Response surface optimization in the presence of internal noise with application to optimal alignment of Carbon Nanotubes

Sobambo Sosina[*], E. Marielle Remillard[**], Qiaoying Zhang[**], Chad Vecitis[**] and Tirthankar Dasgupta[***]

[*]Department of Statistics, Harvard University

[**]John A. Paulson School of Engineering and Applied Sciences, Harvard University

[***]Department of Statistics and Biostatistics, Rutgers University

Abstract

Internal noise, which means fluctuation of input factors around their set values, is common in many experiments in the physical and engineering sciences. Existing methods for response surface optimization in the presence of internal noise typically adopt a two-step approach: (a) fitting a response model as a function of the set value and (b) using Monte-Carlo methods to account for internal noise while optimizing the response. In this article, motivated by a problem in optimizing alignment of carbon nanotubes (CNT), we propose a Bayesian approach for response surface optimization in the presence of internal noise. A unit-free and interpretable measure to quantify the strength of internal noise is proposed. Suitable objective functions or performance measures consistent with the overall goal of optimizing the response function are identified, methods for estimating them from available experimental data are suggested, and simulations are conducted to compare them with respect to their ability to account for internal noise in the optimization problem. The loss accrued by ignoring the internal noise in the optimization problem is quantified and studied via simulation. The proposed method is demonstrated through its application in the CNT alignment problem.

1 Introduction

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 inputs. A well-studied field of statistics, and experimental design in particular, is robust parameter design (Taguchi, 1987; Wu and Hamada, 2009), 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 applications such as the automobile industry. 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 (Phadke, 1989; Li and Wu, 1999). Also, Hamada and Steiner (1997) and Hamada et al. (2005) proposed analysis of mixture experiments with mixing errors as internal noise. In engineering literature, a class of problems that share some common features with the robust design problem under internal noise have been studied under the name of "probabilistic design" (Du and Chen, 2004; Yin and Chen, 2006). These problems focus on developing approaches to optimize a given objective function of an input random variable with respect to its set-value. In contrast, the primary goals of a robust design formulation are (a) to identify an appropriate objective function as a function of the set-value, optimizing which would make the process output less sensitive to effect of internal noise, and (b) to estimate such a function from the available experimental data with evaluation of the uncertainty involved in such estimation.

The fact that experiments in nanotechnology involve input factors that are prone to internal noise was noted and discussed by Dasgupta et al. (2008), 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

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon \ , \tag{1}$$

where X is a single input variable, referred to as "control factor" 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 is subject to internal noise, which means X is itself a random function of a true input Z. Examples of X and Z are the actual temperature and set

temperature respectively. Then we can write, for example,

$$X = g_\gamma(Z) + \upsilon \ , \tag{2}$$

where $g_\gamma(\cdot)$ is a deterministic function parametrized by γ, and υ denotes the additive internal noise. Then we can rewrite (1) as

$$Y = \beta_0 + \beta_1 g_\gamma(Z) + \beta_2 g_\gamma^2(Z) + \beta_1 \upsilon + 2\beta_2 \upsilon g_\gamma(Z) + \beta_2 \upsilon^2 + \epsilon \ . \tag{3}$$

The optimization problem therefore boils down to finding a Z^* that maximizes $E(Y)$ while keeping $var(Y)$ 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 Taguchi (1987) or other measures by Leon et al. (1987).

Data from experiments involving internal noise are typically of the form $(Y_i, Z_i)_{i=1,\ldots,n}$, permitting experimenters to fit models like (3) that connect Y directly to Z and do not permit estimation of (2). Information on the internal noise model as (2) are often independently obtained from surrogate experiments or observational studies that yield pairwise data $(X_j, Z_j)_{j=1,\ldots,m}$. The approach adopted by Dasgupta et al. (2008) consisted of the following two steps: (i) fitting a regression model of Y on Z using the data on (Y, Z), and (ii) subsequently incorporating the surrogate information about (2) obtained from the data on (X, Z) for optimization of (3) using Monte-Carlo simulations. Clearly this approach suffers from the problem that the fitted model connecting Y and Z is mis-specified and not a fitted version of the true model (3). Also, the response surface modeling and its optimization are treated as two separate problems. 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 that typically involves 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. (2016). Data on Z_i was ignored in the analysis, because it was felt that the internal noise was not very strong. Optimization was done with respect to X. 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, motivated by a problem in optimizing alignment of Carbon nanotubes (CNT), we propose

a 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 hierarchical model that captures the mutual relationship among Y, X and Z, (c) Identifying suitable objective functions or performance measures consistent with the overall goal if 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.

We first describe a motivating example in Section 2 from a CNT alignment experiment. In Section 3, 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 4 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 5, we perform extensive simulations to compare the proposed objective functions. Section 6 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 7, and some concluding remarks are presented in Section 8.

2 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 electronics (Park et al., 2013), energy storage devices (De Las Casas and Li, 2012), optical displays (Lee et al., 2001; Liu et al., 2009), chemical sensors (Gao et al., 2012), and water treatment technologies (Das et al., 2014; Vecitis et al., 2011). 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. (2016) utilize an external electric field, one of the most promising techniques for industrial-scale alignment, to orient CNTs, and their work provides a basis for investigating the role of internal noise.

In their paper, Remillard et al. (2016) conduct an experiment to determine the optimal level of electric field strength that maximizes alignment of CNTs. Figure 1 shows an SEM image of multi-walled carbon

nanotubes aligned using a combination of these factors compared to a randomly oriented sample (Remillard et al., 2016).

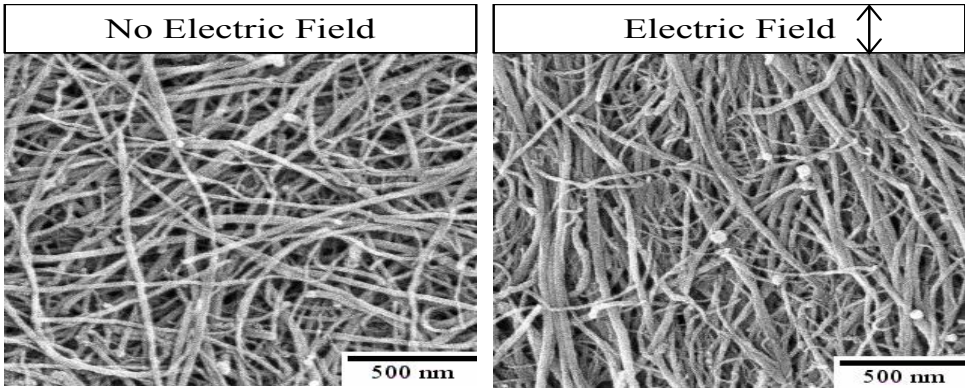


Figure 1: Examples of randomly oriented CNT and CNT aligned in the presence on an electric field.

Experiments were also conducted by simultaneously varying two other factors in addition to electrical field strength (X): CNT mass (M), and CNT concentration in suspension fluid (C). Initial exploration of the factor space led to determination of a response model of the form

$$R = \psi\left(X, M, C\right) + \epsilon \ , \tag{4}$$

where R is a response that measures alignment, $\psi(\cdot)$ is a function of the input variables, and $\epsilon \sim \mathcal{N}(0, \sigma^2)$ denotes the observational noise. The expected response function was optimized to determine the desired levels of M, C and X. However, an important aspect that was not considered in this rudimentary optimization approach was the fact that X cannot be controlled precisely at a desired value: the electrical field X is generated by applying a fixed input voltage Z to the CNT samples suspended in fluid. For a fixed value of Z, different values of X are possible due to fluctuations arising from the power supply machine. Thus X, as an input to the alignment process, is inherently noisy. Figure 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 the distribution of X possesses a mean-dependent standard deviation, leading to a heteroscedastic and asymmetric marginal distribution for R, although neither the magnitude of the noise nor the level of heteroscedasticity appears to be threatening. However, the experimenters were

curious to know (a) whether such noise, however small its magnitude may be, is capable of affecting the optimal solution, and (b) whether there exists a way to quantify the severity of the noise. We will first provide an answer to (b) in Section 3. It is important to note that, a frequentist investigation of questions (a) and (b) under an additive iid internal noise model $X = Z + \epsilon$ was conducted by Ardakani et al. (2011). However, it is clear that such a model for internal noise is not appropriate for the scenario described by Figure 2. We consider a generic model for internal noise that includes the homoscedastic additive noise considered by Ardakani et al. (2011) and the heteroscedastic noise implied by Figure 2 as special cases. Our framework also incorporates the uncertainty associated with the sampling of levels of Z contrary to a fixed design assumed by Ardakani et al. (2011).

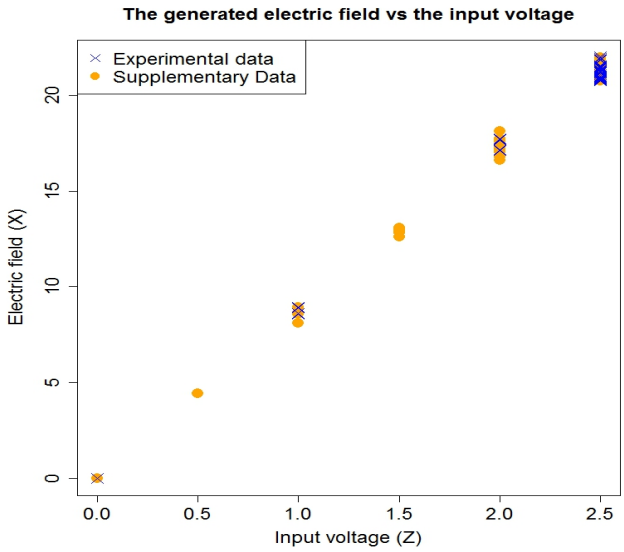


Figure 2: The original (generated) voltage (in volts) versus the amplified voltage (in volts).

