

The optimization domain for $\eta_3(Z, I, \mathcal{M})$, denoted by Ω , is restricted to $[0, 2.5] \times [4, 10] \times [0.001, 0.025]$, the experimental data domain used by Remillard et al. ⁴⁵. Preliminary results from Remillard et al. ⁴⁵ give the following parameter maximum likelihood estimates, coinciding in our model with the posterior modes: $\hat{\beta}_0 = 2.816 \times 10^2$, $\hat{\beta}_1 = 5.921 \times 10^{-2}$, $\hat{\beta}_2 = -3.107 \times 10^{-1}$, $\hat{\beta}_3 = 2.845$, $\hat{\beta}_4 = -1.411 \times 10^{-3}$ and $\hat{\sigma} = 0.082$. Based on these estimates, we numerically calculate the threshold to be $\gamma^* = 0.744$.

We obtain $\hat{\phi} = (\hat{\alpha}, \hat{\gamma})$ based on the gamma noise model proposed in (2.18), and the data used to generate Figure 2.2. Maximizing the gamma likelihood, which in our case is proportional to the posterior density, gives the posterior modes $\hat{\alpha} = 1254.19$ and $\hat{\gamma} = 146.12$, using *optim* in *R*⁴³. We also generate posterior samples for (α, γ) using the *MCMCpack* package³⁵ in *R*. We now proceed with optimizing $\eta_3(Z, I, \mathcal{M})$.

Due to the difficulty of visualizing a response over a three-dimensional domain, we visualize $\eta_3(Z, I, \mathcal{M})$ as a function of Z , when (\mathcal{M}, I) are fixed at some points within their joint domain. For example, we select

$$(\mathcal{M}, I) \in \{(0.001, 4), (0.001, 10), (0.025, 4), (0.009, 6)\},$$

where the first three points are corners of the joint domain, and the last point is the center of those three. Figure 2.11 suggest $\eta_3(Z, I, \mathcal{M})$ will be well behaved (unimodal) over the entire domain, Ω , and we can now carry out the optimization.

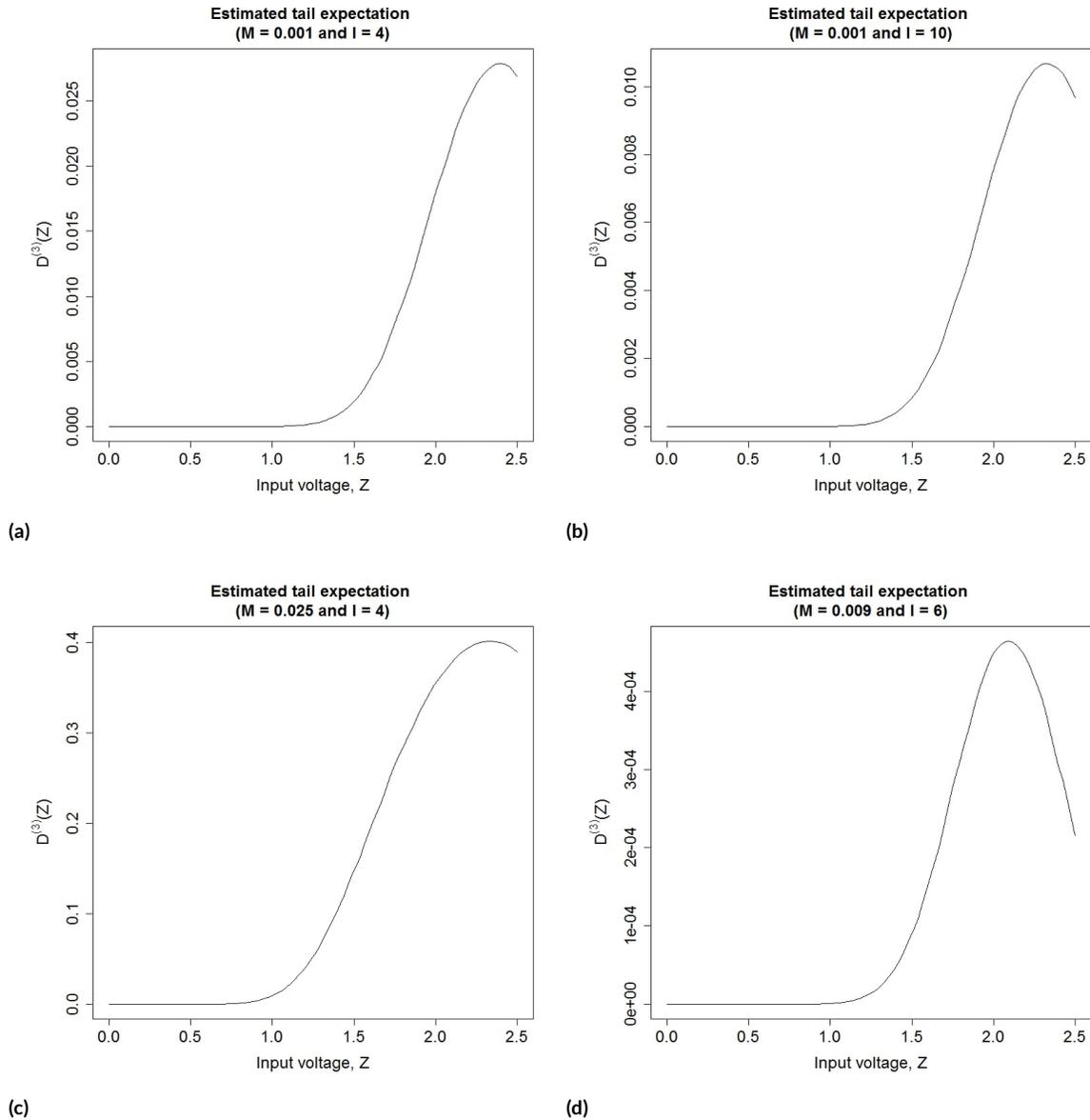


Figure 2.11: Profiles of $\eta_3(Z, I, \mathcal{M})$ as a function of Z , for four different values of (\mathcal{M}, I) .

Using the optimization tool, *optim*, in *R*, we obtain the following optimal design values for maximizing CNT alignment

$$\tilde{z}_{\text{True}} = 2.33, \tilde{i}_{\text{True}} = 4, \text{ and } \tilde{m}_{\text{True}} = 0.025.$$

\tilde{i}_{True} and \tilde{m}_{True} are in perfect agreement with the results obtained by Remillard et al. ⁴⁵. Based on joint posterior samples of (α, γ) , $\tilde{z}_{\text{True}} = 2.33$ yields a posterior predictive electric field strength expectation, $E(X|Z = \tilde{z}_{\text{True}})$, of 20.02, with a 95% prediction interval (19.28, 20.78).

This is essentially the same as $\tilde{x} = 20.05$ obtained by Remillard et al. ⁴⁵ by optimizing $\hat{\mu}_X$ directly. The results match so well in this example both because the authors account for internal noise somewhat by optimizing $\hat{\mu}_X$ directly, and also because the internal noise level is itself very small (a 95% posterior interval for $S(\alpha, \gamma)$, is calculated as (0.0007, 0.0013)).

If we assume instead that the intermediate electric field, X , is not observed/recorded, and that the data generating model is given as in (2.16),

$$R|Z, I, \mathcal{M}, \mathcal{P} \sim \mathcal{N}(\mu(Z, I, \mathcal{M}), \nu^2)$$

$$\mathcal{P} \sim \pi(\mathcal{P}) \propto \mathbf{1},$$

where $\mathcal{P} = (C_0, C_1, C_2, C_3, C_4, \nu)$, and

$$\mu(Z, I, \mathcal{M}) = C_0 \mathcal{M}^2 + (C_1 + C_2 \mathcal{M} I + C_3 \mathcal{M}^2 I^2) Z + C_4 Z^2,$$

then we obtain the following parameter estimates: $\hat{C}_0 = 280.95$, $\hat{C}_1 = 0.5203$, $\hat{C}_2 = -2.5417$, $\hat{C}_3 = 23.5128$, $\hat{C}_4 = -0.1102$ and $\hat{\nu} = 0.0813$.

Optimizing $\hat{\mu}(Z, I, \mathcal{M})$ gives us the following optimal design values for maximizing CNT align-

ment

$$\tilde{z}_{\text{Mis}} = 2.275, \tilde{i}_{\text{Mis}} = 4, \text{ and } \tilde{m}_{\text{Mis}} = 0.025.$$

Once again, \tilde{i}_{Mis} and \tilde{m}_{Mis} are in perfect agreement with the results obtained by Remillard et al. ⁴⁵, as well as with \tilde{i}_{True} and \tilde{m}_{True} . $\tilde{z}_{\text{Mis}} = 2.275$ differs from \tilde{z}_{True} , but only slightly. Based on the assumption of zero internal noise, any \tilde{z}_{Mis} is equivalent to an X value of $(\alpha/\gamma)\tilde{z}_{\text{Mis}}$, which can be estimated either by $(\hat{\alpha}/\hat{\gamma})\tilde{z}_{\text{Mis}}$ or the average over posterior samples of (α, γ) . In our case, both estimates yield 19.53, which falls within the 95% posterior interval for X (given that $Z = \tilde{z}_{\text{True}}$), given earlier as (19.28, 20.78). Thus all three methods (optimizing $\hat{\mu}_Z, \hat{\mu}_X$, or some hybrid of both based on the full hierarchical model) can be said to yield equivalent results when the internal noise level is minimal. If however, α and γ had been several orders of magnitude smaller, while leaving α/γ constant, then the inference (and thus the optimization) resulting from the marginal model would have been unreliable compared to inference based on the full model.

2.7 CONCLUDING REMARKS

We have introduced the general problem of carrying out optimization in the presence of internal noise. Developing a solution for this problem necessitates the creation of appropriate objective functions relevant to our goals, in addition to some performance functions for evaluating those objective functions under internal homoscedasticity or heteroscedasticity. We find that one of the objective functions, the expected tail area, provides the most reliable optima when compared to the others especially the signal to noise ratio. Furthermore, we observe that accounting for possible internal noise improves the optimization process in the homoscedastic case, albeit very slightly. In order to guide future decision making about the measurement and analysis of internal noise, we have proposed a unit-free and response-independent measure of internal noise.

On the other hand, in the heteroscedastic case, the misspecified model which ignores the pres-

ence of internal noise performs poorly as the noise level increases, when compared to the full model which accounts for internal noise. Applying the results of our simulation-based analyses to the CNT data yields expected results: our optimization results coincide almost perfectly with the results of Remillard et al. ⁴⁵, due to the minimal internal noise present in the CNT dataset. We should however expect more substantial gains under relatively high internal noise levels.

Truth is the most valuable thing we have. Let us economize it.

Mark Twain

3

Optimal Experimental Design in the Presence of Internal Noise

IN EXPERIMENTS CONDUCTED to infer relationships between inputs and outputs, it is common for measurements obtained from physical instruments to contain noise. The effect of external noise (e.g., noise due to output measurement error) on statistical inference is a well-studied one, as mentioned in Chapter 2. Most of the work done by statisticians on robust design in the 1980s and 1990s dealt with external noise and were motivated by manufacturing (e.g., automobile) applications. However, as noted by Kang & Joseph²², there has been little interest on the role of internal noise,

typically associated with noisy input measurements, especially as it pertains to experimental designs.

Modern optimal experimental designs, pioneered by Kiefer & Wolfowitz²³, arose as a flexible strategy for reducing the cost of experimentation. The choice of an optimal design is often linked to a model, and is with respect to some chosen statistical criterion, related to objectives such as response surface exploration, prediction, variable screening etc.⁶³. The optimal design points are chosen so as to achieve the highest possible precision in estimating a quantity of interest⁹, and thus provide the best possible statistical inference about that quantity. Because specific information, such as the experimental region, model framework and prior parameter beliefs, is usually available prior to experimentation, Bayesian methods can play an important role in determining optimal designs⁹.

For example, consider the hypothetical response surface model

$$\begin{aligned}
 Y &\sim \mathcal{N}(\beta_0 + \beta_1 X + \beta_2 X^2, \sigma^2) \\
 (\beta_0, \beta_1, \beta_2) &\sim \pi(\cdot),
 \end{aligned}
 \tag{3.1}$$

where we have denoted a single covariate of interest as X , and the prior parameter distribution as $\pi(\cdot)$. The D-optimal design (which corresponds to a response surface exploration objective) for this linear model is the set of input points, X , which maximizes the expected Kullback-Leibler distance between the posterior and prior distributions for $\vartheta = (\beta_0, \beta_1, \beta_2)$, or equivalently maximizes the expected gain in Shannon information⁹. For the quadratic mean response in (3.1), the D-optimal design is an equally weighted 3-point design (at both boundaries and in the center of X 's domain) when σ is known and the prior, $\pi(\cdot)$, is flat. Now suppose the covariate, X , is subject to internal noise, i.e., X is itself a random function of a true input Z and we can write for example,

$$X \sim \mathcal{N}(g(Z), \nu_z),$$

where $g(\cdot)$ is a deterministic and invertible function, and ν_z denotes the internal noise variance, also referred to as the internal noise magnitude. The equally weighted 3-point D-optimal design based on X , generally will not translate to a similar design based on Z , since some distortion will be experienced through $g(\cdot)$, and based on the internal noise magnitude τ . Therefore the traditional optimality criteria have to be modified in order to appropriately account for the presence of internal noise.

In the field of experimental design, the focus on internal noise has mainly been with regard to robust parameter designs^{22,54,63}, i.e., the choice of optimal levels of process inputs that make the response least sensitive to noise. In a departure from robust parameter design goals, we seek to minimize the estimation uncertainty instead of the response sensitivity to noise. Therefore, our main objective in this chapter is to develop practical approaches to model-based optimal experimental design in the presence of internal noise. It should be noted that our work in this chapter builds on the results in Chapter 2, which primarily dealt with parameter estimation, and optimization in the presence of internal noise, when the experiment has already been designed and run. To achieve our main goal, we first discuss the identification of a design optimality criterion in Section 3.1, and computational strategies in Section 3.2. We examine the design optimality criterion under homoscedastic and heteroscedastic internal noise conditions in Section 3.3, apply our results to the design of a real-world experiment in Section 3.4, and conclude in Section 3.5.

3.1 IDENTIFICATION OF DESIGN OPTIMALITY CRITERION

We define a general relationship between a response Y , a single noise input X , and the controllable input Z through following hierarchical model, similar to that proposed in Chapter 2,

$$Y|X, \vartheta \sim \mathcal{F}(Y|X, \vartheta)$$

$$X|Z, \varphi \sim \mathcal{G}(X|Z, \varphi)$$

$$(\vartheta, \varphi) \sim \pi(\vartheta, \varphi),$$

where the distributions \mathcal{F} and \mathcal{G} are parametrized by ϑ and φ respectively, and $\pi(\cdot)$ denotes a prior distribution for the parameters.

To set up the optimal design problem in a simplified way, we assume that the (X, Z) tuples can be generated relatively cheaply resulting in the estimation of φ with high precision. Thus, the focus of the experimenter is on efficient estimation of ϑ . This is a reasonable assumption that holds in many experimental settings. For example, in the experiment reported by Dasgupta et al.¹¹, the parameters of the distributions of the noisy inputs temperature, pressure and distance from the center of the furnace could be estimated fairly precisely from large sets of observations on (X, Z) that were easy to obtain. The main goal is therefore to obtain a set of design points \mathcal{Z} that will maximize the estimation precision of ϑ . Our proposed research is along the lines of the development in the field of Bayesian optimal designs⁹ over the past couple of decades, and recent applications in nanotechnology and manufacturing (see Kang & Joseph²², Zhu et al.⁶⁸).

Under the assumption of zero internal noise, the traditional Bayesian D-optimal design can be defined as the design which maximizes the expected gain in Shannon information, or equivalently maximizes the Kullback-Leibler distance between the posterior and the prior distributions of ϑ . The Bayesian D-optimality criterion is thus given by:

$$\int \left[\int \log \frac{\pi(\vartheta|\mathbf{y}_n, \mathcal{X}_n)}{\pi(\vartheta)} \pi(\vartheta|\mathbf{y}_n, \mathcal{X}_n) d\vartheta \right] \pi(\mathbf{y}_n|\mathcal{X}_n) d\mathbf{y}_n, \quad (3.2)$$

where $\mathcal{X}_n = \{X_1, \dots, X_n\}$ denotes a set of n design points and \mathbf{y}_n denotes the corresponding vector of observed response. Maximizing (3.2) is equivalent to maximizing

$$\int \left[\int \log \pi(\vartheta|\mathbf{y}_n, \mathcal{X}_n) \pi(\vartheta|\mathbf{y}_n, \mathcal{X}_n) d\vartheta \right] \pi(\mathbf{y}_n|\mathcal{X}_n) d\mathbf{y}_n. \quad (3.3)$$

The above idea works particularly well for linear models with normal residuals, where nice closed-form expressions of the D-optimality criterion can be obtained. In the absence of internal noise, if we consider the following normal hierarchical model:

$$\begin{aligned} \mathbf{Y}|\beta, \sigma^2 &\sim \mathcal{N}(U\beta, \sigma^2 \mathbf{I}) \\ \beta|\sigma^2 &\sim \mathcal{N}(\tilde{\beta}, \sigma^2 \Sigma^{-1}), \end{aligned}$$

where U is an $n \times k$ design matrix, with rows \mathbf{u}_i (e.g. if $E(Y|X, \beta) = \beta_0 + \beta_1 X + \beta_2 X^2$, then $k = 3$, and $\mathbf{u}_i = (1, x_i, x_i^2)^T$), \mathbf{I} denotes the $n \times n$ identity matrix, and σ^2 is assumed known. Then, the posterior distribution of β given $(\mathbf{y}_n, \mathcal{X}_n)$ is:

$$\beta|\mathbf{y}_n, \mathcal{X}_n \sim \mathcal{N}(\beta^*, V^*),$$

where β^* denotes the posterior mode and V^* denotes the inverse of the expected Fisher information $\mathcal{I}(\mathcal{X}_n)$, and both are given by

$$\beta^* = U^T U (U^T U + \Sigma)^{-1} ((U^T U)^{-1} U^T \mathbf{y}_n) + \Sigma (U^T U + \Sigma)^{-1} \tilde{\beta} \quad \text{and,} \quad V^* = \sigma^2 (U^T U + \Sigma)^{-1}.$$

The Bayesian D-optimality criterion in (3.3) reduces to

$$q(\mathcal{X}_n) = \det\{\mathcal{I}(\mathcal{X}_n)\}. \tag{3.4}$$

Extension of the above D-optimality criterion to account for internal noise is clearly a non-trivial problem that can be addressed in a variety of ways. One naive method is to take the expectation of the criterion over the distribution of X , which, quickly raises serious computational issues. Another approximate method that we plan to explore is to work with posterior distribution of \mathfrak{D} after

marginalizing out X .