3 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. An obvious choice is the dispersion of X under the assumption of zero-internal-noise, relative to the true dispersion of X. The dispersion of X under the assumption of zero-internal-noise is directly related to the dispersion of Z, which is itself governed by the experimental design or 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 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,

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

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 is a parameter (scalar or vector) associated with $E(X|Z)$ in model (2), and τ is a variance parameter, typically associated with the variance of the internal noise ν in (2), that measures the degree of internal noise. Thus, assuming fixed values of γ and τ,

$$\sigma_f^2(X, \tau, \gamma) = E_f var(X|Z, \tau, \gamma) + var_f E(X|Z, \tau, \gamma) \ . \tag{5}$$

The ratio of the true variance of X explained by the zero-internal-noise model is $\sigma_f^2(X, 0, \gamma)/\sigma_f^2(X, \tau, \gamma)$, where $\sigma_f^2(X, \tau, \gamma)$ is given by (5). Thus a reasonable unit-free measure of internal noise, conditional on γ and τ, is the *ratio of unexplained variance* under a zero-internal-noise assumption, given by:

$$S_f(\gamma, \tau) = 1 - \frac{\sigma_f^2(X, 0, \gamma)}{\sigma_f^2(X, \tau, \gamma)} \ , \tag{6}$$

Note that $S_f(\gamma, \tau)$, as defined by (6), 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), $g_\gamma(z) = \gamma Z$ and $var(\nu) = \tau^2$ for specific values of γ and τ, and let the pdf associated with the experimental design be $f_2(z) = \frac{1}{b-a}\mathbb{I}\left(a \leq z \leq b\right)$ as in the uniform case introduced earlier. Then, from (5),

we have

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

Substitution of the above in (6) yields:

$$S_f(\gamma,\tau) = \frac{12\tau^2}{\gamma^2(b-a)^2 + 12\tau^2} \ . \tag{7}$$

In the case of heteroscedastic internal noise where we have a mean-dependent internal standard deviation, $Var(X|Z,\tau,\gamma) = \tau^2 Z^2$, we have that $E(Var(X|Z,\tau,\gamma)) = (\tau^2/3)\cdot(a^2+b^2+ab)$, and thus

$$S_f(\gamma,\tau) = \frac{4\tau^2(a^2+b^2+ab)}{\gamma^2(b-a)^2 + 4\tau^2(a^2+b^2+ab)} \ . \tag{8}$$

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 5, and our CNT alignment model in Section 7. It may be noted that the proposed measure $S_f(\gamma,\tau)$ may be considered a generalization of the heterogeneity ratio (or its function) proposed by Vuchkov and Boyadjieva (2001) and used by Ardakani et al. (2011) under an additive homoscedastic model $X = Z + \epsilon$ and a fixed design.

We utilize a Bayesian framework to draw inference about $S_f(\gamma,\tau)$. In this framework, we postulate prior distributions for γ and τ, that allow the experimenter to incorporate any available prior knowledge about the internal noise into the analysis. Consequently, $S_f(\gamma,\tau)$ is a function of the random variables γ and τ. Because in most situations closed form expressions of the posterior distributions of γ and τ cannot be obtained, the Bayesian framework will be implemented via Markov chain Monte-Carlo sampling. As a result, posterior draws of γ and τ will be used to generate the posterior distribution of $S_f(\gamma,\tau)$.

In the remaining part of this paper, we simplify the notation for the internal noise measure by dropping the suffix f and denoting it by $S(\gamma,\tau)$ henceforth.

4 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

$$\begin{aligned} Y|X,\boldsymbol{\theta} &\sim \mathcal{F}_{(\boldsymbol{\theta},X)} \equiv \mathcal{F}(Y|X,\boldsymbol{\theta}) \\ X|Z,\boldsymbol{\phi} &\sim \mathcal{G}_{(\boldsymbol{\phi},Z)} \equiv \mathcal{G}(X|Z,\boldsymbol{\phi}) \\ (\boldsymbol{\theta},\boldsymbol{\phi}) &\sim \pi(\boldsymbol{\theta},\boldsymbol{\phi}) \ , \end{aligned} \tag{9}$$

where the distributions $\mathcal{F}(\cdot)$ and $\mathcal{G}(\cdot)$ are parametrized by $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ 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 $\mathcal{F}(Y|X,\boldsymbol{\theta})$ and $\mathcal{G}(X|Z,\boldsymbol{\phi})$ as density functions later on. We restrict the distribution $\mathcal{G}$ such that $E(X|Z,\boldsymbol{\phi})$, denoted by $g_{\boldsymbol{\phi}}(Z)$ is an invertible function of Z.

Without loss of generality, let us assume henceforth that the main objective is the maximization of the response. 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

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

where $(\boldsymbol{Y}^{\text{obs}},\boldsymbol{X}^{\text{obs}})$ 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 $E(Y|X,\boldsymbol{Y}^{\text{obs}},\boldsymbol{X}^{\text{obs}})$ refers to the mean of the posterior predictive distribution (see Gelman et al. (2003)) of Y for a given value of X, observed responses $\boldsymbol{Y}^{\text{obs}}$ and corresponding observed inputs $\boldsymbol{X}^{\text{obs}}$, and is given by:

$$E(Y|X,\boldsymbol{Y}^{\text{obs}},\boldsymbol{X}^{\text{obs}}) = \int E(Y|X,\boldsymbol{\theta}) \ \pi(\boldsymbol{\theta}|\boldsymbol{Y}^{\text{obs}},\boldsymbol{X}^{\text{obs}}) \ d\boldsymbol{\theta}, \tag{10}$$

where $\pi(\boldsymbol{\theta}|\boldsymbol{Y}^{\text{obs}},\boldsymbol{X}^{\text{obs}})$ denotes the posterior distribution of $\boldsymbol{\theta}$. 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.

If in addition to X, there are other non-noisy experimental factors $\boldsymbol{W}$ as in the CNT example, then the first line of (9) will change to $Y|X,\boldsymbol{W},\boldsymbol{\theta} \sim \mathcal{F}_{(\boldsymbol{\theta},\boldsymbol{W},X)} \equiv \mathcal{F}(Y|X,\boldsymbol{W},\boldsymbol{\theta})$, and the optimization problem will

become

$$(X^*, \boldsymbol{W}^*) = \arg\max_{X,\boldsymbol{W}} E(Y|X, \boldsymbol{W}, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{X}^{\text{obs}}, \boldsymbol{W}^{\text{obs}}). \tag{11}$$

Such a set-up will be adopted and demonstrated in Section 7. In this paper, we assume that the variables X and $\boldsymbol{W}$ are already identified through some model selection procedure.

4.1 Objective Functions

One common objective function that can be borrowed from robust design literature (Wu and Hamada, 2009) is the signal-to-noise (S/N) ratio, the inverse of the coefficient of variation. Applying this to (9), we have the objective function

$$\eta_1(Z) = \frac{E(Y|Z, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{X}^{\text{obs}}, \boldsymbol{Z}^{\text{obs}})}{\sqrt{Var(Y|Z, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{X}^{\text{obs}}, \boldsymbol{Z}^{\text{obs}})}}. \tag{12}$$

Here, $(\boldsymbol{Y}^{\text{obs}}, \boldsymbol{X}^{\text{obs}}, \boldsymbol{Z}^{\text{obs}})$ 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 in Remillard et al. (2016) or as two separate pairs (Y, Z) and (X, Z) from two different experiments/sources, as in Dasgupta et al. (2008). The numerator and denominator of (12) 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:

$$\eta_1(Z) = \frac{\int E(Y|Z, \boldsymbol{\theta}, \boldsymbol{\phi})\ \pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathcal{D})\ d\boldsymbol{\theta} d\boldsymbol{\phi}}{\sqrt{\int Var(Y|Z, \boldsymbol{\theta}, \boldsymbol{\phi})\ \pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathcal{D})\ d\boldsymbol{\theta} d\boldsymbol{\phi}}}, \tag{13}$$

where $\pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathcal{D})$ is the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\phi})$ given the data.

Note that this objective function is only appropriate when the expected posterior predictive response, $E(Y|Z, \boldsymbol{Y}^{\text{obs}}, \boldsymbol{X}^{\text{obs}}, \boldsymbol{Z}^{\text{obs}})$, 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. For example, when the conditional distributions in (9) are normal, it has a closed-form expression. However, several researchers have investigated the role of η_1, defined in (12), as an objective function in robust design and have found it to be appropriate only under specific strong assumptions. See, for example, Leon et al. (1987), Box (1988), Nair (1992), Bérubé and Wu (2000). Our preliminary investigation, discussed later, shows that it can be highly sensitive to internal noise, especially when the level of such noise is moderate to high. 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

$$\eta_2(Z) = Pr(Y > y^*|Z, \mathcal{D}) = \int Pr(Y > y^*|Z, \boldsymbol{\theta}, \boldsymbol{\phi})\ \pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathcal{D})\ d\boldsymbol{\theta} d\boldsymbol{\phi}. \tag{14}$$

The criterion η_2 is the marginal probability of the response exceeding some predefined threshold, y^*. Similar probability-based criteria were considered by Wu and Hamada (2009). 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

$$\eta_3(Z) = E(I_{\{Y>y^*\}}Y|Z, \mathcal{D}) = \int E(I_{\{Y>y^*\}}Y|Z, \boldsymbol{\theta}, \boldsymbol{\phi})\ \pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathcal{D})\ d\boldsymbol{\theta} d\boldsymbol{\phi}. \tag{15}$$

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

$$y^* = \max_{Z \in \Omega}(E(Y|Z, \mathcal{D}))\ . \tag{16}$$

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.

4.2 Computation of Objective Functions

The computation of the three objective functions given by (13), (14) and (15) involve computation of terms of the following form:

$$\int E\left[H(Y)|Z, \boldsymbol{\theta}, \boldsymbol{\phi}\right]\ \pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathcal{D})\ d\boldsymbol{\theta} d\boldsymbol{\phi}, \tag{17}$$

where,

$$H(Y) = \begin{cases} Y & \text{corresponding to the numerator of (13)} \\ I_{\{Y>y^*\}} & \text{corresponding to (14)} \\ I_{\{Y>y^*\}}Y & \text{corresponding to (15)} \end{cases}$$

The computational challenges in (17) arise from:

(i) The need to "integrate out" the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, that is, compute the average of $E\left[H(Y)|Z, \boldsymbol{\theta}, \boldsymbol{\phi}\right]$ with respect to the posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathcal{D})$.