In the presence of internal noise, let $\mathcal{Z}_n = \{Z_1, \dots, Z_n\}$ denote a set of n design points in terms of the controllable input Z . Denoting the posterior mode β after marginalization of X as $\hat{\beta}_{PM}(\mathcal{Z}_n)$, and the expected Fisher information after marginalization as $\mathcal{I}(\mathcal{Z}_n)$, we can approximate the marginalized posterior distribution of β by

$$\beta | \mathbf{y}_n, \mathcal{Z}_n \sim \mathcal{N} \left(\hat{\beta}_{PM}(\mathcal{Z}_n), \mathcal{I}^{-1}(\mathcal{Z}_n) \right),$$

where “the marginalized information” has the following closed-form expression

$$\mathcal{I}(\mathcal{Z}_n) = \sigma^{-2} [E(U^T U | \mathcal{Z}_n) + \Sigma]. \quad (3.5)$$

Consequently, it can be shown that maximizing the determinant of $\mathcal{I}(\mathcal{Z}_n)$ will yield the D-optimal design, under the assumption that σ^2 is known. In other words, the D-optimal criterion in the presence of internal noise becomes

$$\tilde{q}(\mathcal{Z}_n) = \det\{\mathcal{I}(\mathcal{Z}_n)\}. \quad (3.6)$$

Note that under zero internal noise, this reduces to the traditional Bayesian D-optimality criterion, $q(\mathcal{X}_n)$ in (3.4).

3.2 COMPUTATIONAL STRATEGIES FOR FINDING OPTIMAL DESIGNS UNDER THE LINEAR MODEL

Finding continuous optimal designs under the linear model in the presence of no internal noise is a well researched problem⁴², and it is fairly straightforward to extend such algorithms to the case with internal noise when the conditional model of Y given X is linear with normal residuals. The main idea is to use Carathéodory’s theorem (stated later) to obtain an upper bound for the number of

unique Z_i 's we can expect to find in any design¹⁴, and to use suitable optimization tools to find the continuous D-optimal design.

The search for the optimal design, \mathcal{X}_n^* , proceeds by first noting that we can write the Fisher information, $\mathcal{I}(\mathcal{X}_n)$, as

$$\mathcal{I}(\mathcal{X}_n) = \frac{1}{\sigma^2} (U^T U + \Sigma) = \frac{1}{\sigma^2} \left(\sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i^T + \Sigma \right) = \frac{1}{\sigma^2} \left(\sum_{j=1}^r n_j (\mathbf{u}_j \mathbf{u}_j^T) + \Sigma \right),$$

where $r \leq n$ denotes the number of unique x_i 's, and $\sum_{j=1}^r n_j = n$. Thus a search for the optimal design can be seen as a search for both the unique design points x_j 's and their corresponding proportions $w_j = n_j/n$ which optimize the chosen criteria. Designs where each $n \cdot w_j$ is required to be an integer, are known as exact designs and finding the optimal designs is generally a hard integer programming problem⁹. On the other hand, designs where such requirements are relaxed are known as approximate or continuous designs. Our focus will be on such continuous designs for the remainder of this paper.

In order to carry out the search, Carathéodory's theorem provides an upper bound for the number of unique x_j 's we can expect to find in any design. The theorem states that for any m -dimensional point, x , in the convex hull of a set B , there exists a subset of B , say \mathcal{A} , consisting of only $m + 1$ points such that x also lies in the convex hull of \mathcal{A} . The implication of the theorem is that the quantity $(1/n) \cdot U^T U$ (henceforth denoted by $\mathcal{M}(\mathcal{X}_n)$), which is a convex combination of n points, only requires some $m + 1$ number of unique design points to be fully described, where $m = \dim(\mathcal{M}(\mathcal{X}_n))$. Note that for k -dimensional \mathbf{u}_i vectors, $\mathcal{M}(\mathcal{X}_n)$ lies in $k(k + 1)/2$ dimensional space, since that is the number of terms needed to fully describe the symmetric matrix, $\mathcal{M}(\mathcal{X}_n)$.

The search for the optimal design based on the chosen optimality criteria then proceeds by initializing $(x_1, \dots, x_m, x_{m+1})$ and their corresponding probability measures $(w_1, \dots, w_m, w_{m+1})$, subject to $\sum_i w_i = 1$, and then employing an optimization tool, say *optim* in R , to find the continuous D-

optimal design. This continuous design can be subsequently rounded off to an exact design, by employing an efficient rounding algorithm, such as that proposed by Pukelsheim ⁴², and implemented in the *R* package, AlgDesign ⁶².

3.3 SIMULATIONS TO ASSESS LOSS ACCRUED BY IGNORING INTERNAL NOISE

We now set up simulations to visualize how the modified D-optimality criterion, $\tilde{q}(\mathcal{Z}_n)$ in (3.6), behaves and performs as the magnitude of internal noise changes. We assume the true data generating model is given by

$$\begin{aligned} Y|X &\sim \mathcal{N}(\beta_0 + \beta_1 X + \beta_2 X^2, \sigma^2) \\ X|Z &\sim \mathcal{N}(\gamma Z, \nu_z) \\ (\beta_0, \beta_1, \beta_2, \sigma, \gamma, \nu_z) &\sim \pi(\cdot), \end{aligned} \tag{3.7}$$

where $\pi(\cdot)$ denotes the joint prior distribution for the parameters. Considering two scenarios for the data generating mechanism: (i) homoscedastic internal noise ($\nu_z = \tau^2$), and (ii) a special case of heteroscedastic internal noise, using a mean dependent variance ($\nu_z = \tau^2 Z$), we now identify some measures for assessing performance and behavior of the modified D-optimality criterion.

3.3.1 PERFORMANCE AND CHARACTERISTIC MEASURES

The primary objective of these simulations is to assess the precision gain (in the response surface estimation) that occurs when a D-optimal design is generated using the modified D-optimality criterion, $\tilde{q}(\mathcal{Z}_n)$, and the (correctly) specified model in (3.7), instead of the following commonly used misspecified data generating model

$$Y|Z \sim \mathcal{N}(C_0 + C_1 Z + C_2 Z^2, \kappa^2)$$

$$(C_o, C_1, C_2, \kappa) \sim \pi(\cdot), \quad (3.8)$$

where $\pi(\cdot)$ denotes the joint prior distribution for the parameters. Note that the D-optimal design under the misspecified model in (3.8) is a 3-point balanced design at a , $(a + b)/2$ and b , if the interval $[a, b]$ denotes the experimental domain of Z , and the joint prior for C_o , C_1 and C_2 is flat. Evaluation of the estimation precision then proceeds under three degrees of misspecification:

1. Correctly specified models for both design and inference stages, using (3.7). We term this the *Correctly Specified* model.
2. A misspecified model for the design stage using (3.8), and a correctly specified model for inference using (3.7). In an experimental setting, this may occur if the presence of internal noise is only realized after the experiments have been carried out. We call this combination the *Partially Misspecified* model.
3. Misspecified models for both design and inference stages, using (3.8). This is the *Fully Misspecified* model.

Comparison of the estimation uncertainty resulting from the *Correctly Specified*, and *Partially Misspecified* models enables us to assess the precision gain resulting from D-optimal designs when internal noise is correctly accounted for. On the other hand, the resulting precision loss when the inference model is also misspecified is evaluated by comparing the estimation uncertainty resulting from the *Partially Misspecified* and *Fully Misspecified* models. For these comparisons, we propose the following measures of estimation uncertainty

1. Model RMSE with respect to X (Conditional RMSE):

$$\Psi_x = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(E(Y_i | x_i) - \hat{Y}(x_i) \right)^2},$$

where $\hat{Y}(x)$ denotes the estimated mean response at x , and $\{x_1, \dots, x_n\}$ denotes a set of domain points chosen on the conditional response surface.

2. Model RMSE with respect to Z (Marginal RMSE):

$$\Psi_z = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(E(Y_i | z_i) - \hat{Y}(z_i) \right)^2},$$

where $\hat{Y}(z)$ denotes the estimated mean response at z , and $\{z_1, \dots, z_n\}$ denotes a set of domain points chosen on the marginal response surface.

Note that both the conditional and marginal RMSEs can be used for comparing the *Correctly Specified*, and *Partially Misspecified* models. On the other hand, only the marginal RMSE can be used for comparing the *Fully Misspecified* model with the other models, since the *Fully Misspecified* model has no intermediate ‘conditional’ surface.

For the secondary goal of visualizing the behavior of $\tilde{q}(\mathcal{Z}_n)$, the modified D-optimality criterion, we propose the following measures:

1. The average design point location:

$$\Psi_a = \frac{1}{n} \sum_{i=1}^n z_i = \sum_{j=1}^r w_j \cdot \tilde{z}_j,$$

for r distinct values, $\tilde{z}_1, \dots, \tilde{z}_r$, of z_i 's, where w_j denotes the fraction of z_i 's which are equal to \tilde{z}_j .

2. The design balance:

$$\Psi_b = \sqrt{\prod_{j=1}^r (1 - w_j)},$$

for r distinct values of z_i 's where w_j once again denotes the fraction of z_i 's which are equal to \tilde{z}_j . For example, $\Psi_b = (n - 1)/n$ (indicating high-balance) when every z_i is distinct, and $\Psi_b = 0$ (lowest possible balance) when all the design z_i 's are the same.

Note that the two measures, Ψ_a and Ψ_b , are purely descriptive and no specific value for either measure is preferred. We now examine first, the effects of accounting for internal noise in the design and inference stages when the internal noise structure is homoscedastic.

3.3.2 SIMULATION I: HOMOSCEDASTIC INTERNAL NOISE

For a homoscedastic internal noise model based on (3.7), the internal variance, v_z is equivalent to τ^2 , some constant. We set the following true parameter values

$$\beta_0 = 200, \beta_1 = 1, \beta_2 = -1, \sigma = 0.05, \text{ and } \gamma = 10,$$

and vary τ over the interval $[0, 1]$. The domain of Z , denoted by Ω_Z , is set to the interval $[0, 0.1]$.

The first four moments of X conditional on Z are given by

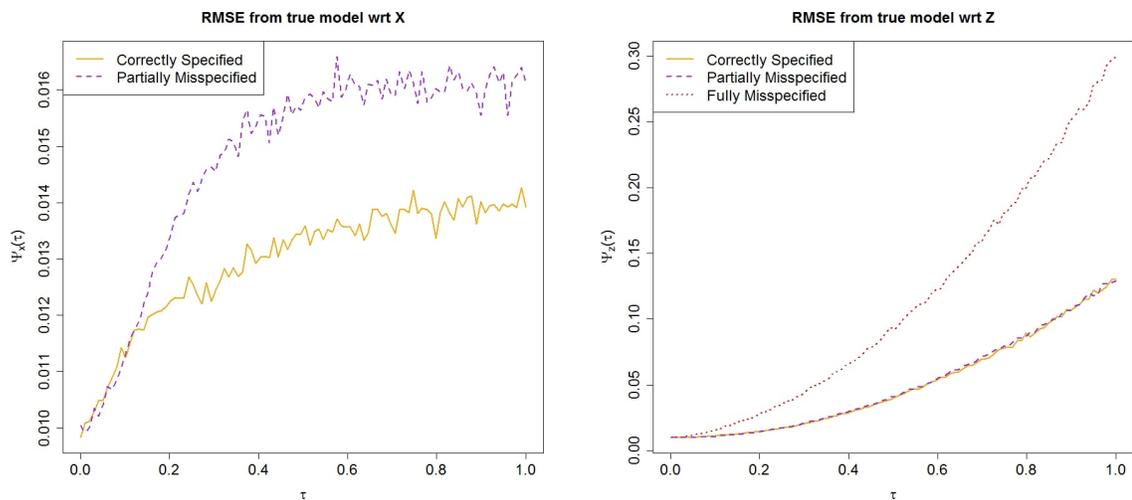
$$E(X|Z) = \gamma Z, \quad E(X^2|Z) = \gamma^2 Z^2 + \tau^2,$$

$$E(X^3|Z) = \gamma^3 Z^3 + 3\gamma\tau^2 Z, \quad E(X^4|Z) = \gamma^4 Z^4 + 6\gamma^2\tau^2 Z^2 + 3\tau^4.$$

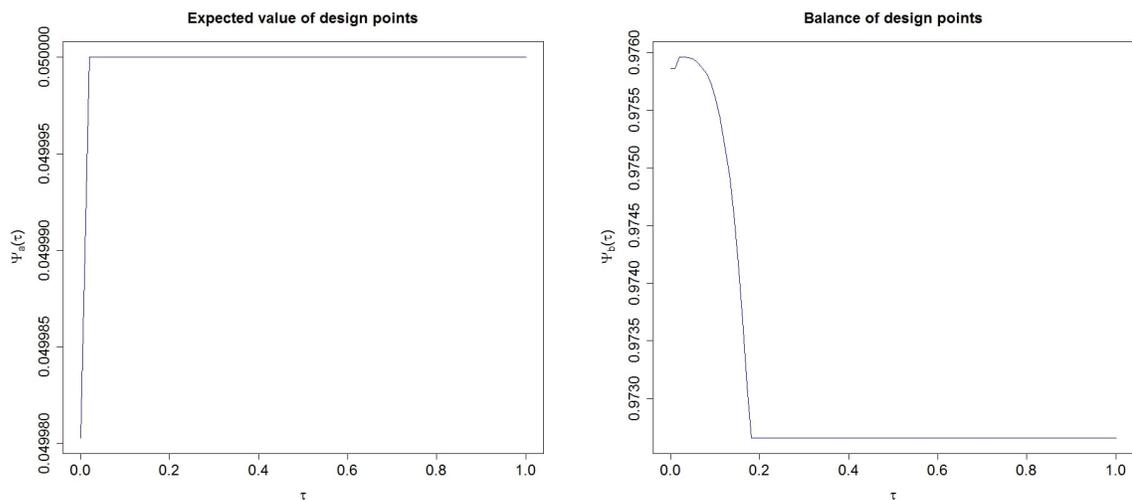
Substituting these into (3.5), along with a joint flat prior for β_0 , β_1 , and β_2 , allows us to calculate the modified D-optimal criterion function in (3.6) for every design set, $\mathcal{Z}_n = \{z_1, \dots, z_n\}$, and thus calculate the noise-modified D-optimal design at each value of τ . Note that Σ is a matrix of zeros by virtue of the joint flat prior for β_0 , β_1 , and β_2 .

For assessing estimation uncertainty, we generate 50 triplets of (Z, X, Y) , based on each set of estimated design points. We also generate a further 50 tuples of (Z, X) , representing supplementary

data, by using equi-spaced z 's within Ω_Z , to generate corresponding x 's. Parameter estimation and response prediction proceeds in similar fashion as in chapter 2. Figure 3.1 shows the performance and characteristic measures defined in Section 3.3.1.



(a) Conditional RMSE for *Correctly Specified*, and *Partially Misspecified* models (b) Marginal RMSE for *Correctly Specified*, *Partially Misspecified*, and *Fully Misspecified* models



(c) Average design point location (d) Design balance

Figure 3.1: Homoscedastic internal noise measures averaged over 2000 runs.

Figure 3.1a shows less estimation uncertainty, in general, when the *Correctly Specified* model is used instead of the *Partially Misspecified* model, especially as the internal noise magnitude increases. This signifies dominance of the noise-modified D-optimality criterion, $\tilde{q}(\cdot)$ in (3.6), over the traditional D-optimal criterion, $q(\cdot)$ in (3.4). Figure 3.1b shows that the *Fully Misspecified* model performs badly when compared to the other two models, and seems to suggest that a misspecified inference model carries a bigger penalty than a misspecified design model when the internal noise is homoscedastic. Figures 3.1c and 3.1d show the evolution of the noise-modified D-optimal design from the traditional 3-point balanced design when the internal noise magnitude is zero, to a 2-point balanced design, resulting from a flattening of the signal (mean function) when the internal noise becomes high.