(ii) The evaluation of $E\left[H(Y)|Z,\boldsymbol{\theta},\boldsymbol{\phi}\right]$.

To address (i), we by-pass the integration by multiplying the posterior density, $\pi(\boldsymbol{\theta},\boldsymbol{\phi}|\mathcal{D})$, with zero everywhere, except at the mode. This is equivalent to plugging in the posterior modes, denoted by $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$, and reduces (17) to $E\left(H(Y)|Z,\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\phi}}\right)$. The effect of such approximations is studied through carefully designed simulation studies in Section 5.

To address (ii), first note that evaluation of the marginal expectation, $E\left[H(Y)|Z,\boldsymbol{\theta},\boldsymbol{\phi}\right]$ 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 A as I_A, and obtain

$$\begin{aligned} E\left(H(Y)|Z,\boldsymbol{\theta},\boldsymbol{\phi}\right) &= E\left[E\left[H(Y)|X,Z,\boldsymbol{\theta},\boldsymbol{\phi}\right]|Z,\boldsymbol{\phi}\right] \\ &= \int_{-\infty}^{\infty} E\left[H(Y)|X,\boldsymbol{\theta}\right]\cdot\mathcal{G}(x|Z,\boldsymbol{\phi})\;dx\;. \end{aligned} \tag{18}$$

In general, the integral in (18) 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 (18), that $E\left[H(Y)|X,\boldsymbol{\theta}\right]$, is the same for all Z, while the density function, $\mathcal{G}(X|Z,\boldsymbol{\phi})$, changes with respect to Z. Thus the naive approach of first sampling a predetermined number of input points, say N, from $\mathcal{G}(X|Z,\boldsymbol{\phi})$, and then calculating the approximate quantity

$$E\left(H(Y)|Z,\boldsymbol{\theta},\boldsymbol{\phi}\right)\approx\frac{1}{N}\sum_{i=1}^{N}E\left(H(Y)|X=x_i,\boldsymbol{\theta}\right)\;,$$

every time Z changes, is computationally inefficient due to the unavoidably large number of times $\mathcal{G}(X|Z,\boldsymbol{\phi})$ will need to be sampled from. A more efficient approach, inspired by importance sampling ideas (Liu, 2008), involves a batch-sequential algorithm. This means that the X's are obtained in a single batch for all possible Z sequences, rather than obtaining new Xs for each member of a Z sequence.

1. Sample a single representative set of N X's, denoted by $\boldsymbol{\mathcal{X}}_N$, from some distribution independent of Z, denoted by $\mathcal{H}(X|\boldsymbol{\phi})$.

2. For each candidate optimal point, Z, calculate the approximate quantity

$$E\left(H(Y)|Z,\boldsymbol{\theta},\boldsymbol{\phi}\right) \approx \frac{\sum\limits_{x\in\boldsymbol{\mathcal{X}}_N} E\left(H(Y)|X=x,\boldsymbol{\theta}\right)\cdot\frac{\mathcal{G}(x|Z,\boldsymbol{\phi})}{\mathcal{H}(x|\boldsymbol{\phi})}}{\sum\limits_{x\in\boldsymbol{\mathcal{X}}_N}\frac{\mathcal{G}(x|Z,\boldsymbol{\phi})}{\mathcal{H}(x|\boldsymbol{\phi})}} \ . \tag{19}$$

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

$$\mathcal{H}(X|\boldsymbol{\phi}) = \int\limits_{\Omega_Z} \mathcal{G}(X|Z=z,\boldsymbol{\phi})\cdot f(z)\ dz \ , \tag{20}$$

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 $\mathcal{H}(X|\boldsymbol{\phi})$ in (20) is not always available and therefore a further approximation is needed. Denoting a set of N Z's sampled from $f(z)$ by $\boldsymbol{\mathcal{Z}}_N$, we can write

$$\mathcal{H}(X|\boldsymbol{\phi}) \approx \frac{1}{N}\sum_{z\in\boldsymbol{\mathcal{Z}}_N}\mathcal{G}(X|Z=z,\boldsymbol{\phi}) \ . \tag{21}$$

This means any member of $\boldsymbol{\mathcal{X}}_N$ can be generated by first sampling Z from Ω_Z using $f(z)$, and then sampling a single X from $\mathcal{G}(X|Z,\boldsymbol{\phi})$. 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 set Ω_Z are equally permissible in the design, then we can set $f(z)$ to be the uniform density over Ω_Z, that is,

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

Since the paired representative set $\{\boldsymbol{\mathcal{Z}}_N, \boldsymbol{\mathcal{X}}_N\}$ 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 $N \to \infty$, the difference due to approximations go to zero.

5 Simulation Studies for Assessing Sensitivity of Objective Function to Noise

We now assess the performance of each objective function under different types and magnitudes of internal noise, using simulations. We consider three scenarios or settings for the data generating mechanism: (i)

zero internal noise, (ii) homoscedastic internal noise, and (iii) a special case of heteroscedastic internal noise, using a mean dependent standard deviation.

We consider the following data generating model:

$$\begin{aligned} &Y|(X,\beta_0,\beta_1,\beta_2,\sigma) \sim \mathcal{N}(\mu(X),\ \sigma^2) \ \text{ where } \mu(X) = \beta_0 + \beta_1 X + \beta_2 X^2, \\ &X|(Z,\gamma,\tau) \sim \mathcal{N}(\gamma Z,\ V(\tau,Z)). \end{aligned} \tag{22}$$

Here we assume $\gamma > 0$, and $\beta_2 < 0$ (so that the optimization corresponds to maximization). The variance function $V(\tau, Z)$ equals τ^2 for the homoscedastic noise setting, and $\tau^2 Z^2$ for the heteroscedastic noise setting.

In our data generation mechanism, we consider fixed values of the parameters $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \sigma)$ and $\boldsymbol{\phi} = (\gamma, \tau)$. The design region Ω_Z is set to the interval $(0, 0.1)$. With each setting of the parameters, 100 triplets of (Z, X, Y), denoted by $\{\boldsymbol{Z}^{(1)}, \boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)}\}$ are generated. In addition, 50 tuples of (Z, X), denoted by $\{\boldsymbol{Z}^{(2)}, \boldsymbol{X}^{(2)}\}$, are generated to represent supplementary data.

Table 1: Parameter values under different simulation settings

Simulation	$\boldsymbol{\theta}$				$\boldsymbol{\phi}$		Variance function	Unit-free measure
Setting	β_0	β_1	β_2	σ	γ	τ	$V(\tau, Z)$	$S(\gamma, \tau)$
Zero internal noise	200	1	-1	0.05	10	0	0	0
Homoscedastic internal noise	200	1	-1	0.05	10	2.87	2.87^2	0.99
Heteroscedastic internal noise	200	1	-1	0.05	10	5	$25Z^2$	0.50

The parameter values in the three simulation settings are shown in Table 1. Note that the three simulation scenarios have the same parameter settings except τ, which determines the magnitude of noise. The last column of Table 1 shows the true values of the unit-free internal noise measure $S(\gamma, \tau)$, calculated by substituting values of γ and τ in (7) for homoscedastic noise and (8) for heteroscedastic noise. Thus our simulation settings consider high homoscedastic noise and moderate heteroscedastic noise.

Because the fixed parameter values can be visualized as point mass priors, the true objectives functions $\eta_1(Z)$, $\eta_2(Z)$ and $\eta_3(Z)$ can be obtained by removing the integrals in (13), (14) and (15) respectively. Although the objective functions η_2 and η_3 do not have closed-form expressions, under model (22), objective function η_1 has a closed form expression (derivation in Appendix A) given by

$$\eta_1(Z) = \frac{\beta_0 + \beta_2 V(\tau, Z) + (\beta_1\gamma)Z + (\beta_2\gamma^2)Z^2}{\sqrt{\sigma^2 + 2\beta_2^2 V(\tau, Z)\left\{V(\tau, Z) + 2\left(\gamma Z + \frac{\beta_1}{2\beta_2}\right)^2\right\}}}\ . \tag{23}$$

Knowing the reality, we now proceed to applying the proposed Bayesian procedure. In order to calculate

the objective functions, we assume the joint weakly-informative prior for the transformed parameters $\boldsymbol{\varpi} = (\beta_0, \beta_1, \beta_2, \log(\sigma), \log(\gamma), \log(\tau))$,

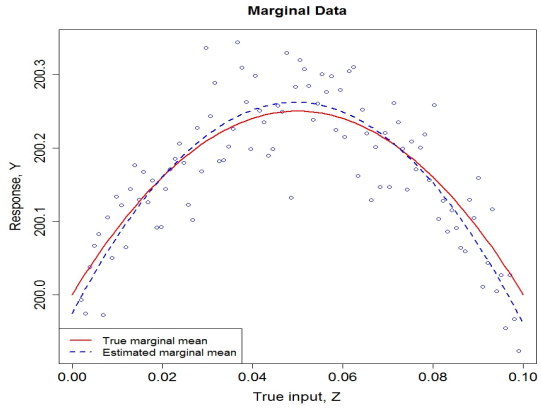
$$\boldsymbol{\varpi} \sim \pi(\boldsymbol{\varpi}) \equiv \mathcal{N}(\mathbf{0},\ 10^{10}\boldsymbol{I}) \ , \tag{24}$$

where $\boldsymbol{I}$ here represents the identity matrix. This is to ensure that the posterior is proper, while being only weakly influenced by the prior. Denoting $\{\boldsymbol{Z}^{(1)}, \boldsymbol{X}^{(1)}, \boldsymbol{Y}^{(1)}, \boldsymbol{X}^{(2)}, \boldsymbol{Y}^{(2)}\}$ by $\mathcal{D}$, the posterior (up to the normalizing constant) is obtained from (22) and (24) as,