3.3.3 SIMULATION 2: HETEROSCEDASTIC INTERNAL NOISE

For the heteroscedastic case, we repeat the same data-generation and estimation procedure used for the homoscedastic case (in Section 3.3.2), except for the following changes

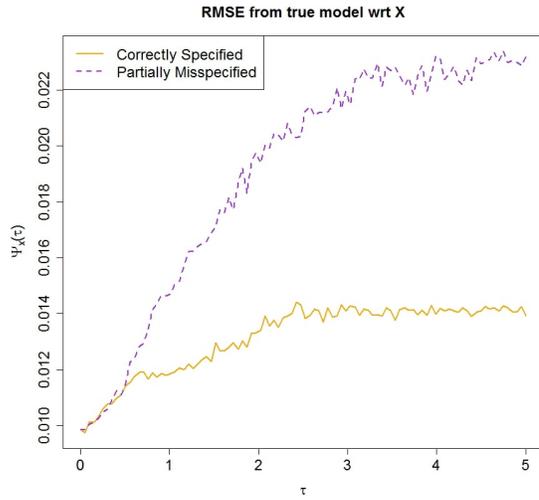
1. The internal variance, ν_z is now equivalent to $\tau^2 Z$.
2. We vary τ over the interval $[0, 5]$

The first four moments of X conditional on Z , which are used to obtain the noise-modified D-optimal design, are now given by

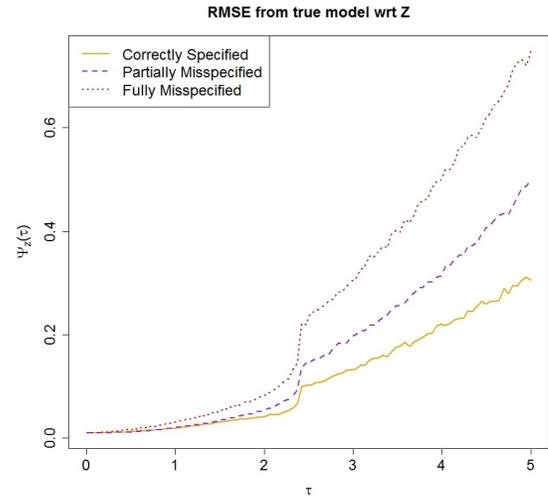
$$E(X|Z) = \gamma Z, \quad E(X^2|Z) = \gamma^2 Z^2 + \tau^2 Z,$$

$$E(X^3|Z) = \gamma^3 Z^3 + 3\gamma\tau^2 Z^2, \quad E(X^4|Z) = \gamma^4 Z^4 + 6\gamma^2\tau^2 Z^3 + 3\tau^4 Z^2.$$

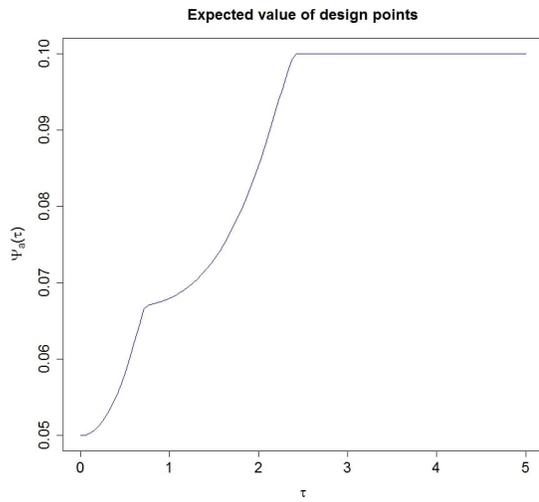
Figure 3.2 now shows the performance and characteristic measures defined in Section 3.3.1, for this heteroscedastic internal noise scenario.



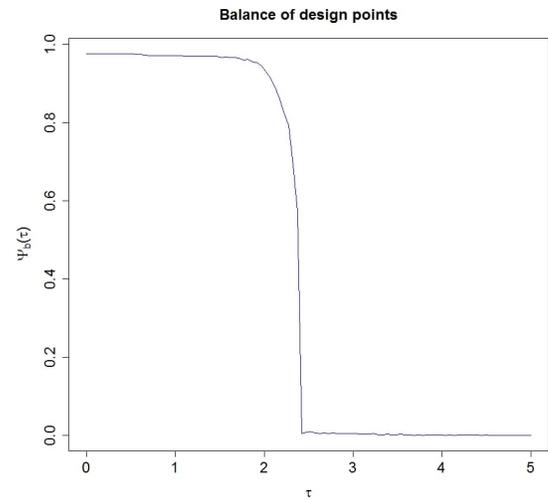
(a) Conditional RMSE for *Correctly Specified*, and *Partially Misspecified* models.



(b) Marginal RMSE for *Correctly Specified*, *Partially Misspecified*, and *Fully Misspecified* models.



(c) Average design point location.



(d) Design balance.

Figure 3.2: Heteroscedastic internal noise measures averaged over 2000 runs.

As in the homoscedastic case, Figure 3.2a shows a gain in estimation precision, when the *Correctly Specified* model is used instead of the *Partially Misspecified* model, once again signifying dominance of the noise-modified D-optimality criterion, $\tilde{q}(\cdot)$ in (3.6), over the traditional D-optimal criterion, $q(\cdot)$ in (3.4). Similar to the homoscedastic case, Figure 3.2b shows that the *Fully Misspecified* model performs badly when compared to the other two models, although it now seems that a misspecified design model carries almost the same penalty as a misspecified inference model when the internal noise is heteroscedastic. Thus it is even more important, when the internal noise is heteroscedastic, to appropriately account for internal noise where possible.

Figures 3.2c and 3.2d show the evolution of the noise-modified D-optimal design from the traditional 3-point balanced design when the internal noise magnitude is zero, to an unbalanced 2-point design at the boundaries of Ω_Z , and finally a 1-point design, when the internal noise becomes high. The unbalanced 2-point design assigns more weight to the boundary point where internal noise is expected to be higher, while the single design point location corresponds to the input location generating the most internal noise. Thus the noise-modified D-optimal design correctly assigns more weight to regions of the input where internal noise is expected to be high, thus ensuring that the response surface is well explored.

3.4 APPLICATION TO CNT ALIGNMENT

Carbon nanotubes (CNTs) play a primary role in flexible electronics⁴⁰, energy storage devices¹², optical displays^{27,32}, chemical sensors¹⁵, and water treatment technologies^{10,58}, due to some of their unique properties such as high electrical and thermal conductivity, high mechanical strength, and optical anisotropy. Making use of CNTs, however, requires their orientation in a single direction.

Motivated by a search for industrial scale alignment techniques for orienting CNTs, Remillard et al.⁴⁵ conducted an experiment to determine the optimal level of electric field strength which maximizes the alignment of carbon nanotubes (CNTs). Figure 3.3 shows an SEM image of multi-walled

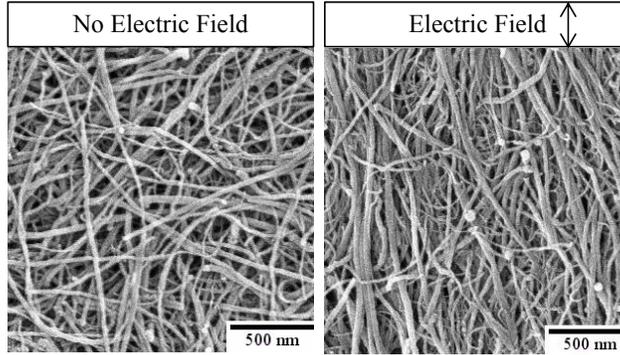


Figure 3.3: Examples of randomly oriented CNT and CNT aligned in the presence on an electric field.

carbon nanotubes aligned using an electric field compared to a randomly oriented sample⁴⁵.

The experiment was conducted by simultaneously varying two other factors in addition to electrical field strength (X): CNT mass (\mathcal{M}), and volume of suspension fluid (measured by its inverse I). Initial exploration of the factor space led to determination of a response model of the form

$$R = \psi(X, \mathcal{M}, I) + \varepsilon, \quad (3.9)$$

where R is a response that measures alignment, $\psi(\cdot)$ is a function of the input variables, and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ denotes the observational noise. In this experiment, the electrical field was generated amplifying a voltage input, Z , to some desired target strength, X , and applying it to the CNT samples suspended in fluid. This electric field amplification process is prone to random fluctuations around the desired value (due to the machine), and thus the amplified electric field, as an input to the alignment process, is inherently noisy. Figure 3.4 shows the relationship between original input voltages (Z) and their corresponding electric field strength (X) from data generated during the experiment, as well as from supplementary (pre-experiment and post-experiment) data. From the figure, it appears that X possess a mean-dependent variance, leading to a heteroscedastic and asymmetric marginal distribution for R .

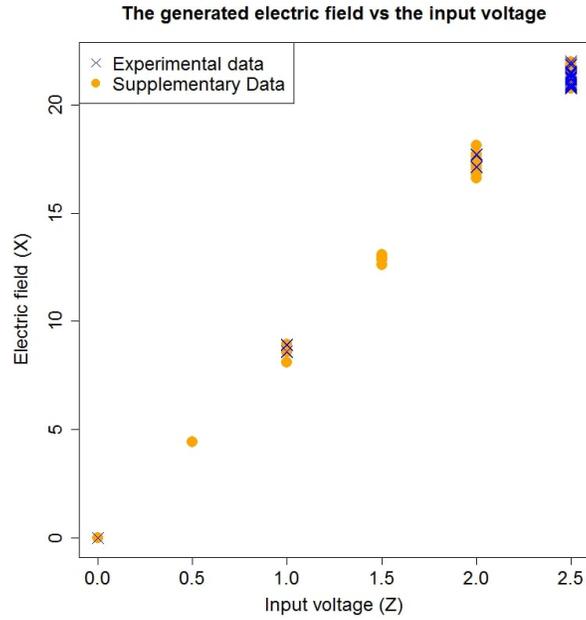


Figure 3.4: The original (generated) voltage (in volts) versus the amplified voltage (in volts).

Applying the insights we have gained thus far in Sections 3.1 – 3.3, to the generation of a D-optimal design for a future CNT alignment experiment, we propose the following Bayesian hierarchical model for the alignment response R ,

$$\begin{aligned}
 R|X, I, \mathcal{M}, \beta, \sigma^2 &\sim \mathcal{N}(\psi(X, I, \mathcal{M}), \sigma^2) \\
 X|Z, \varphi &\sim \Gamma(\alpha Z, \gamma) \equiv \frac{1}{\gamma} \Gamma(\alpha Z) \\
 \beta|\sigma^2 &\sim \mathcal{N}(\tilde{\beta}, \sigma^2 \Sigma^{-1}) \\
 (\sigma^2, \varphi) &\sim \Pi(\sigma^2, \varphi) \propto \mathbf{I},
 \end{aligned} \tag{3.10}$$

where Γ denotes the gamma distribution, $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$, $\varphi = (\alpha, \gamma)$, $\sigma^{-2}\Sigma$ denotes the

prior precision matrix, and

$$\psi(X, I, \mathcal{M}) = \beta_0 \mathcal{M} + (\beta_1 + \beta_2 MI + \beta_3 \mathcal{M}^2 I^2)X + \beta_4 X^2,$$

following the results in Remillard et al. ⁴⁵.

Thus $U^T U$ from (3.5) can be written as $\sum_i \mathbf{u}_i \mathbf{u}_i^T$, where $\mathbf{u}_i^T = (\mathcal{M}_i^2, X_i, \mathcal{M}_i I_i X_i, \mathcal{M}_i^2 I_i^2 X_i, X_i^2)$,

so that,

$$\mathbf{u}_i \mathbf{u}_i^T = \begin{pmatrix} \mathcal{M}_i^4 & \mathcal{M}_i^2 X_i & \mathcal{M}_i^2 I_i X_i & \mathcal{M}_i^4 I_i^2 X_i & \mathcal{M}_i^2 X_i^2 \\ \mathcal{M}_i^2 X_i & X_i^2 & \mathcal{M}_i I_i X_i^2 & \mathcal{M}_i^2 I_i^2 X_i^2 & X_i^3 \\ \mathcal{M}_i^2 I_i X_i & \mathcal{M}_i I_i X_i^2 & \mathcal{M}_i^2 I_i^2 X_i^2 & \mathcal{M}_i^2 I_i^3 X_i^2 & \mathcal{M}_i I_i X_i^3 \\ \mathcal{M}_i^4 I_i^2 X_i & \mathcal{M}_i^2 I_i^2 X_i^2 & \mathcal{M}_i^2 I_i^3 X_i^2 & \mathcal{M}_i^4 I_i^4 X_i^2 & \mathcal{M}_i^2 I_i X_i^3 \\ \mathcal{M}_i^2 X_i^2 & X_i^3 & \mathcal{M}_i I_i X_i^3 & \mathcal{M}_i^2 I_i^2 X_i^3 & X_i^4 \end{pmatrix}$$

The first four moments of X conditional on Z are given by

$$\begin{aligned} E(X|Z) &= \frac{\alpha}{\gamma} Z, \\ E(X^2|Z) &= \frac{\alpha}{\gamma^2} Z + \left(\frac{\alpha}{\gamma} Z\right)^2, \\ E(X^3|Z) &= \frac{2\alpha Z \sqrt{Z}}{\gamma^3} + 3 \frac{\alpha^2}{\gamma^3} Z^2 + \left(\frac{\alpha}{\gamma} Z\right)^3, \\ E(X^4|Z) &= \frac{6\alpha}{\gamma^4} Z^2 + 3 \frac{\alpha^2}{\gamma^4} Z^2 + 4E(X|Z)E(X^3|Z) - 6(E(X^2|Z))^2 + 3(E(X|Z))^4. \end{aligned}$$

Note that each $\mathbf{u}_i \mathbf{u}_i^T$ lives in a 15 dimensional space, and thus, with the application of Carathéodory's theorem, we can expect at most 16 unique combinations of $(\mathcal{M}_i, I_i, Z_i)$ in the optimal design.

If the goal is to design a completely new experiment, we set $\Sigma = \mathbf{o}_{5 \times 5}$ (the zero matrix), which is equivalent to placing a flat prior on β in (3.10). On the other hand, if the main design goal is to improve inference based on a previous experiment, then we can choose $\sigma^2 \Sigma^{-1}$ to be the posterior

covariance matrix for β from that experiment.

Substituting α and γ with their posterior modes, $\hat{\alpha}$ and $\hat{\gamma}$, and optimizing the noise-modified D-optimal criterion, $\tilde{q}(\mathcal{Z}_n)$, over the domain of (M, I, Z) , determined as $\Omega = [0.001, 0.025] \times [4, 10] \times [0, 2.5]$ by Remillard et al. ⁴⁵, we get two D-optimal designs in (3.11), corresponding to each experiment goal. The 7-point design, $\mathcal{Z}_n^{(1)}$, corresponds to the D-optimal design for a completely new experiment, while the single-point design, $\mathcal{Z}_n^{(2)}$, corresponds to the D-optimal design for a continuing experiment using the posterior covariance matrix for β from the CNT alignment results obtained by ⁴⁵. Note that the single-point design, $\mathcal{Z}_n^{(2)}$, is at the location associated with the most internal noise. Each row of the designs, $\mathcal{Z}_n^{(1)}$ and $\mathcal{Z}_n^{(2)}$, contains a unique design point, along with a corresponding proportion, w . As mentioned earlier, these proportions can be converted to an exact design, based on a pre-specified number of experimental runs, by employing an efficient rounding algorithm such as that proposed by Pukelsheim ⁴².

$$\mathcal{Z}_n^{(1)} = \left(\begin{array}{ccc|c} M & I & Z & w \\ \hline 0.0010 & 4.0000 & 2.5000 & 0.1825 \\ 0.0010 & 5.8080 & 1.4029 & 0.0600 \\ 0.0116 & 10.0000 & 1.5426 & 0.1128 \\ 0.0250 & 4.8874 & 2.5000 & 0.1755 \\ 0.0250 & 7.8786 & 0.0000 & 0.1583 \\ 0.0250 & 10.0000 & 1.1717 & 0.1266 \\ 0.0250 & 10.0000 & 2.5000 & 0.1843 \end{array} \right), \quad \mathcal{Z}_n^{(2)} = \left(\begin{array}{ccc|c} M & I & Z & w \\ \hline 0.0250 & 10.0000 & 2.5000 & 1 \end{array} \right) \quad (3.11)$$

3.5 CONCLUDING REMARKS

In this section, we have presented a framework for modifying the traditional optimality criterion, so as to adjust for possible internal noise within the planned experiment. In particular, we have focused solely on D-optimality criterion since those are in line with our main objective of inference. However performing a similar modification of other criteria, say A-optimality, should be straightforward in general. The modified D-optimality criterion behaves as expected: When internal noise levels are very low (or non-existent), the resulting optimal designs match the traditional D-optimal designs- a balanced three point design at the end-points and mid-point of the domain. On the other hand, with high levels of homoscedastic internal noise, the modified criterion leads to a balanced two-point design at the domain end-points, while high levels of heteroscedastic internal noise yields a one-point optimal design at the most noisy location.

As an application example, we use the modified criterion to design future experiments for the alignment of carbon nanotubes, under two scenarios: a completely new experiment, or an experiment to supplement inference based on past experiments. In the completely new experiment scenario, we obtain a seven point design, with a good balance in their proportions (a calculated balance of 0.93, out of all possible balance values within $[0, 0.98]$). When considering a new CNT alignment experiment as a supplement to the previous one carried out by Remillard et al.⁴⁵, then the D-optimal design is a single-point design at the location generating the most internal noise.