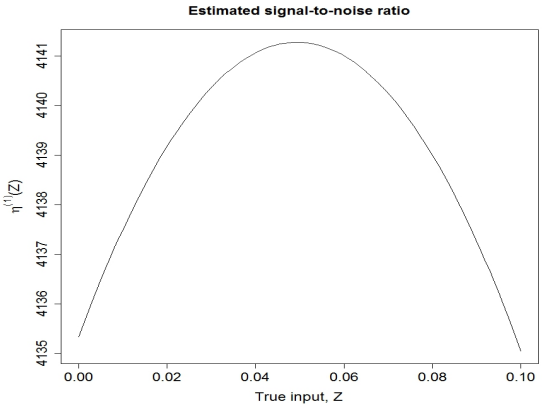
$$\pi\left(\boldsymbol{\varpi}|\mathcal{D}\right) \propto \mathcal{N}(\ \mu(\boldsymbol{X}^{(1)}),\ \sigma^2\boldsymbol{I}) \cdot \mathcal{N}(\gamma\boldsymbol{Z}^{(1)},\ V(\tau, Z)\boldsymbol{I}) \cdot \mathcal{N}(\gamma\boldsymbol{Z}^{(2)},\ V(\tau, Z)\boldsymbol{I}) \cdot \pi(\boldsymbol{\varpi}) \ . \tag{25}$$

Maximizing $\pi\left(\boldsymbol{\varpi}|\mathcal{D}\right)$ from (25) and reversing the parameter transformation yields the posterior modes $\hat{\boldsymbol{\theta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma})$ and $\hat{\boldsymbol{\phi}} = (\hat{\gamma}, \hat{\tau})$, which may now be used to evaluate the objective functions η_1, η_2 and η_3.

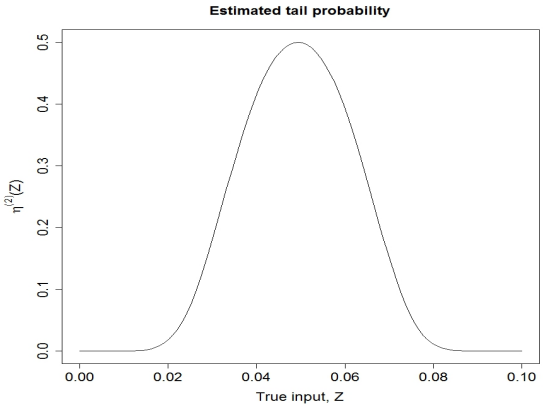
The generated data, true model, estimated fit as well as the three objective functions are shown under zero-internal noise in Figure 3, under high homoscedastic internal noise in Figure 4, and under moderate heteroscedastic internal noise in Figure 5



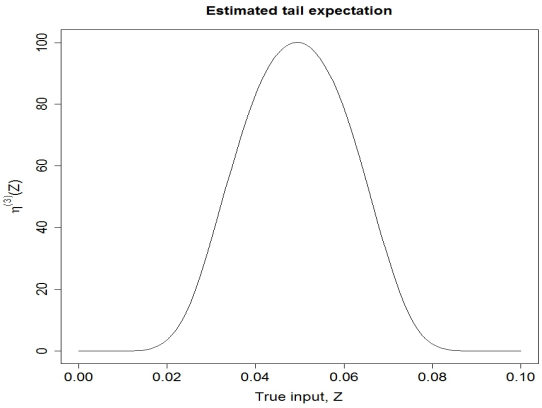
(a) Y vs Z for the simulated data points, true relationship, and estimated relationship.



(b) $\eta_1(Z)$ vs Z for simulated data.



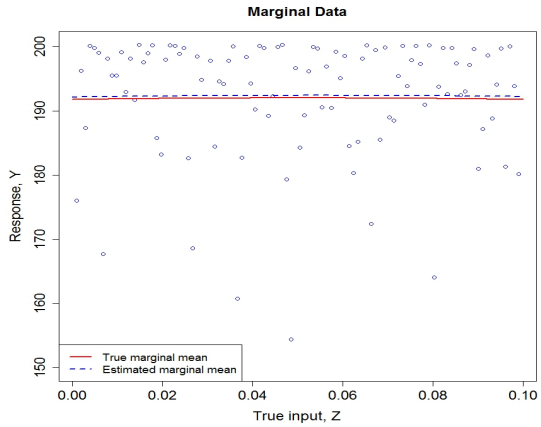
(c) $\eta_2(Z)$ vs Z for simulated data.



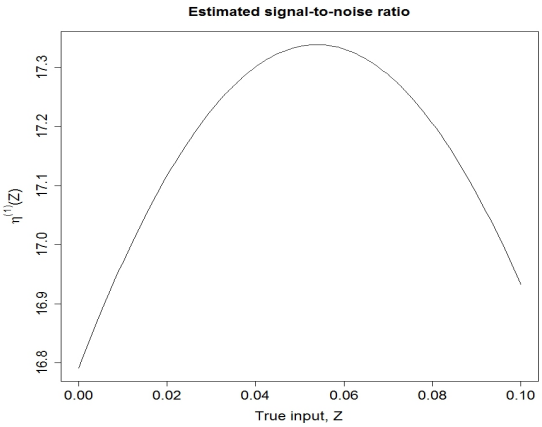
(d) $\eta_3(Z)$ vs Z for simulated data.

Figure 3: Zero internal noise ($\tau = 0$ and $S(\gamma, \tau) = 0$).

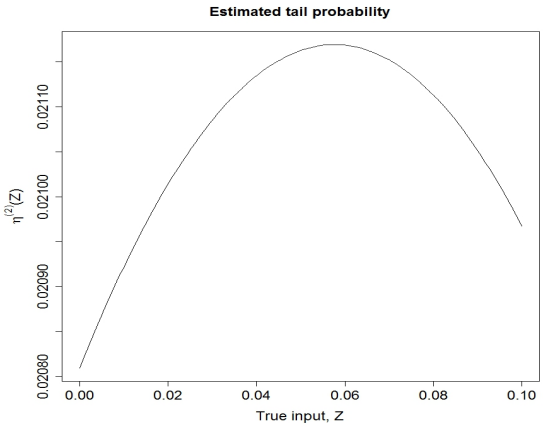
As expected, Figure 3a and Figure 4a show that high internal noise in the data generation mechanism leads to loss of the signal in the sense that the response simply ceases showing any relationship to the input. The rest of the figures show slight distortions of $\eta_1(Z)$ and $\eta_2(Z)$, with a severe distortion of $\eta_3(Z)$, as the internal noise is increased. These may be due to the underestimation of γ, which occurs as the internal noise drowns out the internal signal, and may explain the inclination of the objective functions' maxima towards the right boundary of Ω_Z.



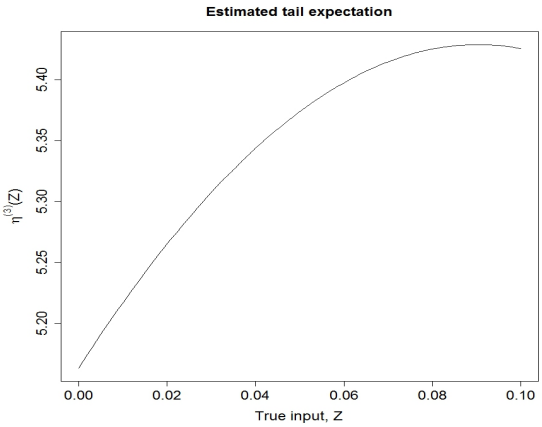
(a) Y vs Z for the simulated data points, true relationship, and estimated relationship.



(b) $\eta_1(Z)$ vs Z for simulated data.



(c) $\eta_2(Z)$ vs Z for simulated data.



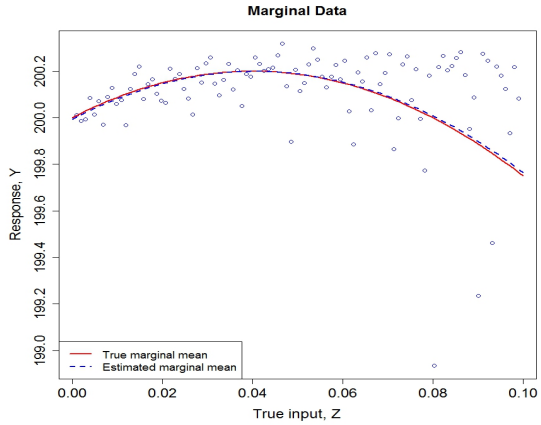
(d) $\eta_3(Z)$ vs Z for simulated data.

Figure 4: High homoscedastic internal noise ($\tau = 2.87$ and $S(\gamma, \tau) = 0.99$).

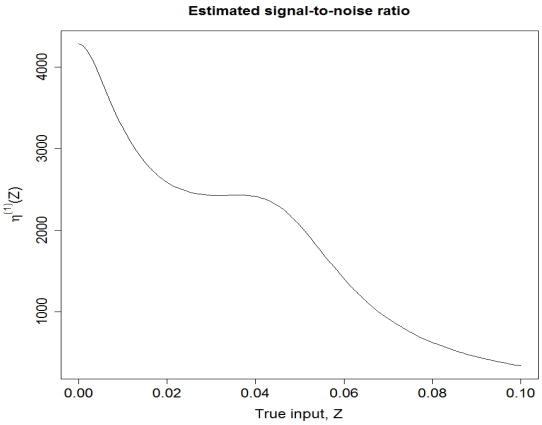
Independent of the underestimation of γ, the internal noise level now directly factors into the calculation of the true optimum, since the optimal point at $\tilde{z}^{(1)} = -\beta_1\gamma/\{2\beta_2(\gamma^2+\tau^2)\}$ represents a shrinkage towards zero, with the degree of shrinkage being directly 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 3a and Figure 5a, we can see the effect of the internal heteroscedasticity on the response. The observed variance increases as Z increases, and the signal is weakened and distorted, but not completely lost. Figure 3b and Figure 5b 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

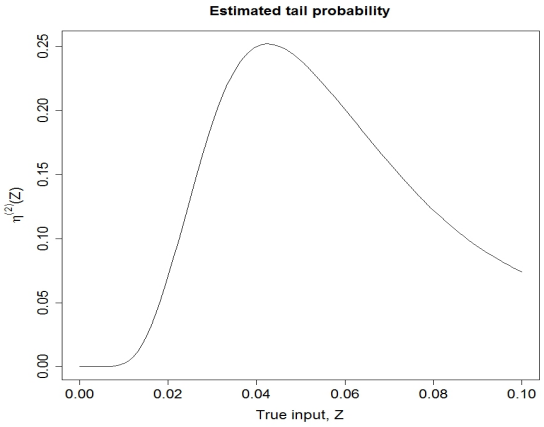
figures depict similar behavior for $\eta_2(Z)$ and $\eta_3(Z)$ when the noise level is increased, which means generally reduced maximum values, and a general shrinkage towards zero which is not as acute as $\eta_1(Z)$'s shrinkage. This makes $\eta_2(Z)$ and $\eta_3(Z)$ more reliable than the signal-to-noise ratio, $\eta_1(Z)$, in this scenario where the internal variance is a linear function of the internal mean function. Based on the reliable performance of $\eta_2(Z)$ in both the homoscedastic and heteroscedastic cases, we can settle on the estimated tail probability, $\eta_2(Z)$, as the objective function of choice in the rest of this paper.



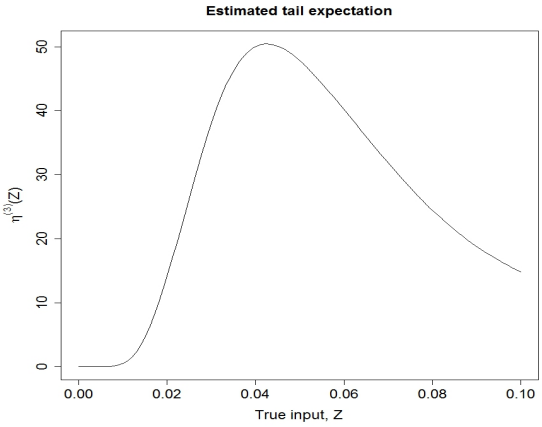
(a) Y vs Z for the simulated data points, true relationship, and estimated relationship.



(b) $\eta_1(Z)$ vs Z for simulated data.



(c) $\eta_2(Z)$ vs Z for simulated data.



(d) $\eta_3(Z)$ vs Z for simulated data.

Figure 5: Medium internal noise ($\tau = 5$ and $S(\gamma, \tau) = 0.5$).

6 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 recored tuples (Z, Y). This is a fairly common scenario, therefore it is important 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 (9), for inferring the relationship between Z and Y when X is missing, is the following misspecified model

$$\begin{aligned} Y|Z, \boldsymbol{\theta} &\sim \mathcal{F}_{(\boldsymbol{\theta}, Z)} \equiv \mathcal{F}(Y|Z, \boldsymbol{\theta}) \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}) \ . \end{aligned} \tag{26}$$

Specifically, the misspecified model based on the simulations in Section 5 is given by

$$\begin{aligned} &Y|(Z, \beta_0, \beta_1, \beta_2, \sigma) \sim \mathcal{N}(\mu(Z),\ \sigma^2) \ \text{ where } \mu(Z) = \beta_0 + \beta_1 Z + \beta_2 Z^2 \\ &(\beta_0, \beta_1, \beta_2, \log(\sigma)) \sim \mathcal{N}(\mathbf{0},\ 10^{10}\boldsymbol{I}) \ . \end{aligned} \tag{27}$$

We now introduce an approach for assessment of the price that an experimenter has to pay to utilize the misspecified model (26) for optimization. Let $\tilde{Z}_{\text{True}}(\tau)$ denote the optimal Z under the true model (9) 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 (26). Denoting some chosen utility function by $\mathcal{J}(\cdot)$, a comparison of the functions $\mathcal{J}\left(\tilde{Z}_{\text{True}}(\tau)\right)$ and $\mathcal{J}\left(\tilde{Z}_{\text{Mis}}(\tau)\right)$ 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*, $P(Y > y^*|Z, \boldsymbol{\theta}, \boldsymbol{\phi})$, and the *true tail expectation*, $E(I_{\{Y>y^*\}}Y|Z, \boldsymbol{\theta}, \boldsymbol{\phi})$, where y^* is chosen to be $\max_X(E(Y|X, \boldsymbol{\theta}))$. 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:

$$\mathcal{K}(\tau) = P\left[Y\left(\tilde{Z}_{\text{True}}\left(\tau\right)\right) > Y\left(\tilde{Z}_{\text{Mis}}\left(\tau\right)\right)\right], \tag{28}$$

where the probability is defined with respect to the true model (9).

6.1 Studying the Effect of Ignoring Homoscedastic Internal Noise through Simulation

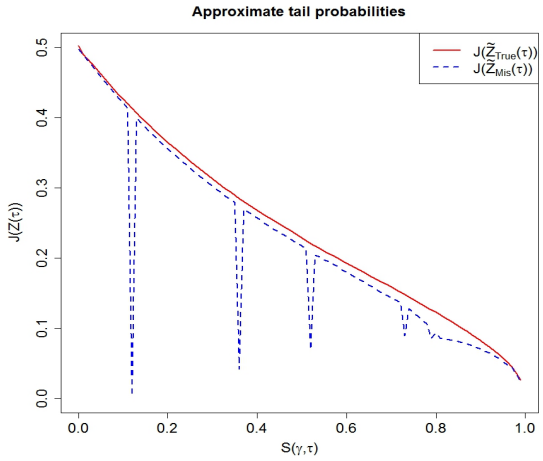
We now explore the loss of performance under the homoscedastic internal noise scenario. For the simulations in this section, we once again use the same simulation set-up as in Section 5, except that we select τ by inverting 100 equi-spaced $S(\gamma = 10, \tau)$ points within the interval $[0,\ 1)$ using (7) for the homoscedastic noise case. For each of the 100 variations of the model parameters, we generate data as in Section 5 under the homoscedastic internal noise setting. 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, $P(Y > y^*|Z, \boldsymbol{\theta}, \boldsymbol{\phi})$ (the true tail probability), and the loss function, $\mathcal{K}(\tau)$ defined by (28). Note that these two functions cannot be written in closed form and have to be numerically approximated using the Monte-Carlo approximation:

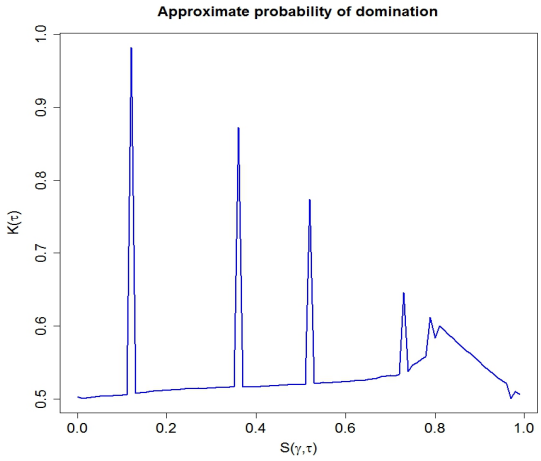
$$P(A) \approx \frac{1}{n} \sum_{i=1}^{n} I(A_i) \ ,$$

where A denotes either the event $Y > y^*$, or the event $Y\left(\tilde{Z}_{\text{True}}(\tau)\right) > Y\left(\tilde{Z}_{\text{Mis}}(\tau)\right)$, $I(\cdot)$ denotes the indicator function, and A_i denotes the occurrence of event A for the ith observation. The two measures in a single simulated dataset are shown in Figure 6, while results averaged over 1000 simulated datasets are shown in Figure 7. In these plots, the unit-free measure of internal noise $S(\gamma, \tau)$ is on the horizontal axis.

The plots in Figure 7 seem to suggest that there is little to be gained on average, from explicitly accounting for homoscedastic internal noise. The plots in Figure 6, however, tell a somewhat different, yet 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. The spikes occur when the truly concave $\mu(Z)$ is estimated to be non-concave, a situation which happens more often as the noise level increases, but is overtaken at higher 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.

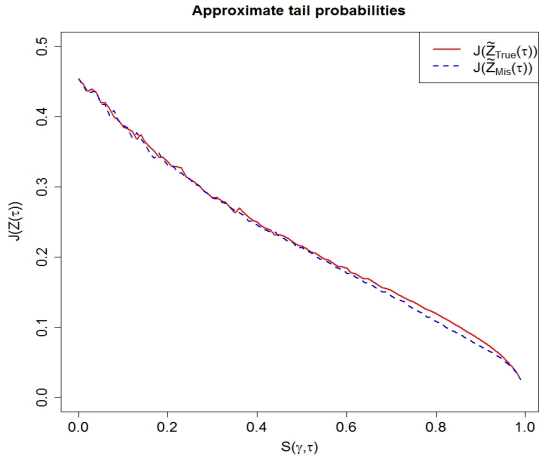


(a) $\mathcal{J}\left(\tilde{Z}_{\text{True}}(\tau)\right)$ and $\mathcal{J}\left(\tilde{Z}_{\text{Mis}}(\tau)\right)$ vs $S(\gamma, \tau)$.

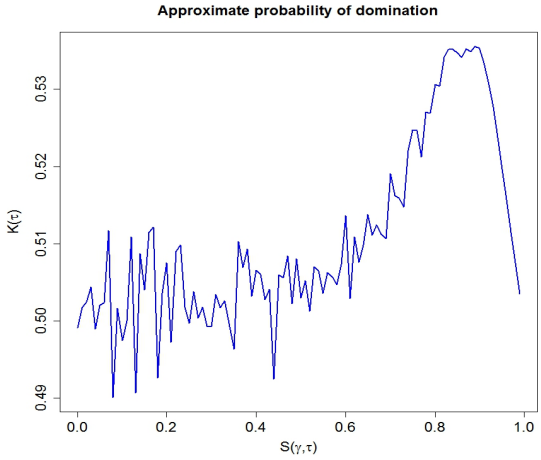


(b) $\mathcal{K}(\tau)$ vs $S(\gamma, \tau)$.