*Uncertainty is a quality to be cherished, therefore – if
not for it, who would dare to undertake anything?*

Villiers de L'Isle-Adam

4

A Missing Data Perspective for Modeling Deformations in 3D Printed Products

IN THE PAST DECADE, additive manufacturing (AM) has attracted a lot of attention from manufacturers for producing physical objects⁷. AM, also referred to as 3D printing, is a highly promising manufacturing technique which allows the layer-by-layer synthesis of physical products from 3D Computer-Aided Designs (CAD)¹⁸. Reliance on CAD templates means that production efficiency is not affected by the complexity of the desired product, since molding construction and fixture tooling designs are unnecessary^{19,18}. However, dimensional accuracy control currently poses a challenge

in AM, often due to shrinkage resulting from phase changes in the printing material^{61,21}.

Both empirical and physical models have been proposed for understanding and predicting shrinkage^{6,33,55,67}, but as Huang et al.²¹ note, these models have been limited by inadequate physical understanding, high computational complexity, or fail to capture shape-dependent local shrinkage.

Defining deformation as in-plane profile deviations observed in Polar coordinates, Huang et al.²¹ developed a comprehensive deformation model for 3D printed cylinders which was used to design a compensation plan for minimizing deformation in new unobserved cylinders. i.e., denoting the radii of the nominal (desired), and observed circles at an observed angle, \mathcal{G}^{obs} , by r_o and $r^{obs}(\mathcal{G}^{obs})$ respectively, Huang et al.²¹ developed a model for the deformation, $\Delta^{(c)}$:

$$\Delta^{(c)}(\mathcal{G}^{obs}, r_o) = r^{obs}(\mathcal{G}^{obs}) - r_o . \quad (4.1)$$

Building on that work, Huang et al.²⁰ developed a cookie-cutter approach for modeling regular polygons. i.e., denoting deformation of polygons by $\Delta^{(p)}$, Huang et al.²⁰ proposed:

$$\Delta^{(p)}(\mathcal{G}^{obs}, r_o) = \Delta^{(c)}(\mathcal{G}^{obs}, r_o) + f_I(\mathcal{G}^{obs}) , \quad (4.2)$$

where r_o now denotes the radius of the polygon's circumcircle, and $f_I(\cdot)$ represents the deformation due to the deviation of the polygon from the circumcircle (cookie-cutter difference).

The choice of the Polar coordinates is a natural one since the Cartesian representation, which has been studied^{57,56}, faces a practical issue of correctly identifying shape deviation and capturing dependence between the axes²¹.

The work by Huang et al.²¹ and Huang et al.²⁰, form the basis and inspiration for this paper. Although the cookie-cutter models developed by Huang et al.²⁰, perform well particularly for squares, it requires remodeling the cookie-cutter function, $f_I(\mathcal{G}^{obs})$, for any new shape. This is prohibitive in

terms of resources since any new shape first has to be printed and the deformation modeled prior to designing compensation plans. This is a limitation we attempt to overcome in this paper. Our main objective is thus to develop a single comprehensive model for all shapes, by developing a cookie-cutter approach from first principles. During this process, we encounter missing data, which we describe in section 4.1. Section 4.2 details models for handling the missing data, and describing point-to-point deformation, while the results based measured printed shapes are presented in Section 4.3. We discuss how a compensation plan based on the estimated relationships may be developed in Section 4.4 and conclude in Section 4.5.

4.1 DEFORMATION AS A MISSING DATA PROBLEM

We argue that deformation, as defined by (4.1), is inherently a missing data problem. To see this, consider the hypothetical in-plane deformation of a printed cube in Figure 4.1. Denoting the nominal angle, nominal radius function, observed angle and observed radius function by \mathcal{G}^{nom} , $r^{nom}(\cdot)$, \mathcal{G}^{obs} , and $r^{obs}(\cdot)$ respectively, any given point on the nominal profile can be represented by the tuple, $(\mathcal{G}^{nom}, r^{nom}(\mathcal{G}^{nom}))$, and the corresponding point after deformation has occurred is represented by $(\mathcal{G}^{obs}, r^{obs}(\mathcal{G}^{obs}))$. \mathcal{G}^{nom} , $r^{nom}(\cdot)$, \mathcal{G}^{obs} , and $r^{obs}(\cdot)$ have traditionally been defined with respect to a marked origin. However, marked origins are not always feasible, especially for donut shaped objects (objects with a central hole where the origin should be). We thus depart from that definition and redefine the angles and radii with respect to the centroid of the given shape. For the nominal shapes, the centroid and marked origin coincide, although this does not necessarily hold for the observed shapes, due to possibly asymmetric deformation.

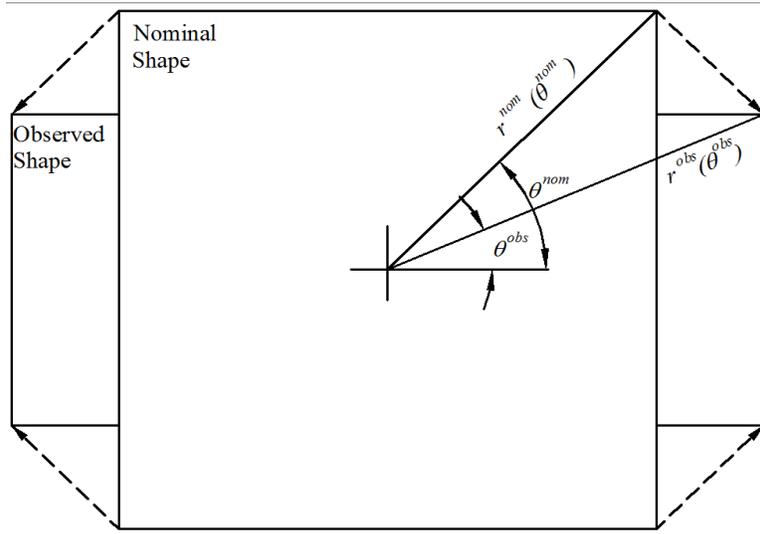


Figure 4.1: In-plane nominal and observed angles and radii for a hypothetical 3D printed cube.

The point-to-point difference is given by the vector $(\Delta_t, \Delta_r) = (\mathcal{G}^{obs} - \mathcal{G}^{nom}, r^{obs}(\mathcal{G}^{obs}) - r^{nom}(\mathcal{G}^{nom}))$. To relate this notion of deformation to the Polar-coordinate-based definition as in (4.1), note that the deformation in (4.1) can be written as $\Delta = r^{obs}(\mathcal{G}^{obs}) - r^{nom}(\mathcal{G}^{obs})$, where $r^{nom}(\mathcal{G}^{obs}) = r_o$ in (4.1). Performing a Taylor series expansion of $r^{nom}(\mathcal{G}^{obs})$ around $r^{nom}(\mathcal{G}^{nom})$, and denoting the first two derivatives of $r^{nom}(\cdot)$ by $\dot{r}^{nom}(\cdot)$ and $\ddot{r}^{nom}(\cdot)$ respectively, the deformation can thus be re-written in terms of Δ_t and Δ_r as

$$\Delta = \Delta_r - \Delta_t \cdot \dot{r}^{nom}(\mathcal{G}^{nom}) - \frac{\Delta_t^2}{2} \cdot \ddot{r}^{nom}(\mathcal{G}^{nom}) - \dots \quad (4.3)$$

Thus Δ is simply a specific 1-dimensional function of the 2-dimensional point-to-point deformations. This expansion in (4.3) is an important one because it gives insights on how the cookie-cutter model may be developed. For a cylinder the corresponding in-plane shape is simply a circle, the nominal radius function ($r^{nom}(\cdot)$) is constant with zero-valued derivatives, and thus $\Delta = \Delta_r$. For any general shape with nominal radius function, $r^{nom}(\cdot)$, (4.3) implies that the corresponding cookie

cutter function should depend on derivatives of $r^{nom}(\cdot)$, as well as the angle deformation, Δ_r . For example, consider measurements obtained from a printed square in Fig. 4.2 and a free-form concave shape in Fig. 4.3.

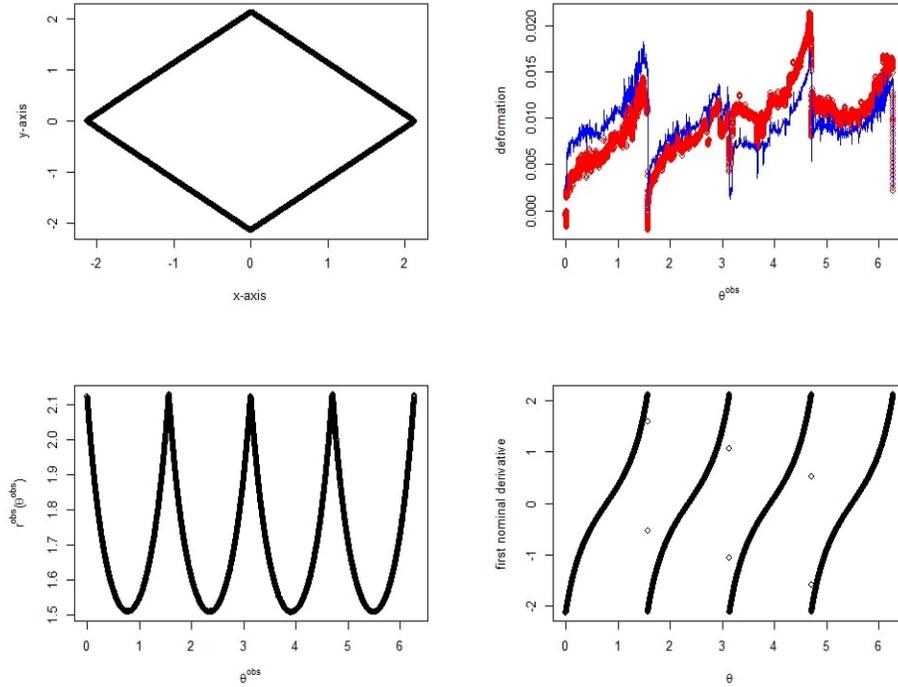


Figure 4.2: In-plane square shape in Cartesian (top) and Polar coordinates (bottom) on the left side. On the top right side, the deformation Δ , based on a marked origin in red, and the deformation based on centroid center in blue. The bottom right shows the first derivative of $r^{nom}(\cdot)$.

It can be seen that the first derivative of $r^{nom}(\cdot)$ plays a major role in determining the shape of the deformation, with the specific shape determined by Δ_r , the angle deformations. The top right plots of Figs. 4.2 and 4.3 show that this argument holds whether we define \mathcal{G}^{nom} , $r^{nom}(\cdot)$, \mathcal{G}^{obs} , and $r^{obs}(\cdot)$ based on a marked origin (in red) or we define them based on the centroid of the respective nominal and observed shapes.

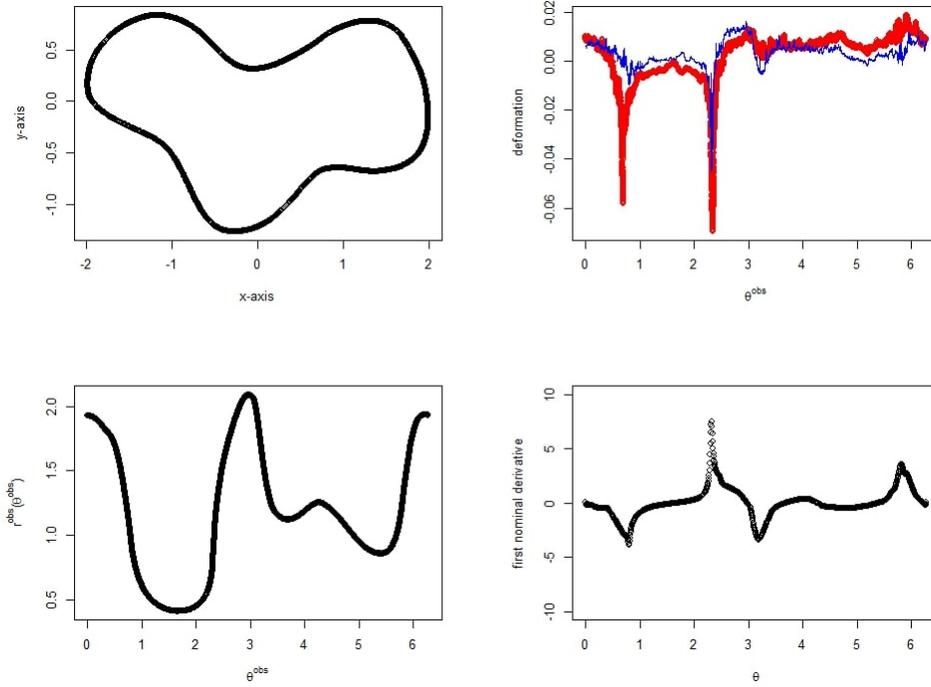


Figure 4.3: In-plane concave shape in Cartesian (top) and Polar coordinates (bottom) on the left side. On the top right side, the deformation Δ , based on a marked origin in red, and the deformation based on centroid center in blue. The bottom right shows the first derivative of $r^{nom}(\cdot)$.

Prior to manufacturing the 3D printed shape, the nominal radius function, $r^{nom}(\cdot)$, is known. After the printing process is complete, measurements are made along the profile of the object, thus giving the tuple, $(\mathcal{S}^{obs}, r^{obs})$, for each profile point. Deformation is thus a missing data problem because the corresponding nominal angle, \mathcal{S}^{nom} , for each measured point is almost always unknown.

Estimating \mathcal{S}^{nom} is often possible due to a combination of supplementary data (in the form of tracked markers), and landmark information (e.g., location of corners/peaks). Specifically, we employ multiple imputation techniques⁴⁸, since they account for the full uncertainty created by the missing data. Note that for 3D printed cylinders, $r^{nom}(\cdot)$ is always constant. Therefore, estimation of \mathcal{S}^{nom} only becomes necessary when we depart from cylindrical shapes.

4.2 MODELS

From (4.3), it is clear that model building for the deformation, Δ , should proceed in two dimensions: (i) a model for $\Delta_t = \mathcal{Y}^{obs} - \mathcal{Y}^{nom}$, and (ii) another for $\Delta_r = r^{obs}(\mathcal{Y}^{obs}) - r^{nom}(\mathcal{Y}^{nom})$. Since Δ_r depends on Δ_t through the unknown \mathcal{Y}^{nom} , Δ_t has to be modeled first. Our modeling strategy is thus:

1. Fit a shape-independent model for Δ_t using a combination of available landmarks and markers whose positions are tracked from the nominal to the observed shape.
2. Fit a model for Δ_r using profile data for each shape, imputing $\mathcal{Y}^{nom} = \mathcal{Y}^{obs} - \Delta_t$ for each point in each shape.

We now proceed with the model for Δ_t .

4.2.1 EMPIRICAL BAYES IMPUTATION MODEL FOR MISSING ANGLES

We denote the list of available nominal marker and landmark locations, and corresponding observed locations by $\{\mathcal{Y}_M^{nom}, \mathcal{Y}_M^{nom}, \mathcal{Y}_M^{obs}, \mathcal{Y}_M^{obs}\}$. Denoting the list of nominal and observed angles and radii for points on each profile by \mathcal{Y}^{nom} , \mathcal{Y}^{nom} , \mathcal{Y}^{obs} , and \mathcal{Y}^{obs} respectively, the imputation model has the form

$$\mathcal{Y}^{nom} | \mathcal{Y}^{obs}, \mathcal{Y}^{obs}, \{\mathcal{Y}_M^{nom}, \mathcal{Y}_M^{obs}, \mathcal{Y}_M^{obs}\}, \varphi \sim \mathcal{F}(\cdot), \quad (4.4)$$

where $\mathcal{F}(\cdot)$ is some distribution parametrized by φ . Thus if we assume the following data generating mechanism under a Bayesian framework,

$$\begin{aligned} (\mathcal{Y}^{obs}, \mathcal{Y}^{obs}) | \mathcal{Y}^{nom}, \mathcal{Y}^{nom}, \varphi &\sim \mathcal{G}(\cdot) \\ (\mathcal{Y}^{nom}, \mathcal{Y}^{nom}) &\sim \Pi(\cdot), \end{aligned} \quad (4.5)$$

where $\mathcal{G}(\cdot)$ and $\Pi(\cdot)$ represent data and prior distributions respectively, it can be seen that $\mathcal{F}(\cdot)$ in (4.4) is simply the posterior predictive distribution for \mathcal{Y}^{nom} based on (4.5), and observed data, $\{\mathcal{Y}_M^{nom}, \mathcal{Y}_M^{obs}, \mathbf{r}_M^{obs}\}$.

For the purposes of imputation, there is no need to select models for $\mathcal{G}(\cdot)$ and $\Pi(\cdot)$. Instead we can directly model

$$\mathcal{Y}^{nom} | \mathcal{Y}^{obs}, \mathbf{r}^{obs}, \varphi \sim \mathcal{H}(\cdot), \quad (4.6)$$

where $\mathcal{H}(\cdot)$ represents the data distribution parametrized by φ . Thus \mathcal{F} from (4.4) once again becomes the posterior predictive distribution for \mathcal{Y}^{nom} based on (4.6), and observed data, $\{\mathcal{Y}_M^{nom}, \mathcal{Y}_M^{obs}, \mathbf{r}_M^{obs}\}$.