Figure 6: Results for a single simulated dataset.



(a) $\mathcal{J}\left(\tilde{Z}_{\text{True}}(\tau)\right)$ and $\mathcal{J}\left(\tilde{Z}_{\text{Mis}}(\tau)\right)$ vs $S(\gamma, \tau)$.



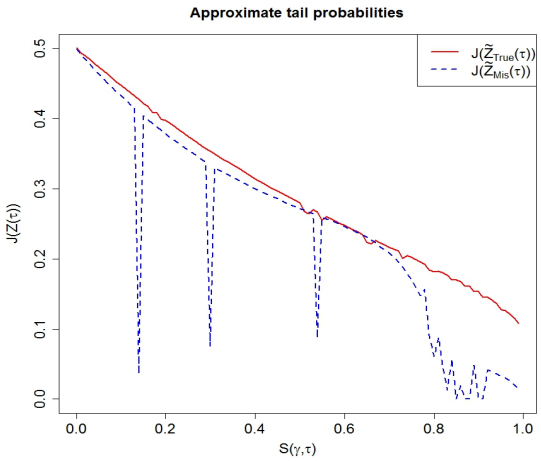
(b) $\mathcal{K}(\tau)$ vs $S(\gamma, \tau)$.

Figure 7: Results averaged over 1000 simulated datasets.

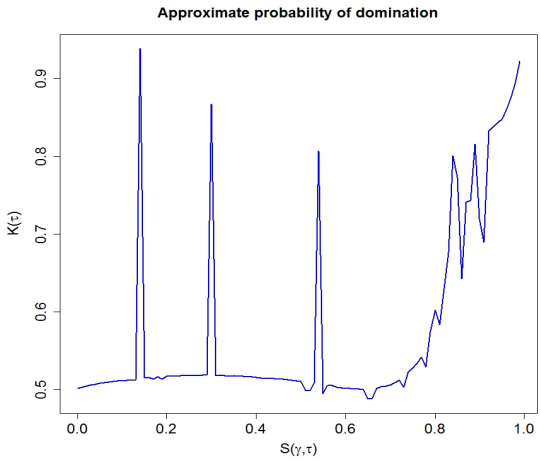
6.2 Studying the Effect of Ignoring Heteroscedastic Internal Noise through Simulation

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. We only need to change the distribution of X to $\mathcal{N}(X|\ \gamma Z,\ \tau^2 Z^2)$, and make the necessary adjustments to the performance measures. The two measures in

a single simulated dataset are shown in Figure 8, and averaged over 1000 simulated datasets in Figure 9.

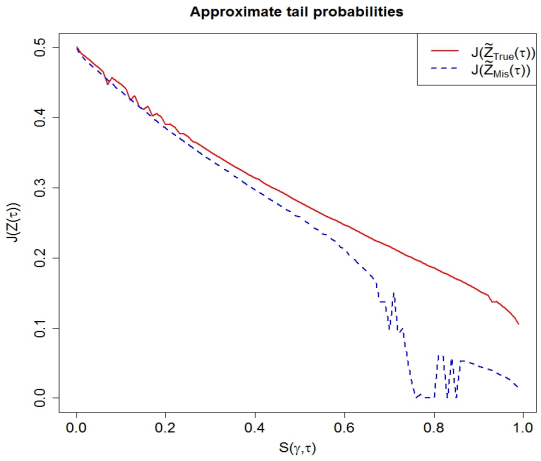


(a) $\mathcal{J}\left(\tilde{Z}_{\text{True}}(\tau)\right)$ and $\mathcal{J}\left(\tilde{Z}_{\text{Mis}}(\tau)\right)$ vs $S(\gamma, \tau)$.

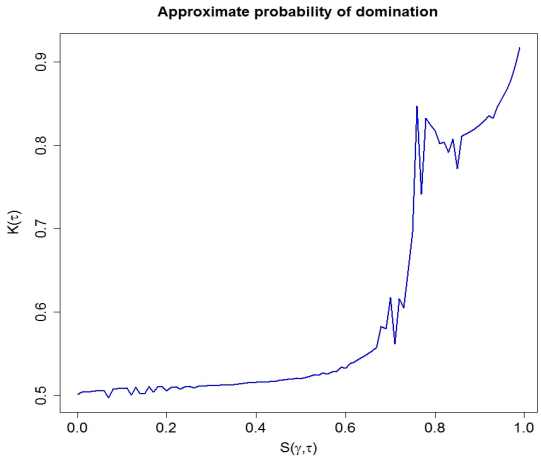


(b) $\mathcal{K}(\tau)$ vs $S(\gamma, \tau)$.

Figure 8: Results for a single simulated dataset.



(a) $\mathcal{J}\left(\tilde{Z}_{\text{True}}(\tau)\right)$ and $\mathcal{J}\left(\tilde{Z}_{\text{Mis}}(\tau)\right)$ vs $S(\gamma, \tau)$.



(b) $\mathcal{K}(\tau)$ vs $S(\gamma, \tau)$.

Figure 9: Results averaged over 1000 simulated datasets.

In Figure 8, the noise-dependent departures of 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 9 also shows the relatively sub par optimization performance based on the misspecified model on average, when compared to the correctly specified (full) model, as the noise level increases. Thus there is clear benefit in accounting for the internal noise, both in individual instances and on average, particularly

when the shape of the true relationship between the input and response is unknown prior to experimentation.

7 Application to CNT Alignment

7.1 Inference and Optimization

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 $\eta_2(Z)$ from (14), and the internal noise measure, $S(\gamma, \tau)$ accordingly. Based on the experiment described in Section 2, the following Bayesian hierarchical model is now proposed to model the response R

$$\begin{aligned} &R|X, C, M, \boldsymbol{\theta} \sim \mathcal{N}\left(\psi(X, C, M),\ \sigma^2\right) \\ &X = GZ \text{ with } G|\boldsymbol{\phi} \sim \Gamma(\alpha,\ \gamma) \equiv \frac{1}{\gamma}\Gamma(\alpha) \\ &(\boldsymbol{\theta}, \boldsymbol{\phi}) \sim \mathcal{N}(\mathbf{0},\ 10^{10}\boldsymbol{I}) \ , \end{aligned} \tag{29}$$

where Γ denotes the gamma distribution, $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \log(\sigma))$, $\boldsymbol{\phi} = (\log(\alpha), \log(\gamma))$, and

$$\psi(X, C, M) = \beta_0 C^2 + (\beta_1 + \beta_2 T_b + \beta_3 T_c + \beta_4 C)X + (\beta_5 + \beta_6 T_c + \beta_7 M)X^2 \ , \tag{30}$$

following the results in Remillard et al. (2016). The 112 CNT experimental samples from Remillard et al. (2016) were generated in three batches denoted by labels: a, b, and c. Using observations from sample a as the baseline, Remillard et al. (2016) thus define T_b and T_c in (30) as indicator variables, representing membership of observations within sample b and sample c respectively. Remillard et al. (2016) also generate 81 supplementary observations for (Z, X) pairs for measuring the internal noise- the experimental and supplementary (Z, X) pairs are depicted in Fig. 2. The gamma distribution model for the internal noise, our addition to the original model by Remillard et al. (2016), ensures that the generated electric field is non-negative and has a linear mean and a mean-dependent standard deviation in agreement with Fig. 2. Thus, we have

$$E(X|Z) = \frac{\alpha}{\gamma}Z \text{ and } Var(X|Z) = \frac{\alpha}{\gamma^2}Z^2 \ . \tag{31}$$

With Z restricted to the interval $[0,\ 2.5]$, the unit-free internal noise measure from (8) now has the form

$$S(\alpha, \gamma) = \frac{4}{\alpha + 4} \ . \tag{32}$$

The updated definition of $\eta_2(Z, C, M)$ based on (14) now becomes

$$\eta_2(Z, C, M) = Pr(Y > y^*|(Z, C, M), \mathcal{D}) \ , \tag{33}$$

where the observed data, $\mathcal{D}$ is now defined as $(\boldsymbol{Y}^{\text{obs}}, \boldsymbol{X}^{\text{obs}}, \boldsymbol{Z}^{\text{obs}}, \boldsymbol{C}^{\text{obs}}, \boldsymbol{M}^{\text{obs}})$. Based on the batch algorithm for computing our objective functions in Section 4.2, we approximate $\eta_2(Z, C, M)$ specifically using the steps:

1. Sample a single representative set of $N(=1000)$ X's, denoted by $\boldsymbol{\mathcal{X}}_N$, by first sampling N Z's uniformly from the interval $[0,\ 2.5]$ volts, and then sampling the respective X's using the gamma density, $\Gamma(G|\ \hat{\alpha},\ \hat{\gamma})$ and the relationship $X = GZ$, with $\hat{\alpha}$ and $\hat{\gamma}$ representing posterior modes.

2. Based on the posterior modes, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$, calculate the approximate quantity

$$\eta_2(Z, C, M) \approx \frac{\sum\limits_{x \in \boldsymbol{\mathcal{X}}_N} Pr(Y > y^*|(x, C, M), \hat{\boldsymbol{\theta}}) \cdot \omega(x, Z|\hat{\boldsymbol{\phi}})}{\sum\limits_{x \in \boldsymbol{\mathcal{X}}_N} \omega(x, Z|\hat{\boldsymbol{\phi}})} \ , \tag{34}$$

   where, by denoting the gamma conditional density function for X by $\Gamma(x|\ \hat{\alpha}Z,\ \hat{\gamma})$, the weights $\omega(x, Z|\hat{\boldsymbol{\phi}})$ are defined as

$$\omega(x, Z|\hat{\boldsymbol{\phi}}) = \frac{\Gamma((x/Z)\ |\ \hat{\alpha},\ \hat{\gamma})}{\sum\limits_{x \in \boldsymbol{\mathcal{X}}_N} \Gamma((x/Z)\ |\ \hat{\alpha},\ \hat{\gamma})} \ .$$

The optimization domain for $\eta_2(Z, C, M)$, denoted by Ω, is restricted to $[0,\ 2.5]$ volts $\times$ $[0.02,\ 0.1]$ mg/mL $\times$ $[0.005,\ 0.025]$ mg, the experimental data domain used by Remillard et al. (2016). Model fit based on (29) and (30), as well as the CNT dataset in Remillard et al. (2016), gives the following posterior modes: $\hat{\beta}_0 = 1.914 \times 10^1,\ \hat{\beta}_1 = 6.026 \times 10^{-2},\ \hat{\beta}_2 = -6.628 \times 10^{-3},\ \hat{\beta}_3 = -3.776 \times 10^{-2},\ \hat{\beta}_4 = -1.499 \times 10^{-1},\ \hat{\beta}_5 = -1.805 \times 10^{-3},\ \hat{\beta}_6 = 1.200 \times 10^{-3}$, and $\hat{\beta}_7 = 3.383 \times 10^{-2}$. These *maximum a posteriori* (MAP) estimates nearly match the regression estimates obtained by Remillard et al. (2016). From our model fit, we also obtain the MAP estimates: $\hat{\sigma} = 6.888 \times 10^{-2}$, $\hat{\alpha} = 1.251 \times 10^3$, and $\hat{\gamma} = 1.458 \times 10^2$. All the above MAP estimates are obtained using the optimization tool *optim*, in *R*.

Posterior samples for the parameters are generated using the *MCMCmetrop1R* tool (Martin et al., 2011) in *R* (with trace plots provided in Appendix B). The means of posterior distributions of all the parameters are fairly close to the MAP estimates reported in the previous paragraph.

Based on (29), (30), and (31), we have that

$$E(Y|Z,C,M,\mathcal{D}) = \beta_0 C^2 + (\beta_1 + \beta_2 T_b + \beta_3 T_c + \beta_4 C)\left(\frac{\alpha}{\gamma} Z\right) + (\beta_5 + \beta_6 T_c + \beta_7 M)\left(\left(\frac{\alpha}{\gamma^2} + \frac{\alpha^2}{\gamma^2}\right) Z^2\right) \ .$$

Thus from (16), and using *optim* to maximize over Z, C, and M, we numerically calculate the threshold, y^*, to be 0.721 for CNT sample a ($T_b = 0$ and $T_c = 0$). Note that y^* nearly matches the predicted maximum response value of 0.72 obtained by Remillard et al. (2016).

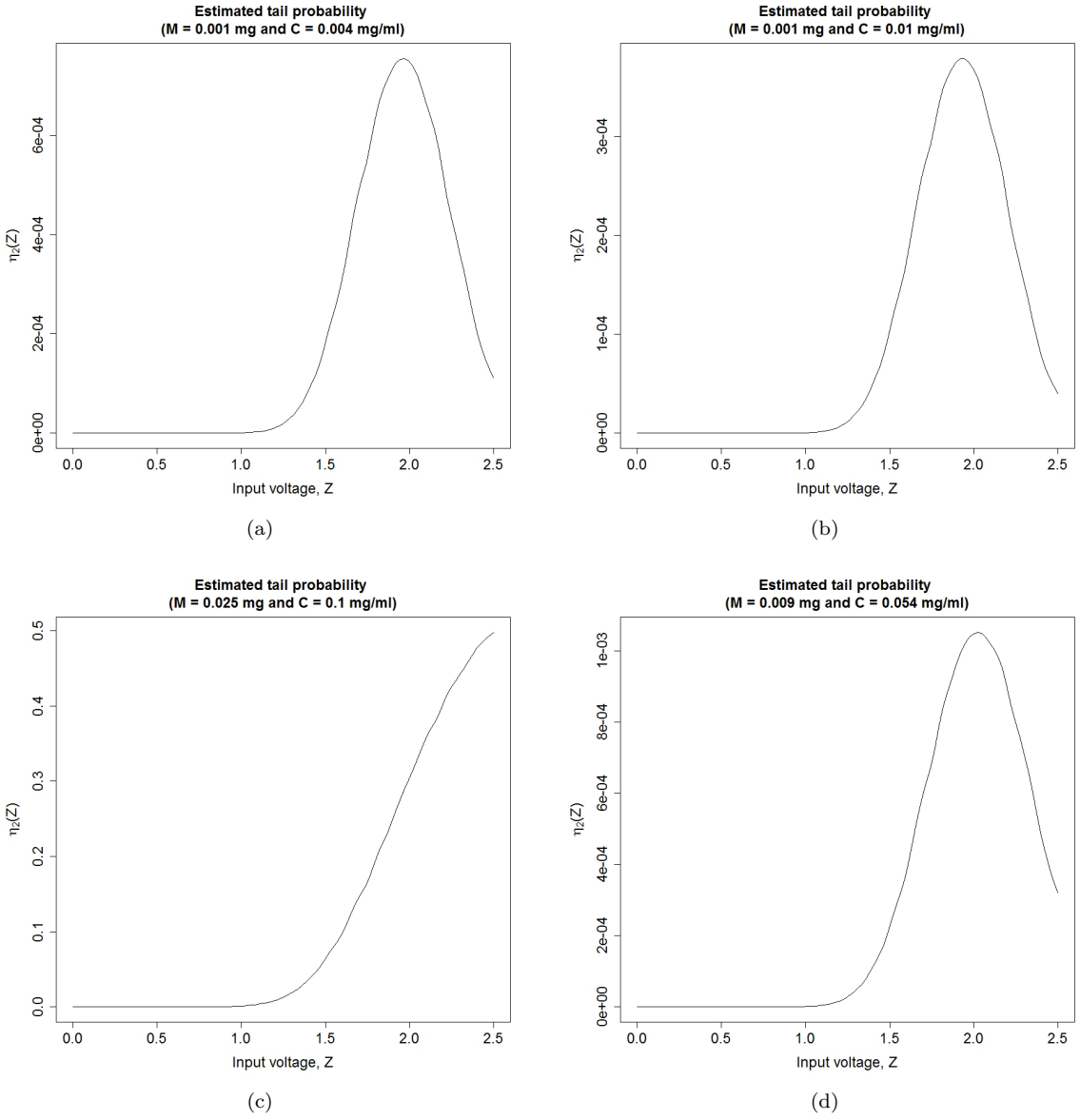


Figure 10: Profiles of $\eta_2(Z, C, M)$ as a function of Z, for four different values of (C, M).

Our task is now to maximize $\eta_2(Z, C, M)$ over the joint domain of (Z, C, M). Due to the difficulty of visualizing a response over a three-dimensional domain, we visualize $\eta_2(Z, C, M)$ as a function of Z, when (C, M) are fixed at some points within their joint domain. For example, we select

$$(C, M) \in \{(0.004, 0.001),\ (0.01, 0.001),\ (0.1, 0.025),\ (0.054, 0.009)\}\ ,$$

where the first three points are corners of the joint domain of (C, M), and the last point is the center of those three. Figure 10 suggests $\eta_2(Z, C, M)$ will be well behaved (unimodal) over the entire domain, Ω, and we can now carry out the optimization.

Using *optim* on the full model in (29), we obtain the following optimal domain values for maximizing CNT alignment

$$\tilde{Z}_{\text{Full}} = 2.5 \text{ volts},\ \tilde{C}_{\text{Full}} = 0.1 \text{ mg/mL},\ \text{and } \tilde{M}_{\text{Full}} = 0.025 \text{ mg}\ .$$

This is the same as obtaining an optimal concentration of 0.01 wt%, and an optimal volume of 0.25 mL, both of which are in perfect agreement with the optimal concentration and volume obtained by Remillard et al. (2016). One advantage of accounting for (and modeling) internal noise, however, is that we are now able to obtain an optimal value of the input voltage, $\tilde{Z}_{\text{Full}}$ as well, instead of the intermediate (and uncontrollable) electric field as obtained by Remillard et al. (2016).

Based on the joint posterior samples of (α, γ), our optimal value of $\tilde{Z}_{\text{Full}} = 2.5$ volts yields a posterior predictive electric field strength expectation, $E(X|Z = \tilde{Z}_{\text{Full}}, \mathcal{D})$, of 21.51 V_{rms}/mm , with a 95% posterior predictive interval (20.65, 22.42) V_{rms}/mm. By optimizing $\hat{\psi}(X, C, M)$ directly, Remillard et al. (2016) obtained the optimal electric field strength, $\tilde{X}$ of approximately 22 V_{rms}/mm. Their value of $\tilde{X}$ is quite close to our value of $E(X|Z = \tilde{Z}_{\text{Full}}, \mathcal{D})$, and falls within our 95% posterior interval for $E(X|Z = \tilde{Z}_{\text{Full}})$. The results match so well in this example because the authors account for internal noise somewhat by recording the intermediate electric field, X, and optimizing $\hat{\psi}(X, C, M)$ directly.

If we assume instead that the intermediate electric field X is not observed/recorded, and that the response model is misspecified as in (26),

$$\begin{aligned} R|Z, C, M, \boldsymbol{\theta} &\sim \mathcal{N}\left(\mu(Z, C, M),\ \nu^2\right) \\ \boldsymbol{\theta} &\sim \mathcal{N}(\mathbf{0},\ 10^{10}\boldsymbol{I})\ , \end{aligned} \tag{35}$$

where $\boldsymbol{\theta} = (B_0, B_1, B_2, B_3, B_4, B_5, B_6, B_7, \log(\nu))$, and

$$\mu(Z, C, M) = B_0C^2 + (B_1 + B_2T_b + B_3T_c + B_4C)Z + (B_5 + B_6T_c + B_7M)Z^2 \ ,$$

then we obtain the following MAP estimates: $\hat{B}_0 = 1.925 \times 10^1$, $\hat{B}_1 = 5.221 \times 10^{-1}$, $\hat{B}_2 = -5.652 \times 10^{-2}$, $\hat{B}_3 = -3.289 \times 10^{-1}$, $\hat{B}_4 = -1.255$, $\hat{B}_5 = -1.330$, $\hat{B}_6 = 8.872 \times 10^{-2} \times 10^{-1}$, $\hat{B}_7 = 2.283$ and $\hat{\nu} = 7.034 \times 10^{-2}$.