Specifically for any given set of profile points, we employ the additive model,

$$\mathcal{Y}^{nom} = \mathcal{Y}^{obs} + \mu \mathbf{1} + \nu(\mathcal{Y}^{obs}, \mathbf{r}^{obs}) + \varepsilon, \quad (4.7)$$

where μ represents the mean of the difference $\mathcal{Y}^{nom} - \mathcal{Y}^{obs}$ over the profile, $\nu(\cdot)$ is a realization from a zero-mean Gaussian process (GP)⁴⁴, and ε is a white noise vector, i.e.,

$$\nu(\vartheta, r) \sim GP(\mathbf{o}, \sigma^2 K((\vartheta, r), (\vartheta', r'))) \quad \text{and} \quad \varepsilon \sim \mathcal{N}(\mathbf{o}, \sigma_\varepsilon^2 \mathbf{I}).$$

Due to the differences in domain for the angle and radius dimensions, the usual covariance functions may be inappropriate. It is possible to warp the dimension for angles to correct for the difference in domain, as is done to obtain the periodic correlation function by warping the angle dimension and employing the squared exponential covariance function⁴⁴. However, this raises the issue of careful selection of appropriate bounds for the range parameter. It is easier instead to define the distance between two locations by their distance in Cartesian coordinates for the GP. Thus we can rewrite

$$\nu(\vartheta, r) \equiv \nu(x, y) \sim GP(\mathbf{o}, \sigma^2 K((x, y), (x', y'))).$$

For selection of the correlation function, $K(\cdot, \cdot)$, we turn to a special class of Matérn correlation functions⁴⁴ defined in 1-dimension by

$$k_a(x, x') = \left(1 + \frac{\sqrt{5}|x - x'|}{\varrho_a} + \frac{5(x - x')^2}{3\varrho_a^2} \right) \exp\left(-\frac{\sqrt{5}|x - x'|}{\varrho_a}\right),$$

where ϱ_a represents the range parameter. We thus complete the model for Δ_r by defining our correlation function as the tensor product

$$K((x, y), (x', y')) = k_a(x, x') \cdot k_b(y, y').$$

The imputation distribution, \mathcal{F} is now simply the conditional multivariate normal distribution. i.e., denoting the covariance matrix between \mathcal{Y}^{nom} and \mathcal{Y}_M^{nom} by $\mathbf{K}_{i,M}$, its transpose by $\mathbf{K}_{M,i}$, the covariance matrix for \mathcal{Y}_M^{nom} by $\mathbf{K}_{M,M}$, and the covariance matrix for \mathcal{Y}^{nom} by $\mathbf{K}_{i,i}$ we have that

$$\mathcal{Y}^{nom} | \mathcal{Y}^{obs}, \mathbf{r}^{obs}, \{\mathcal{Y}_M^{nom}, \mathcal{Y}_M^{obs}, \mathbf{r}_M^{obs}\}, \varphi \sim MVN(\mathbf{U}, \mathbf{V}), \quad (4.8)$$

where

$$\mathbf{U} = \mathcal{Y}^{obs} + \mu \mathbf{1} + \mathbf{K}_{i,M} \mathbf{K}_{M,M}^{-1} (\mathcal{Y}_M^{nom} - \mathcal{Y}_M^{obs} - \mu \mathbf{1}),$$

and

$$\mathbf{V} = \mathbf{K}_{i,i} - \mathbf{K}_{i,M} \mathbf{K}_{M,M}^{-1} \mathbf{K}_{M,i}.$$

Note that we have assumed $\varphi = (\mu, \varrho_x, \varrho_y, \sigma, \sigma_z)$ is known (or given) in (4.8), although this is often not the case. In our analysis, we simply replace φ with its maximum likelihood estimate (MLE) $\hat{\varphi}$, thus implementing an empirical Bayes procedure^{13,38,34,44}. We now turn to the task of modeling the radii deformation, Δ_r .

4.2.2 BAYESIAN MODEL FOR DEFORMATION

One implicit justification for the cookie-cutter model developed by Huang et al.²⁰ is that regular polygons share some global deformation properties, captured by the deformation of their respective circumcircles. We now formalize this justification by claiming that the shared global behavior of all shapes, in the absence of noise, is of the form

$$\frac{r^{obs}(\mathcal{Y}^{obs})}{r^{nom}(\mathcal{Y}^{nom})} = f(\mathcal{Y}^{nom}, r^{nom}(\mathcal{Y}^{nom})). \quad (4.9)$$

To support this claim, we compare $r^{obs}(\mathcal{Y}^{obs})/r^{nom}(\mathcal{Y}^{nom})$ for three shapes in Fig. 4.4: a circle with radius 3 inches, a pentagon with circumcircle radius of 3 inches and a dodecagon of circumcircle radius 3 inches. In order to make this comparison, we have used \mathbf{U} in (4.8) to estimate and impute \mathcal{Y}^{nom} .

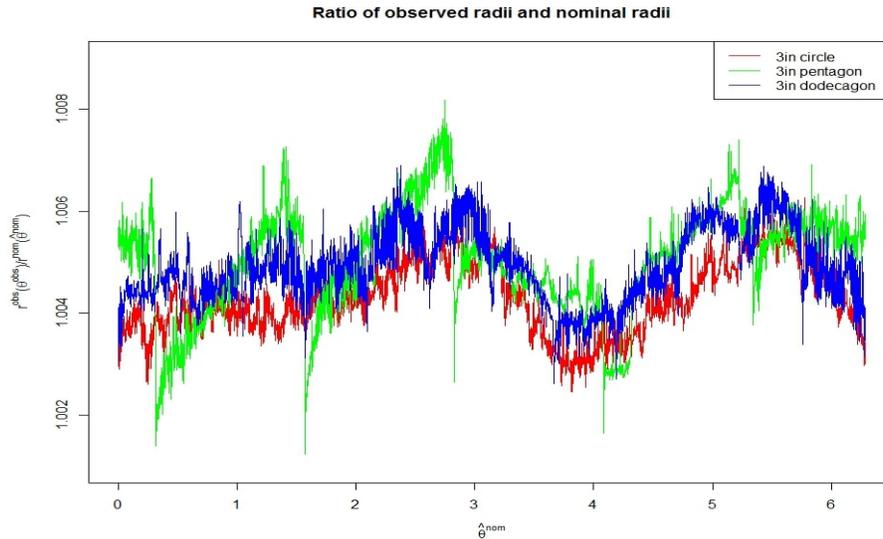


Figure 4.4: Ratio of observed radii and nominal radii (after single imputation) for three different shapes.

It can be seen that the ratio, $r^{obs}(\mathcal{Y}^{obs})/r^{nom}(\mathcal{Y}^{nom})$, reveals the global shrinkage patterns sufficiently well and suggests a sinusoidal functional form for $f(\cdot, \cdot)$ in (4.9). To further reveal the depen-

dence of $f(\cdot, \cdot)$ on the nominal radius, $r^{nom}(\cdot)$, we compare $r^{obs}(\mathcal{Y}^{obs})/r^{nom}(\mathcal{Y}^{nom})$ for three circles of different radii in Fig. 4.5.

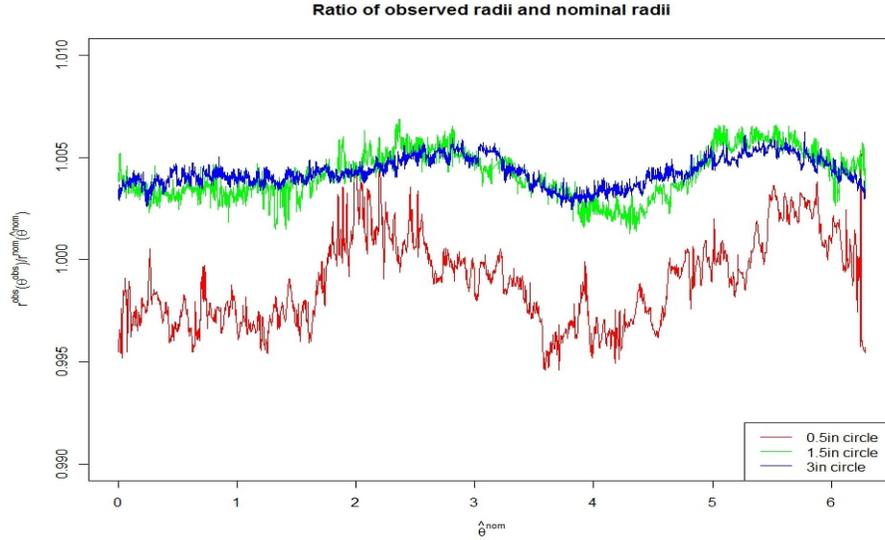


Figure 4.5: Ratio of observed radii and nominal radii for three circles with different sizes.

The sinusoidal pattern remains present, although the magnitude changes with respect to $r^{nom}(\cdot)$. In constructing a model for deformation in cylinders, Huang et al.²¹ observed this same phenomenon and note that the shrinkage of any printed shape should be proportional to the entire volume of the product. Applying the same power models used by Huang et al.²¹ for volume-dependent shrinkage parameters, we propose the following form for $f(\cdot, \cdot)$,

$$f(\mathcal{Y}^{nom}, r^{nom}(\mathcal{Y}^{nom})) = \alpha(r^{nom}(\mathcal{Y}^{nom}))^a + \beta(r^{nom}(\mathcal{Y}^{nom}))^b \cos\left(2\left(\mathcal{Y}^{nom} - \frac{\pi}{4}\right)\right), \quad (4.10)$$

and the following model for the noisy data,

$$r^{obs}(\mathcal{Y}^{obs}) = r^{nom}(\mathcal{Y}^{nom}) \cdot f(\mathcal{Y}^{nom}, r^{nom}(\mathcal{Y}^{nom})) + e, \quad (4.11)$$

where $e \sim \mathcal{N}(0, \sigma_e^2)$ describes the noise distribution. We have defined the constant variance noise on the scale of the observed radii, so that the noisy ratio, $r^{obs}(\mathcal{Y}^{obs})/r^{nom}(\mathcal{Y}^{nom})$ has a noise magnitude inversely proportional to the nominal radii, $r^{nom}(\mathcal{Y}^{nom})$, as observed in Fig 4.5. To obtain either Δ_r or Δ from (4.11), we simply subtract $r^{nom}(\mathcal{Y}^{nom})$ or $r^{nom}(\mathcal{Y}^{obs})$ respectively. Introducing flat priors for the parameters under a Bayesian framework, our final model for $r^{obs}(\mathcal{Y}^{obs})$ is given by

$$r^{obs}(\mathcal{Y}^{obs}) \sim \mathcal{N}\left(\alpha(r^{nom}(\mathcal{Y}^{nom}))^{a+1} + \beta(r^{nom}(\mathcal{Y}^{nom}))^{b+1} \cos\left(2\left(\mathcal{Y}^{nom} - \frac{\pi}{4}\right)\right), \sigma_e^2\right) \\ \pi(\alpha, a, \beta, b, \log(\sigma_e)) \propto \mathbf{1}. \quad (4.12)$$

4.3 ESTIMATION AND RESULTS

In estimating the models, we analyze 10 printed shape datasets: 6 circles (0.5, 1.5, 3, 0.5, 1 and 2 inch radii respectively), 2 squares (2 and 3 inch side lengths), a pentagon (1 inch circumcircle) and a dodecagon (3 inch circumcircle). These are shown in Fig. 4.6. We also have 36 markers recorded (from the nominal shapes to the final shapes after deformation) for each of 2 circles (1 and 3 inch radii) and a square (3 inch side length). Landmarks are chosen to be corner points of the 10 training datasets where they exist.

We first fit the model for Δ_r in section 4.2.1, using the DiceKriging package⁴⁷ in the *R* statistical software. The MLE is obtained as $\hat{\phi} = (\hat{\mu} = 0.0003, \hat{\xi}_a = 0.5630, \hat{\xi}_b = 0.6279, \hat{\sigma} = 0.0009, \hat{\sigma}_z = 0.0013)$. Based on these estimates, we calculate the posterior (conditional) means and standard errors of Δ_r for a grid of new points in Cartesian coordinates using (4.8). The posterior means are shown in Fig. 4.7, while the posterior standard errors are shown in Fig. 4.8.

The complex contour patterns in Fig. 4.7 provide some justification for modeling Δ_r through the non-parametric Gaussian process model we used. In Fig. 4.8, it can be seen that we have more uncertainty in locations where there is a dearth of marker data, and this suggests possible marker

Training datasets

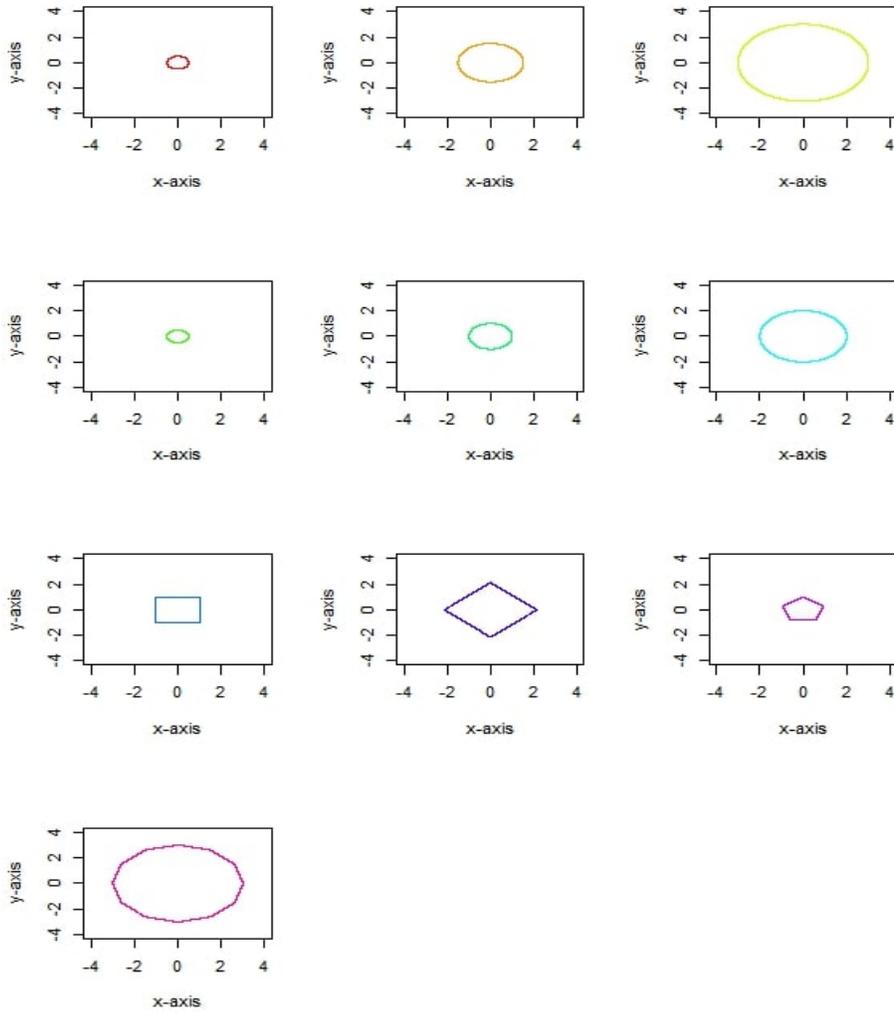


Figure 4.6: Observed profile measurements for 10 shapes: 6 circles (0.5, 1.5, 3, 0.5, 1 and 2 inch radii respectively), 2 squares (2 and 3 inch side lengths), a pentagon (1 inch circumcircle) and a dodecagon (3 inch circumcircle).

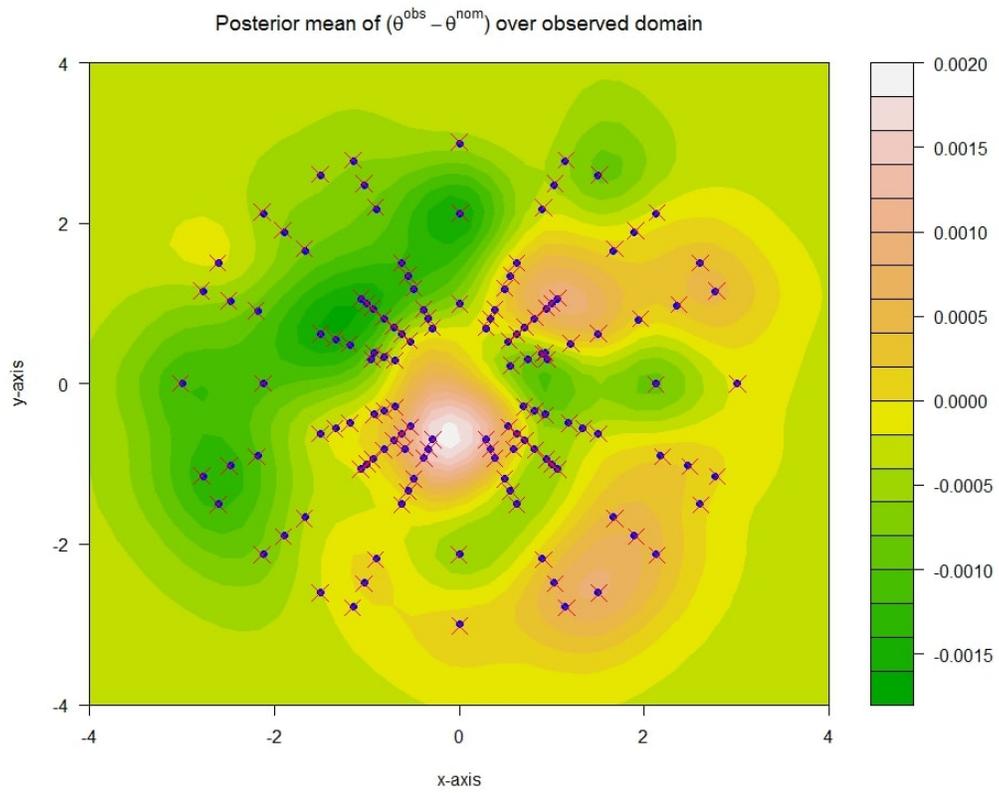


Figure 4.7: Contour plot for the posterior mean of Δ_r , overlaid by blue dots and red crosses representing all available marker (and landmark) locations on nominal and observed shapes respectively.

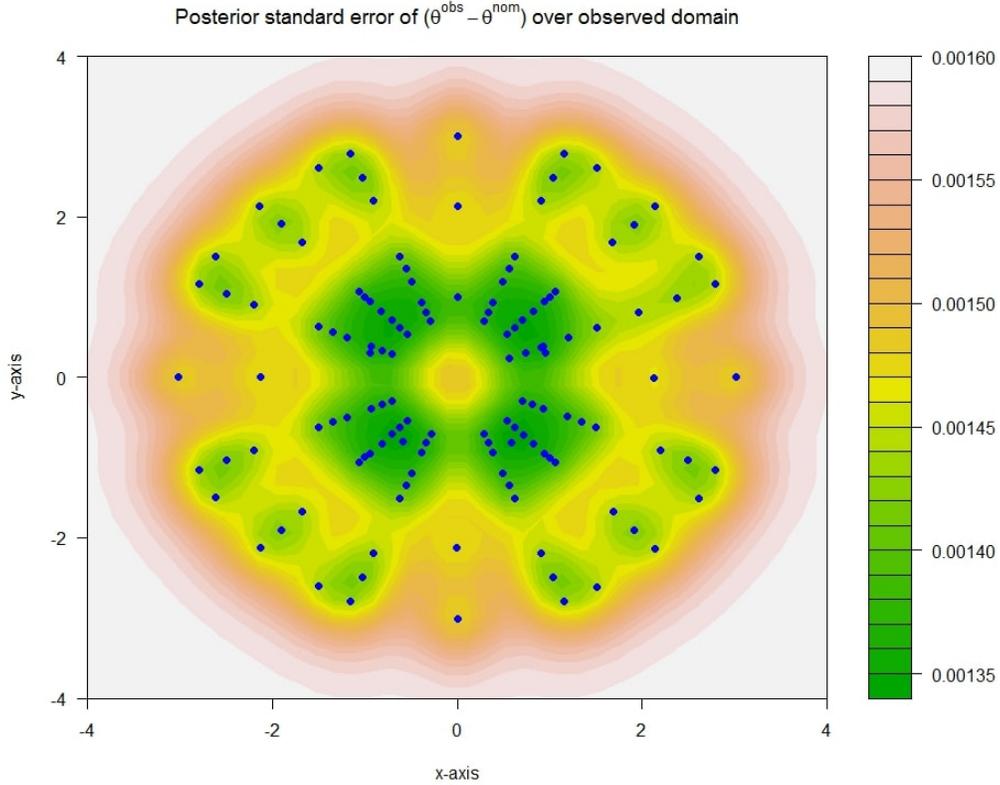


Figure 4.8: Contour plot for the posterior standard error of Δ_t , overlaid by blue dots representing all available marker (and landmark) locations on observed shapes respectively.

locations when planning future printing experiments.

To fit the model for $r^{\text{obs}}(\mathcal{Y}^{\text{obs}})$ in (4.12), we perform multiple imputations for \mathcal{Y}^{nom} under high-dimensionality considerations, using the estimated model for Δ_t and the provided \mathcal{Y}^{obs} . Specifically, due to the high-dimensionality of \mathcal{Y}^{obs} , we only sample Δ_t for some representative set of \mathcal{Y}^{obs} (chosen to equi-spaced in the interval $(0, 2\pi]$). All \mathcal{Y}^{obs} not in the representative set are then obtained by interpolation. Although this raises some issues about correct quantification of the uncertainty in the model, we choose the size of the representative set to be as high as possible under computational constraints, so that the effect of the interpolation on uncertainty quantification is minimized.

The deformation profile at the observed angles, $\Delta = r^{obs}(\mathcal{S}^{obs}) - r^{nom}(\mathcal{S}^{obs})$, for our 10 training datasets are shown in Fig. 4.9

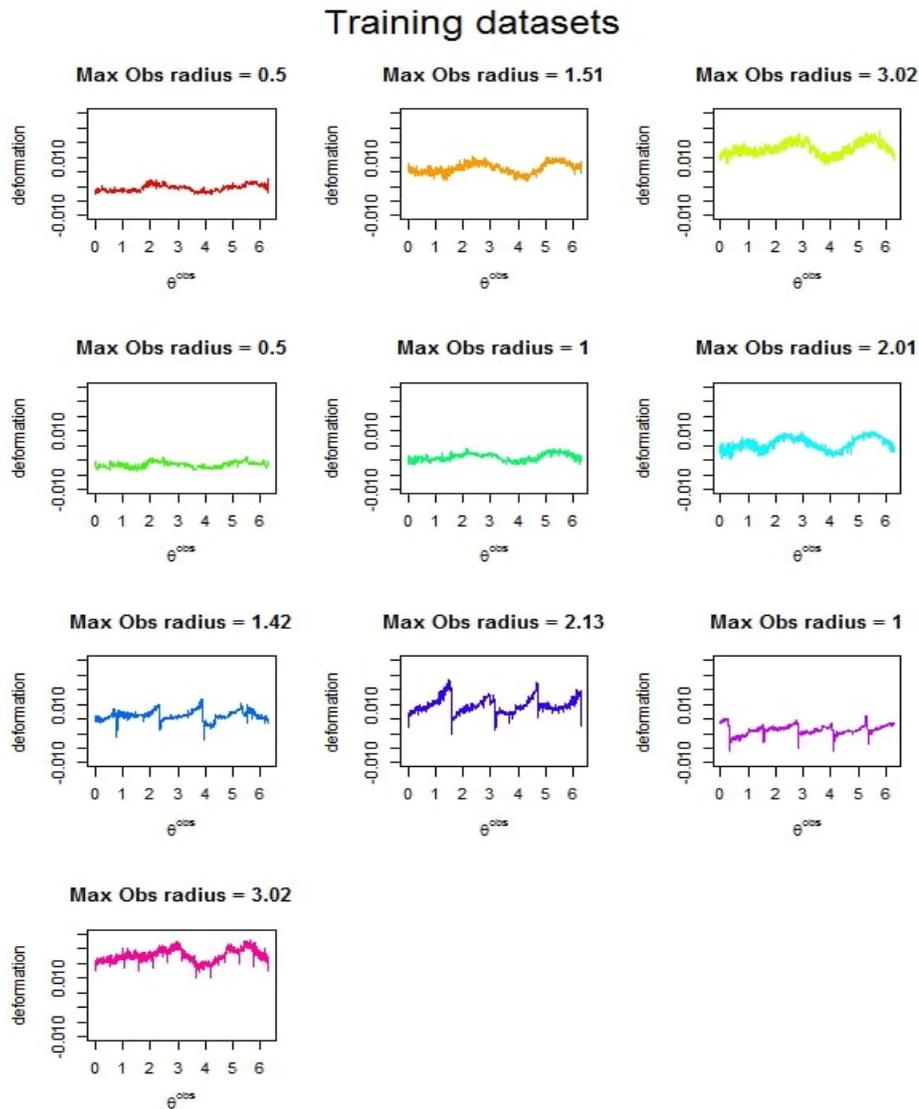


Figure 4.9: Observed profile deformations ($\Delta = r^{obs}(\mathcal{S}^{obs}) - r^{nom}(\mathcal{S}^{obs})$) for 10 shapes: 6 circles (0.5, 1.5, 3, 0.5, 1 and 2 inch radii respectively), 2 squares (2 and 3 inch side lengths), a pentagon (1 inch circumcircle) and a dodecagon (3 inch circumcircle).

Performing Markov Chain Monte-Carlo (MCMC)¹⁷ draws over the posterior distribution of $(\alpha, a, \beta, b, \log(\sigma_\epsilon))$ based on (4.12), we estimate Δ by its posterior predictive mean, and provide 95% posterior intervals (see Appendix C.1 for the MCMC diagnostic plots). The posterior predictive mean and intervals are shown in Fig. 4.10. The model fit appears to be good in general, although it seems to have a little trouble accurately capturing the deformation at corners of the squares.

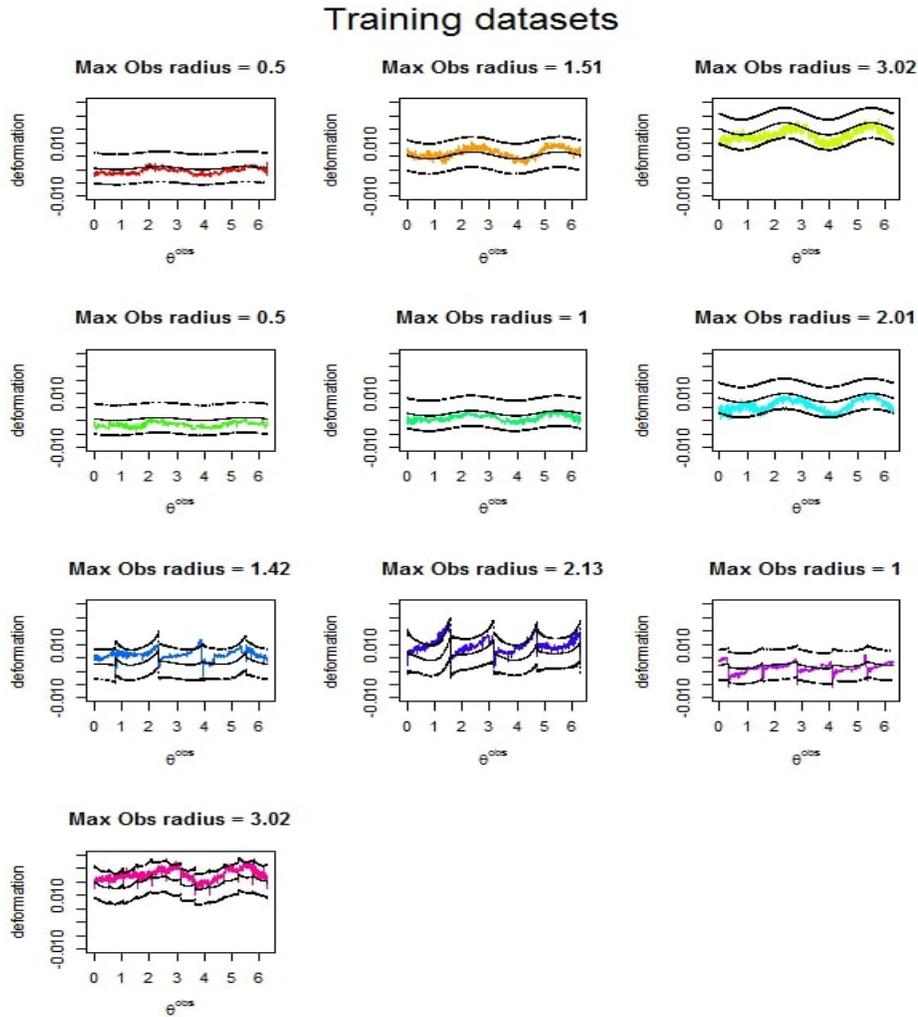


Figure 4.10: Observed profile deformations, posterior mean, and 95% posterior intervals for $(r^{obs}(\mathcal{Y}^{obs}) - r^{nom}(\mathcal{Y}^{obs}))$ based on the 10 training datasets.

We perform an external validation of the model by predicting the deformation of 5 new printed shapes: a pentagon (3 inch circumcircle radius), a convex free-form shape, a rectangle (1.6 by 1.2 inches), a dodecagon (3 inch circumcircle radius), and a concave free-form shape. All that is needed for the prediction of each new shape is the nominal radius function, $r^{nom}(\cdot)$, along with the estimated models from section 4.2. These new shapes are shown in Fig. 4.11, along with model predictions and 95% predictive intervals in Figure 4.12. The predictions for the polygons are generally good, except at the rectangle's corners. Considering that the predictive model was trained on circles and polygons, the predictions for the free-form shapes are especially good, despite the sharp dip not captured by the prediction intervals for the concave free-form deformation.

Test datasets

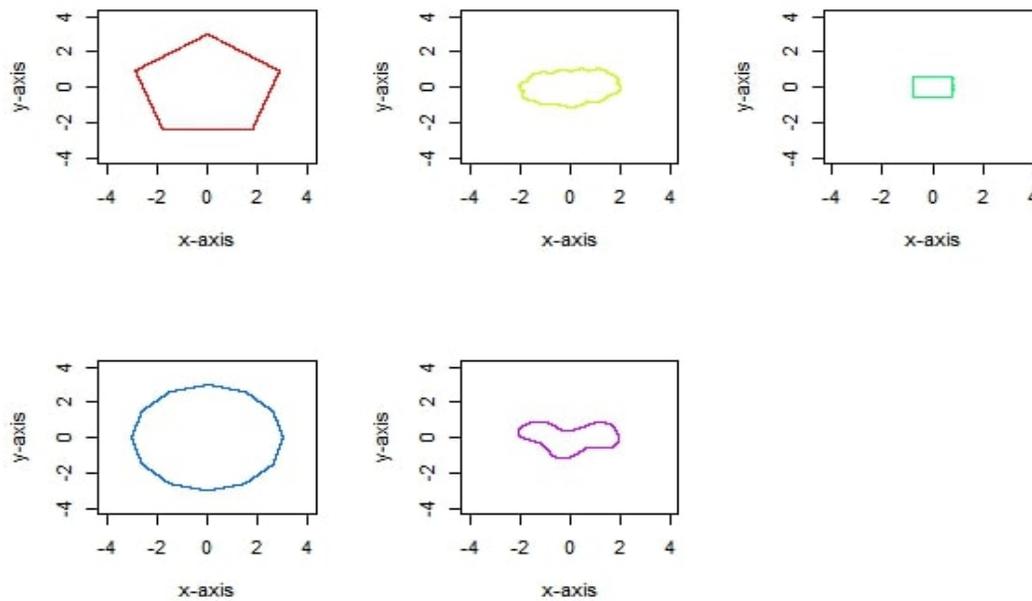


Figure 4.11: Observed profile measurements in Cartesian coordinates for the 5 test shapes: a pentagon (3 inch circumcircle radius), a convex free-form shape, a rectangle (1.6 by 1.2 inches), a dodecagon (3 inch circumcircle radius), and a concave free-form shape.

Test datasets

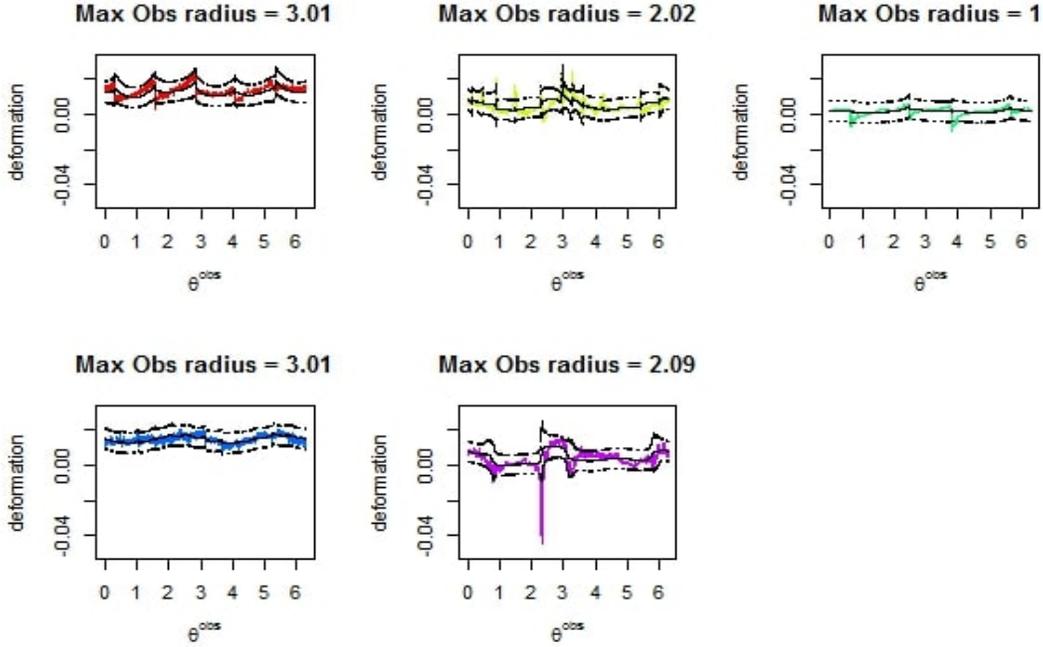


Figure 4.12: Observed profile deformations, posterior mean, and 95% posterior intervals for $(r^{obs}(\mathcal{Y}^{obs}) - r^{nom}(\mathcal{Y}^{obs}))$ based on the 5 test shapes: a pentagon (3 inch circumcircle radius), a convex free-form shape, a rectangle (1.6 by 1.2 inches), a dodecagon (3 inch circumcircle radius), and a concave free-form shape.

4.4 COMPENSATION PLAN DESIGN

The ultimate purpose of the deformation models we propose and estimate in sections 4.2 and 4.3 is to design compensation plans for minimizing the deformation. We propose the following strategy for obtaining a desired radius of r^* at an angle \mathcal{Y}^* :

- i. Obtain an estimate of the \mathcal{Y}^{nom} which deforms to \mathcal{Y}^* , denoted by $\tilde{\mathcal{Y}}$, using the posterior predictive mean based on (4.8)

$$\tilde{\mathcal{Y}} = E(\mathcal{Y}^{nom} | \mathcal{Y}^{obs} = \mathcal{Y}^*, r^{obs} = r^*, \{\mathcal{Y}_M^{nom}, \mathcal{Y}_M^{obs}, \mathbf{r}_M^{obs}\}, \varphi = \hat{\varphi}),$$

where $\hat{\phi}$ denotes the MLE of ϕ .

2. Obtain an estimate of the r^{nom} which deforms to r^* , denoted by \tilde{r} , by solving for r in the equation

$$\hat{\alpha}r^{\hat{a}+1} + \hat{\beta}r^{\hat{b}+1} \cos\left(2\left(\tilde{\vartheta} - \frac{\pi}{4}\right)\right) = r^*,$$

where $\hat{\alpha}$, $\hat{\beta}$, \hat{a} and \hat{b} refer to posterior mean of the respective parameters.

Thus for every point, (ϑ^*, r^*) , in the profile of a desired shape, we obtain corresponding estimates of the compensation, $c(\tilde{\vartheta}) = \tilde{r} - r^{nom}(\tilde{\vartheta})$ to added at the angle $\tilde{\vartheta}$.

We can obtain an estimate of the uncertainty around this compensation plan based on $(\tilde{\vartheta}, \tilde{r})$, by using Monte-Carlo techniques. Specifically, for a pre-specified number of samples N , we use the following strategy

1. For $i = 1, \dots, N$, sample ϑ_i based on the distribution in (4.8).
2. Using the distribution in (4.11), sample $Z_i \sim \mathcal{N}(0, 1)$ and solve for r_i in the equation.

$$\hat{\alpha}r_i^{\hat{a}+1} + \hat{\beta}r_i^{\hat{b}+1} \cos\left(2\left(\vartheta_i - \frac{\pi}{4}\right)\right) + \hat{\sigma}_e Z_i = r^*.$$

3. Convert each (ϑ_i, r_i) to Cartesian coordinates (x_i, y_i) , and convert $(\tilde{\vartheta}, \tilde{r})$ to (\tilde{x}, \tilde{y}) as well.
4. Calculate the quantity

$$D = \frac{1}{N-1} \begin{pmatrix} \sum_{i=1}^N (x_i - \tilde{x})^2 & \sum_{i=1}^N (x_i - \tilde{x})(y_i - \tilde{y}) \\ \sum_{i=1}^N (x_i - \tilde{x})(y_i - \tilde{y}) & \sum_{i=1}^N (y_i - \tilde{y})^2 \end{pmatrix}$$

5. Report $\tau = \sqrt{\text{trace}(D)}$, the Frobenius norm of $D^{1/2}$ (where $(D^{1/2})^T D^{1/2} = D$), as the uncertainty associated with using $(\tilde{\vartheta}, \tilde{r})$. Note that τ is also the root mean squared distance

of possible nominal points (i.e., (x_i, y_i)) for getting the desired observed point (i.e., (θ^*, r^*)), from the nominal point estimate, (\tilde{x}, \tilde{y}) .

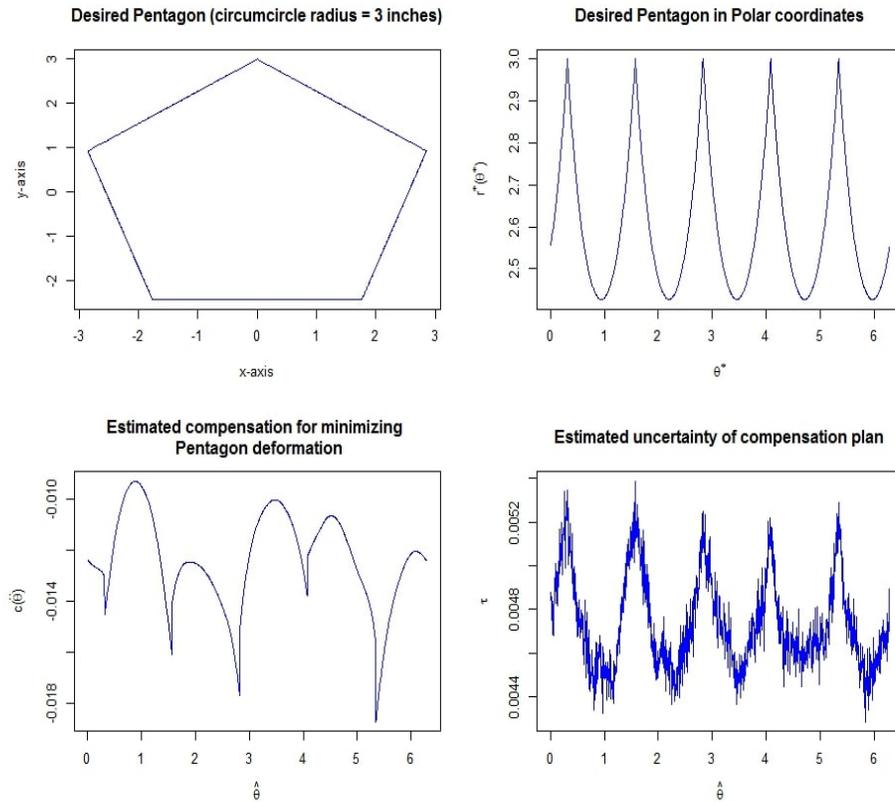


Figure 4.13: Compensation plan for pentagon with 3 inch circumcircle radius.

For example, we design a compensation design for a pentagon with circumcircle radius of 3 inches, based on our estimated models in section 4.3. The desired shape, compensation plan, and uncertainty around that compensation plan are depicted in Fig. 4.13.

It can be seen that compensation around the corners of the pentagon exhibit the most uncertainty. This extra uncertainty near the corners can be explained by smoothing effect of light blurring, as noted by Xu et al.⁶⁶. This smoothing effect, resulting from energy convolution of neigh-

boring pixels⁶⁶, thus has a more drastic effect on shape corners, as detected by our point-to-point deformation models.

4.5 CONCLUDING REMARKS

In this paper we have developed a single framework for modeling the deformation of 3D printed products based on point-wise deformations. We have shown that the printed shapes share global deformation/shrinkage features which can be modeled separately from the local deformation features based on the cookie-cutter idea by Huang et al.²⁰. We have also shown that the local deformation features depend on the derivatives of the nominal radius function, and involve missing data which can be handled using multiple imputation and tracked markers/landmarks.

External validation of our fitted model shows good predictions even for shapes not yet encountered, thus confirming a strong shared global deformation pattern. However, improvements can be made for predicting local deformation patterns through the aid of more markers, particularly in regions where such markers are sparse. Based on the point-wise deformation models, we have proposed a compensation plan procedure and we have shown as an example, a compensation plan design, along with uncertainty quantification, for achieving a pentagon with 3 inch circumcircle radius.

Future research directions include designing a fully Bayesian framework for imputing the missing angles, accounting for dependence along the deformation profile instead of independent errors, and developing strategies for handling the high-dimensional data.



Appendix for Chapter I

A.0 GLOSSARY OF NOTATIONS

- $g_i(\mathcal{D})$ → the random shape of island i on the copper tray.
- $q(t)$ → the kinetics driving the growth on the tray.
We denote $Q(t) = \int_0^t q(s) ds$ and define $Q(0) = 0$.
- t_{oi} → the nucleation time of island i on the tray.
- $v_i(\mathcal{D}, t)$ → the angular dependent growth velocity of island i .
- $N(t)$ → the (random) number of islands on the tray at time t .
- $S_i(t_{oi}, t)$ → the area covered by island i with nucleation time t_{oi} , by time t .
- $S_j(t)$ → the total area covered by the graphene islands on the tray by time t .

Note that $S(t) = \sum_{i=1}^{\infty} S_i(t) \cdot \mathbf{1}\{t_{oi} \leq t\} = \sum_{i=1}^{N(\infty)} S_i(t) \cdot \mathbf{1}\{t_{oi} \leq t\} = \sum_{i=1}^{N(t)} S_i(t)$
 where $\mathbf{1}\{t_{oi} \leq t\}$ is the usual indicator function and $N(t) = \sum_{i=1}^{\infty} \mathbf{1}\{t_{oi} \leq t\}$.

- $S^\circ(t)$ → the total area covered by the graphene islands on the tray by time t ,
 when $t_{oi} = 0$ for every i .
- $r_i(\vartheta, t_{oi}, t)$ → the length/distance from the edge of the i th island with nucleation
 time t_{oi} , to its center, at a specific angle ϑ and time t (for a circular island
 shape, this is simply the radius at time t).
 Also note that $r_i(\vartheta, t_{oi}, t) = 0$ if $t_{oi} > t$ (as implied by⁶⁴).
- $\tilde{v}_i(\vartheta, t)$ → the angular dependent growth velocity of island i on the tray for
 the average growth process (i.e., averaging out the random behavior of
 carbon atom attachment/detachment). Recall that we assumed
 $\tilde{v}_i(\vartheta, t) = g_i(\vartheta)q(t)$.
- $\tilde{r}_i(\vartheta, t_{oi}, t)$ → the length/distance from the edge of the i th island with nucleation
 time t_{oi} , to its center, at a specific angle ϑ and time t , after averaging out
 the random carbon atom attachment/detachment.
 $\tilde{r}_i(\vartheta, t_{oi}, t) = \int_{t_{oi}}^t \tilde{v}_i(\vartheta, s) ds$.
- $\tilde{S}_i(t_{oi}, t)$ → the average area covered by island i with nucleation time t_{oi} , by time t
 i.e. averaging out the random behavior of carbon atom attachment/detachment.
 Note: $\tilde{S}_i(t_{oi}, t) = E(S_i(t_{oi}, t)|t_{oi})$, and $\tilde{S}_i(t_{oi}, t) = \frac{1}{2} \int_0^{2\pi} \tilde{r}_i^2(\vartheta, t_{oi}, t) d\vartheta$.

A.1 PROOF OF THEOREM 1.2.1

Given $t_{oi} \leq t$ for i th graphene island on the copper tray, we can write

$$\begin{aligned}
 \frac{d}{dt} \tilde{S}_i(t_{oi}, t) &= \frac{1}{2} \frac{d}{dt} \int_{\circ}^2 \tilde{r}_i^2(\vartheta, t_{oi}, t) d\vartheta \\
 &= \int_{\circ}^2 \tilde{v}_i(\vartheta, t) \left(\int_{t_{oi}}^t \tilde{v}_i(\vartheta, s) ds \right) d\vartheta \\
 &= \left(\int_{\circ}^2 g_i^2(\vartheta) d\vartheta \right) q(t) \int_{t_{oi}}^t q(s) ds \\
 &= 2A_{nuc} \cdot q(t) \int_{t_{oi}}^t q(s) ds.
 \end{aligned}$$

Now integrating both sides with respect to time t , we obtain

$$\tilde{S}_i(t_{oi}, t) + Const = 2A_{nuc} \left[\frac{1}{2} Q^2(t) - Q(t_{oi})Q(t) \right].$$

When $t = t_{oi}$, we have $\tilde{S}_i(t_{oi}, t) = 0$ and thus

$$Const = -2A_{nuc} \left(\frac{1}{2} Q^2(t_{oi}) \right),$$

so that $\tilde{S}_i(t_{oi}, t) = A_{nuc} [Q(t) - Q(t_{oi})]^2$. Since $Q(0) = 0$, we also get $\tilde{S}_i(0, t_{oi}) = A_{nuc} Q^2(t_{oi})$,

which implies that

$$Q(t_{oi}) = \sqrt{\frac{\tilde{S}_i(0, t_{oi})}{A_{nuc}}}.$$

The conditional mean for $\tilde{S}_i(t_{oi}, t)$ is given as

$$E(\tilde{S}_i(t_{oi}, t) | t_{oi} \leq t) = A_{nuc} \cdot E([Q(t) - Q(t_{oi})]^2 | t_{oi} \leq t)$$

$$= A_{nuc} \cdot [(Q(t) - E(Q(t_{oi})|t_{oi} \leq t))^2 + Var(Q(t_{oi})|t_{oi} \leq t)] .$$

Denoting

$$\begin{aligned} \xi(t) &= E(\tilde{S}_i(t_{oi}, t)|t_{oi} \leq t) , \\ \phi(t) &= \sqrt{A_{nuc}} \cdot E(Q(t_{oi})|t_{oi} \leq t) = E(\sqrt{\tilde{S}_i(0, t_{oi})}|t_{oi} \leq t) , \text{ and} \\ \hat{\sigma}(t) &= A_{nuc} \cdot Var(Q(t_{oi})|t_{oi} \leq t) = Var(\sqrt{\tilde{S}_i(0, t_{oi})}|t_{oi} \leq t) , \end{aligned}$$

We can now write

$$Q(t) = \frac{1}{\sqrt{A_{nuc}}} \left[\sqrt{\xi(t) - \hat{\sigma}(t)} + \phi(t) \right] \text{ and } q(t) = \frac{1}{\sqrt{A_{nuc}}} \left[\frac{\xi'(t) - \hat{\sigma}'(t)}{2\sqrt{\xi(t) - \hat{\sigma}(t)}} + \phi'(t) \right] ,$$

completing the proof.

A.2 DIAGNOSTIC PLOTS

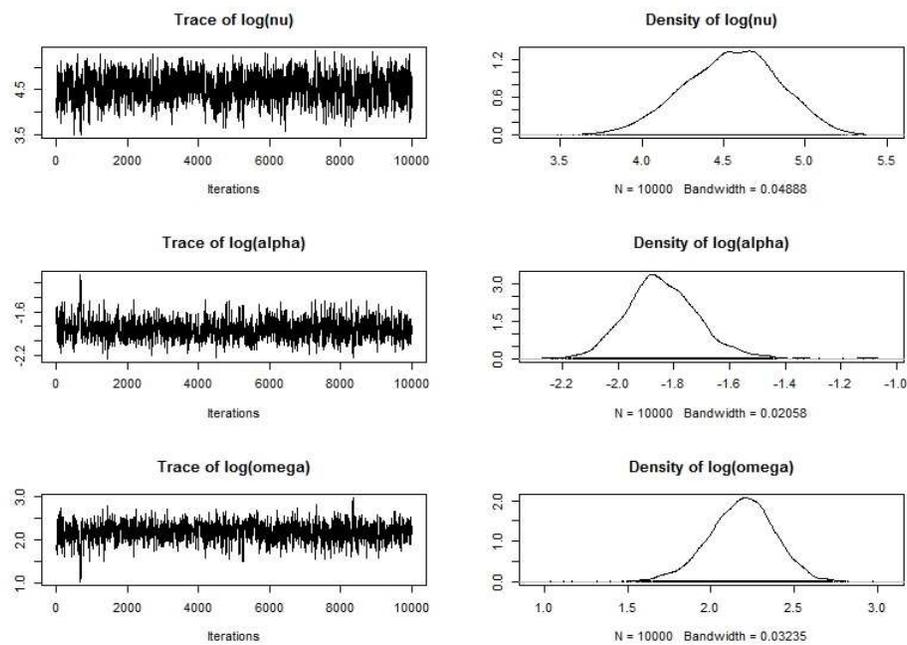


Figure A.1: 10000 MCMC draws for each variable after discarding the burn-in. The plots indicate good mixing.

B

Appendix for Chapter 2

B.1 SIGNAL TO NOISE RATIO

First note from (2.14) that by completing the square, we can rewrite μ_X as

$$\begin{aligned}\mu_X &= \beta_2 \left(X + \frac{\beta_1}{2\beta_2} \right)^2 + \left(\beta_0 - \frac{\beta_1^2}{4\beta_2} \right) \\ &= \beta_2 \tau^2 \eta(X) + \left(\beta_0 - \frac{\beta_1^2}{4\beta_2} \right),\end{aligned}$$

$$\text{where } \sqrt{\eta(X)} = \frac{1}{\tau} \left(X + \frac{\beta_1}{2\beta_2} \right) \sim \mathcal{N} \left(\frac{1}{\tau} \left(\gamma Z + \frac{\beta_1}{2\beta_2} \right), 1 \right),$$

so that $\eta(X)$ is itself a non-central chi-square random variable with $k = 1$ degree of freedom and $\lambda(Z)$ non-centrality parameter

$$\eta(X) \sim \chi^2 \left(k = 1, \lambda(Z) = \frac{1}{\tau^2} \left(\gamma Z + \frac{\beta_1}{2\beta_2} \right)^2 \right),$$

and thus μ_X is simply a scaled and shifted non-central chi-squared random variable, which makes calculating its moments straightforward. Recall that the mean and variance of a non-central chi-square variable with k degrees of freedom and non-centrality parameter λ is given by $k + \lambda$ and $2(k + 2\lambda)$ respectively. Thus we will have the following marginal expectation

$$\begin{aligned} E(Y|Z, \vartheta, \varphi) &= E(\mu_X|Z, \vartheta, \varphi) \\ &= \beta_2 \tau^2 (k + \lambda(Z)) + \left(\beta_0 - \frac{\beta_1^2}{4\beta_2} \right) \\ &= (\beta_0 + \beta_2 \tau^2) + (\beta_1 \gamma)Z + (\beta_2 \gamma^2)Z^2, \end{aligned}$$

after simplification. The marginal variance is also given by

$$\begin{aligned} Var(Y|Z, \vartheta, \varphi) &= E(\sigma^2|Z, \vartheta, \varphi) + Var(\mu_X|Z, \vartheta, \varphi) \\ &= \sigma^2 + 2\beta_2^2 \tau^4 (k + 2\lambda(Z)) \\ &= \sigma^2 + 2\beta_2^2 \tau^2 \left(\tau^2 + 2 \left(\gamma Z + \frac{\beta_1}{2\beta_2} \right)^2 \right). \end{aligned}$$



Appendix for Chapter 4

C.1 MCMC DIAGNOSTIC PLOTS

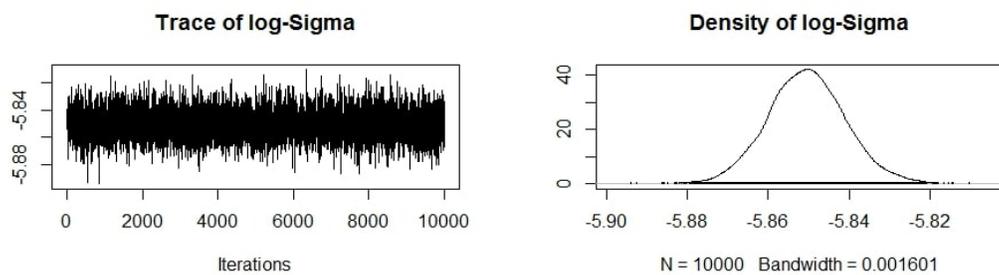


Figure C.1: MCMC trace plots based on (4.12).

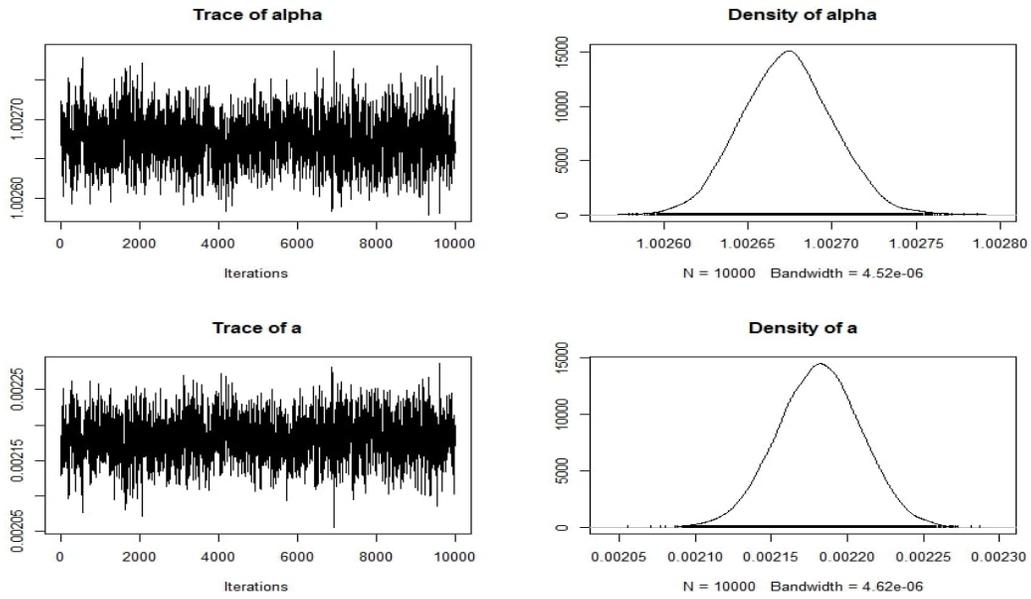


Figure C.2: MCMC trace plots based on (4.12).

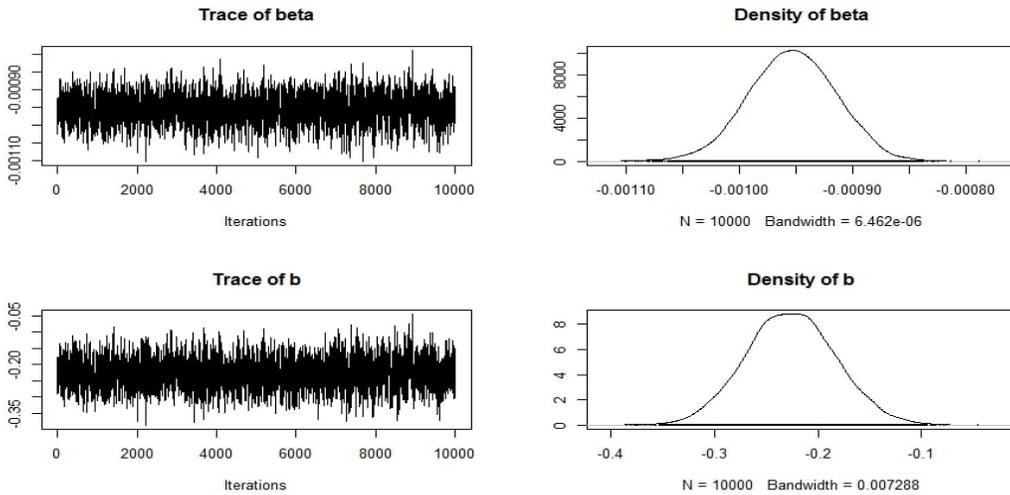


Figure C.3: MCMC trace plots based on (4.12).

References

- [1] Ba, S. (2015). *SLHD: Maximin-Distance (Sliced) Latin Hypercube Designs*. CRAN. R package version 2.1-1.
- [2] Bérubé, J. & Wu, C. F. J. W. (2000). Signal-to-noise ratio and related measures in parameter design optimization: An overview. *Sankhya - Series B*, 62, 417–432.
- [3] Bhaviripudi, S., Jia, X., Dresselhaus, M. S., & Kong, J. (2010). Role of kinetic factors in chemical vapor deposition synthesis of uniform large area graphene using copper catalyst. *Nano Letters*, 10(10), 4128–4133. PMID: 20812667.
- [4] Bolotin, K., Sikes, K., Hone, J., Stormer, H., & Kim, P. (2008). Temperature-dependent transport in suspended graphene. *Physical review letters*, 101(9), 96802.
- [5] Box, G. E. P. (1988). Signal-to-noise ratios, performance criteria, and transformations (with discussions). *Technometrics*, 30, 1–40.
- [6] Bugeđa, G., Cervera, M., Lombera, G., & Onate, E. (1995). Numerical analysis of stereolithography processes using the finite element method. *Rapid Prototyping Journal*, 1(2), 13–23.
- [7] Campbell, T., Williams, C., Ivanova, O., & Garrett, B. (2011). Could 3d printing change the world. *Technologies, Potential, and Implications of Additive Manufacturing*, Atlantic Council, Washington, DC.
- [8] Celebi, K., Cole, M. T., Choi, J. W., Wyczisk, F., Legagneux, P., Rupesinghe, N., Robertson, J., Teo, K. B. K., & Park, H. G. (2013). Evolutionary kinetics of graphene formation on copper. *Nano Letters*, 13(3), 967–974. PMID: 23339597.
- [9] Chaloner, K. & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, 10, 273–304.

- [10] Das, R., Ali, M. E., Hamid, S. B. A., Ramakrishna, S., & Chowdhury, Z. Z. (2014). Carbon nanotube membranes for water purification: A bright future in water desalination. *Desalination*, 336, 97–109.
- [11] Dasgupta, T., Ma, C., Joseph, V., Wang, Z., & Wu, C. (2008). Statistical modeling and analysis for robust synthesis of nanostructures. *Journal of the American Statistical Association*, 103, 594–603.
- [12] De Las Casas, C. & Li, W. (2012). A review of application of carbon nanotubes for lithium ion battery anode material. *Journal of Power Sources*, 208, 74–85.
- [13] Efron, B. & Morris, C. (1972). Empirical bayes on vector observations: An extension of stein's method. *Biometrika*, 59(2), 335–347.
- [14] Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press.
- [15] Gao, C., Guo, Z., Liu, J.-H., & Huang, X.-J. (2012). The new age of carbon nanotubes: An updated review of functionalized carbon nanotubes in electrochemical sensors. *Nanoscale*, 4, 1948.
- [16] Geim, A. K. & Novoselov, K. S. (2007). The rise of graphene. *Nature Materials*, 6(3), 183–191.
- [17] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian Data Analysis*. Chapman and Hall/CRC, 2 edition.
- [18] Gibson, I., Rosen, D. W., & Stucker, B. (2009). *Additive Manufacturing Technologies: Rapid Prototyping to Direct Digital Manufacturing*. Springer Publishing Company, Incorporated, 1st edition.
- [19] Hilton, P. D. & Jacobs, P. F. (2000). *Rapid Tooling: Technologies and Industrial Applications*. New York, NY, USA: Marcel Dekker, Inc.
- [20] Huang, Q., Nouri, H., Xu, K., Chen, Y., Sosina, S., & Dasgupta, T. (2014). Statistical predictive modeling and compensation of geometric deviations of three-dimensional printed products. *Journal of Manufacturing Science and Engineering*, 136(6), 061008–061008.
- [21] Huang, Q., Zhang, J., Sabbaghi, A., & Dasgupta, T. (2015). Optimal offline compensation of shape shrinkage for three-dimensional printing processes. *IIE Transactions*, 47(5), 431–441.

- [22] Kang, L. & Joseph, V. R. (2009). Bayesian optimal single arrays for robust parameter design. *Technometrics*, 51(3), 250–261.
- [23] Kiefer, J. & Wolfowitz, J. (1959). Optimum designs in regression problems. *Ann. Math. Statist.*, 30(2), 271–294.
- [24] Kim, H., Mattevi, C., Calvo, M. R., Oberg, J. C., Artiglia, L., Agnoli, S., Hirjibehedin, C. F., Chhowalla, M., & Saiz, E. (2012). Activation energy paths for graphene nucleation and growth on cu. *ACS Nano*, 6(4), 3614–3623.
- [25] Kim, H., Saiz, E., Chhowalla, M., & Mattevi, C. (2013). Modeling of the self-limited growth in catalytic chemical vapor deposition of graphene. *New Journal of Physics*, 15(5), 053012.
- [26] Kim, K., Zhao, Y., Jang, H., Lee, S., Kim, J., Kim, K., Ahn, J., Kim, P., Choi, J., & Hong, B. (2009). Large-scale pattern growth of graphene films for stretchable transparent electrodes. *Nature*, 457(7230), 706–710.
- [27] Lee, N. S., Chung, D. S., Han, I. T., Kang, J. H., Choi, Y. S., Kim, H. Y., Park, S. H., Jin, Y. W., Yi, W. K., Yun, M. J., Jung, J. E., Lee, C. J., You, J. H., Jo, S. H., Lee, C. G., & Kim, J. M. (2001). Application of carbon nanotubes to field emission displays. *Diamond and Related Materials*, 10, 265–270.
- [28] Leon, R. V., Shoemaker, A. C., & Kacker, R. N. (1987). Performance measures independent of adjustment: An explanation and extension of taguchi’s signal-to-noise ratios (with discussions). *Technometrics*, 29, 253–285.
- [29] Li, W. & Wu, C. F. J. W. (1999). An integrated method of parameter design and tolerance design. *Quality Engineering*, 11, 417–425.
- [30] Li, X., Cai, W., An, J., Kim, S., Nah, J., Yang, D., Piner, R., Velamakanni, A., Jung, I., Tutuc, E., Banerjee, S. K., Colombo, L., & Ruoff, R. S. (2009). Large-area synthesis of high-quality and uniform graphene films on copper foils. *Science*, 324, 1312–1314.
- [31] Liu, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. Springer Publishing Company, Incorporated.
- [32] Liu, P., Liu, L., Wei, Y., Liu, K., Chen, Z., Jiang, K., Li, Q., & Fan, S. (2009). Fast high-temperature response of carbon nanotube film and its application as an incandescent display. *Advanced Materials*, 21, 3563–3566.

- [33] Lynn-Charney, C. & Rosen, D. W. (2000). Usage of accuracy models in stereolithography process planning. *Rapid Prototyping Journal*, 6(2), 77–87.
- [34] Mardia, K. V. & Marshall, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1), 135–146.
- [35] Martin, A. D., Quinn, K. M., & Park, J. H. (2011). MCMCpack: Markov chain monte carlo in R. *Journal of Statistical Software*, 42(9), 1–21.
- [36] McKay, M. D. (1992). Latin hypercube sampling as a tool in uncertainty analysis of computer models. In *Proceedings of the 24th Conference on Winter Simulation, WSC '92* (pp. 557–564). New York, NY, USA: ACM.
- [37] McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2), pp. 239–245.
- [38] Morris, C. N. (1983). Parametric empirical bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(381), 47–55.
- [39] Nair, V. N. (1992). Taguchi's parameter design: A panel discussion. *Technometrics*, 34, 127–161.
- [40] Park, S., Vosguerichian, M., & Bao, Z. (2013). A review of fabrication and applications of carbon nanotube film-based flexible electronics. *Nanoscale*, 5, 1727–52.
- [41] Phadke, M. S. (1989). *Quality Engineering Using Robust Design*. Englewood Cliffs, NJ: Prentice-Hall.
- [42] Pukelsheim, F. (2006). *Optimal Design of Experiments*. New York: Wiley.
- [43] R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [44] Rasmussen, C. E. & Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- [45] Remillard, E. M., Zhang, Q., Sosina, S., Branson, Z., Dasgupta, T., & Vecitis, C. D. (2016). Electric-field alignment of aqueous multi-walled carbon nanotubes on microporous substrates. *Carbon*, 100, 578 – 589.

- [46] Ribeiro, P. J. J. & Diggle, P. J. (2015). *geoR: Analysis of Geostatistical Data*. CRAN. R package version 1.7-5.1.
- [47] Roustant, O., Ginsbourger, D., & Deville, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1), 1–55.
- [48] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.
- [49] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji - an open source platform for biological image analysis. *Nat Methods*, 9(7), 676–682. 22743772[pmid].
- [50] Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). Nih image to imagej: 25 years of image analysis. *Nat Meth*, 9(7), 671–675.
- [51] Shafiei, S., Nourbakhsh, A., Ganjipour, B., Zahedifar, M., & Vakili-Nezhaad, G. (2007). Diameter optimization of VLS-synthesized ZnO nanowires, using statistical design of experiment. *Nanotechnology*, 18, 355708.
- [52] Stankovich, S., Dikin, D., Dommett, G., Kohlhaas, K., Zimney, E., Stach, E., Piner, R., Nguyen, S., & Ruoff, R. (2006). Graphene-based composite materials. *Nature*, 442(7100), 282–286.
- [53] Stoller, M. D., Park, S., Zhu, Y., An, J., & Ruoff, R. S. (2008). Graphene-based ultracapacitors. *Nano letters*, 8(10), 3498–3502.
- [54] Taguchi, G. (1987). *System of Experimental Design: Engineering methods to optimize quality and minimize costs*. White Plains, NY: UNIPUB, Kraus International Publications.
- [55] Tang, Y., Henderson, C., Muzzy, J., & Rosen, D. W. (2004). Stereolithography cure modelling and simulation. *International Journal of Materials and Product Technology*, 21(4), 255–272.
- [56] Tong, K., Joshi, S., & Lehtihet, E. A. (2008). Error compensation for fused deposition modeling (fdm) machine by correcting slice files. *Rapid Prototyping Journal*, 14(1), 4–14.
- [57] Tong, K., Lehtihet, E. A., & Joshi, S. (2003). Parametric error modeling and software error compensation for rapid prototyping. *Rapid Prototyping Journal*, 9(5), 301–313.

- [58] Vecitis, C. D., Gao, G., & Liu, H. (2011). Electrochemical carbon nanotube filter for adsorption, desorption, and oxidation of aqueous dyes and anions. *Journal of Physical Chemistry C*, 115, 3621–3629.
- [59] Vlasiouk, I., Regmi, M., Fulvio, P., Dai, S., Datskos, P., Eres, G., & Smirnov, S. (2011). Role of hydrogen in chemical vapor deposition growth of large single-crystal graphene. *ACS Nano*, 5(7), 6069–6076. PMID: 21707037.
- [60] Vlasiouk, I., Smirnov, S., Regmi, M., Surwade, S. P., Srivastava, N., Feenstra, R., Eres, G., Parish, C., Lavrik, N., Datskos, P., Dai, S., & Fulvio, P. (2013). Graphene nucleation density on copper: Fundamental role of background pressure. *The Journal of Physical Chemistry C*, 117(37), 18919–18926.
- [61] Wang, W., Cheah, C., Fuh, J., & Lu, L. (1996). Influence of process parameters on stereolithography part shrinkage. *Materials & Design*, 17(4), 205 – 213.
- [62] Wheeler, B. (2014). *AlgDesign: Algorithmic Experimental Design*. R package version 1.1-7.3.
- [63] Wu, C. F. J. & Hamada, M. (2009). *Experiments: Planning, Analysis and Optimization*. New York: Wiley.
- [64] Wu, J. & Huang, Q. (2014). Graphene growth process modeling: a physical-statistical approach. *Applied Physics A, Materials Science & Processing*, 116(4), 1747 – 1756.
- [65] Xing, S., Wu, W., Wang, Y., Bao, J., & Pei, S.-S. (2013). Kinetic study of graphene growth: Temperature perspective on growth rate and film thickness by chemical vapor deposition. *Chemical Physics Letters*, 580, 62 – 66.
- [66] Xu, L., Huang, Q., Sabbaghi, A., & Dasgupta, T. (2013). Shape deviation modeling for dimensional quality control in additive manufacturing. In *ASME 2013 International Mechanical Engineering Congress and Exposition* (pp. V02AT02A018–V02AT02A018).: American Society of Mechanical Engineers.
- [67] Zhou, J. G., Herscovici, D., & Chen, C. C. (2000). Parametric process optimization to improve the accuracy of rapid prototyped stereolithography parts. *International Journal of Machine Tools and Manufacture*, 40(3), 363 – 379.
- [68] Zhu, L., Dasgupta, T., & Huang, Q. (2014). A D-optimal design for estimation of parameters of an exponential-linear growth curve of nanostructures. *Technometrics*, 56, 432–442.