Optimizing $\hat{\mu}(Z, C, M)$ gives us the following optimal domain values for maximizing CNT alignment

$$\tilde{Z}_{\text{Mis}} = 2.5 \text{ volts}, \ \tilde{C}_{\text{Mis}} = 0.1 \text{ mg/mL}, \ \text{ and } \tilde{M}_{\text{Mis}} = 0.025 \text{ mg} \ .$$

Note that $\tilde{Z}_{\text{Mis}} = \tilde{Z}_{\text{Full}}$, $\tilde{C}_{\text{Mis}} = \tilde{C}_{\text{Full}}$ and $\tilde{M}_{\text{Mis}} = \tilde{M}_{\text{Full}}$. This is because the internal noise level is itself almost negligible. Using posterior draws of α and γ, the posterior distribution of $S(\alpha, \gamma)$ is obtained. As shown in Figure 11, a 95% posterior interval for $S(\alpha, \gamma)$ based on this distribution is calculated as $(0.0014, \ 0.0026)$). Thus the benefit of accounting for internal noise is barely noticeable. If, however, the underlying α and γ had been several orders of magnitude smaller (while leaving α/γ constant so that $E(X|Z, \alpha, \gamma)$ remains the same), then $S(\alpha, \gamma)$ as given by (32) would be closer to 1, and the benefit of accounting for internal noise would be more obvious. We show this using a data-based simulation study in the next subsection.

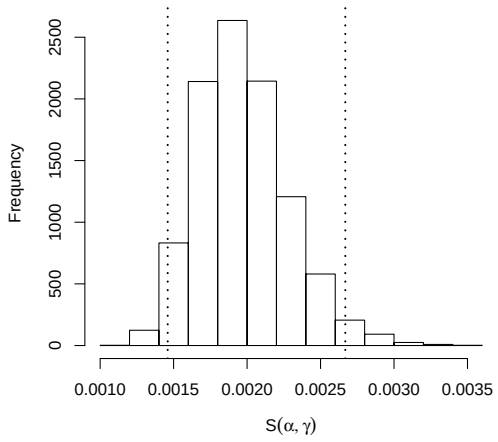


Figure 11: Posterior distribution of $S(\alpha, \gamma)$

7.2 A Data-based Simulation Study for Assessing Performance Loss Under Zero-Internal-Noise Models

In a similar vein to the simulations carried out in Section 6, we now design simulations to assess performance loss when the misspecified model in (35) is used instead of the full model in (29), as the internal noise is increased. A big difference from the simulations in Section 6, however, is that the simulations in this section will be based on the actual observed data points used by Remillard et al. (2016). This means, we retain $(\boldsymbol{Z}^{\text{obs}}, \boldsymbol{C}^{\text{obs}}, \boldsymbol{M}^{\text{obs}})$, and simulate the corresponding $(\boldsymbol{X}^{\text{obs}}, \boldsymbol{Y}^{\text{obs}})$ data as necessary, for each new internal noise level. Specifically,

1. We create a grid of 100 internal noise levels using 100 equi-spaced values for $S(\alpha, \gamma)$ from the interval $(0,\ 1)$.

2. We create corresponding grid values for α and γ by inverting the grid for $S(\alpha, \gamma)$, as well as using the condition that $\alpha/\gamma = \hat{\alpha}/\hat{\gamma}$, where $\hat{\alpha}$ and $\hat{\gamma}$ are the MAP estimates we obtained based on (29). Note from (32), that increasing the magnitude of the $(\alpha,\ \gamma)$ pairs leads to lower internal noise levels.

3. We set $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, where $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \log(\sigma))$, and $\hat{\boldsymbol{\theta}}$ represents the MAP estimates we obtained based on (29).

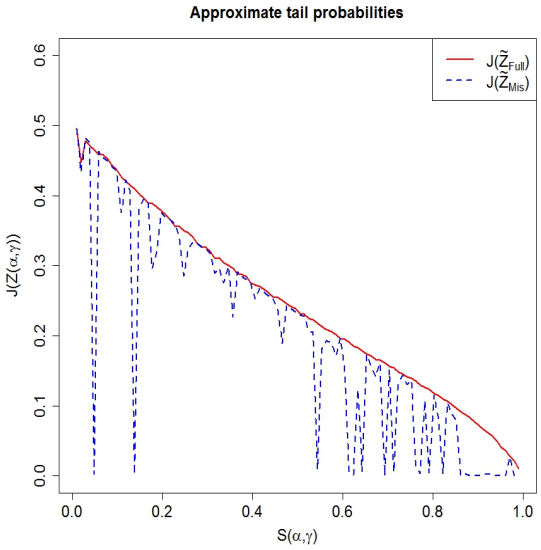
4. For each grid pair of $(\alpha,\ \gamma)$, and conditional on the value of $\boldsymbol{\theta}$ above, generate $(\boldsymbol{X}^{\text{obs}}, \boldsymbol{Y}^{\text{obs}})$ for corresponding $(\boldsymbol{Z}^{\text{obs}}, \boldsymbol{C}^{\text{obs}}, \boldsymbol{M}^{\text{obs}})$ from the CNT dataset, using the full model in (29). We also simulate corresponding X's for the 81 supplementary Z observations from Remillard et al. (2016) as well.

5. Now repeat the analyses in Section 7.1, that is, carry out inference based on the generated data and obtain optimal points, $\tilde{Z}_{\text{Full}}$ and $\tilde{Z}_{\text{Mis}}$, using the full model in (29), and the misspecified model in (35) respectively.

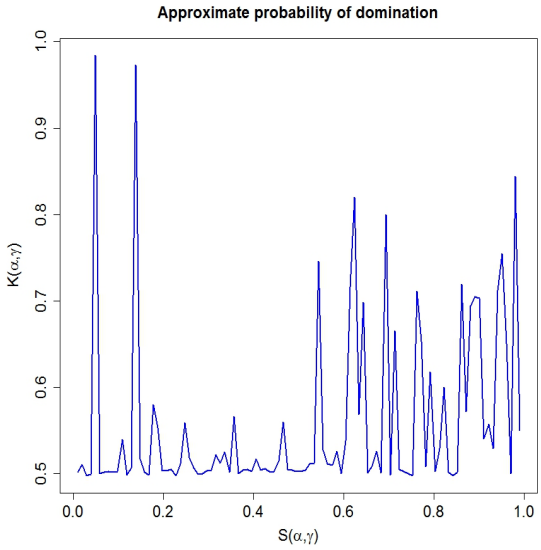
6. As in Section 6, evaluate the optimal points obtained using both the utility function $\mathcal{J}(Z)$, that is, either the *true tail probability*, $P(Y > y^*|Z, C, M, \boldsymbol{\theta}, \boldsymbol{\phi})$, and the *true probability of domination* $\mathcal{K}(\alpha, \gamma)$ defined as $P\left[Y\left(\tilde{Z}_{\text{Full}}(\alpha, \gamma)\right) > Y\left(\tilde{Z}_{\text{Mis}}(\alpha, \gamma)\right)|\boldsymbol{\theta}, \boldsymbol{\phi}\right]$.

Fig. 12 illustrates the performance gains expected when the internal noise increases, assuming the full model in (29) is correct, and the same (Z, C, M) domain points as in Remillard et al. (2016) are used. In particular, we see from Fig. 12a that optimal points obtained using the full model (29) exhibit superior performance on average, when compared to optimal points from the misspecified model (35), as internal noise

level increases. This superior performance is further supported by the increased probability of domination of optimal points from the full model, over optimal points from the misspecifed model, as internal noise increases, as shown in Fig. 12b.



(a) $\mathcal{J}\left(\tilde{Z}_{\text{Full}}(\alpha,\gamma)\right)$ and $\mathcal{J}\left(\tilde{Z}_{\text{Mis}}(\alpha,\gamma)\right)$ vs $S(\alpha,\gamma)$.



(b) $\mathcal{K}(\alpha,\gamma)$ vs $S(\alpha,\gamma)$.

Figure 12: Results averaged over 500 simulated datasets.

8 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 tail probability, 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 presence 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. (2016), due to the minimal internal noise present in the CNT dataset. As the data-based simulation study suggests however, we should expect more substantial gains under relatively high internal noise levels.

The proposed approach has the following limitations that lead to some potential topics for future research:

(1) The computation of the objective functions is based on substituting posterior modes to bypass evaluation of complex integrals. Such a procedure is essentially similar to one in which frequentist estimates are substituted, especially when the priors are vague. Development of efficient algorithms to perform the computation and optimization using a fully Bayesian procedure that makes use of the posterior draws will be an interesting topic for future research.

(2) We have assumed that the variables $\boldsymbol{X}$ and $\boldsymbol{W}$ in the model

$$Y|X, \boldsymbol{W}, \boldsymbol{\theta} \sim \mathcal{F}_{(\boldsymbol{\theta}, \boldsymbol{W}, X)} \equiv \mathcal{F}(Y|X, \boldsymbol{W}, \boldsymbol{\theta})$$

are already identified, which is equivalent to assuming that variable selection has already been performed prior to the Bayesian analysis. A useful extension of the proprosed procedure would be to develop a unified approach in which Bayesian variable selection is included. This can possibly be done along the lines of Chipman et al. (1997) by incorporating indicator variables into the hierarchical model.

Acknowledgment

We are very thankful to two reviewers, an AE and the editor for their many careful, insightful remarks and constructive suggestions, which significantly improved the quality of this paper. This research was supported by NSF grant DMS-1612901.

A Signal to Noise Ratio

First note from (22) that by completing the square, for fixed $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \sigma^2)$ and $\boldsymbol{\phi} = (\gamma, \tau)$ we can rewrite $\mu(X)$ as

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

$$\text{where } \sqrt{\eta(X)} = \frac{1}{\tau} \left(X + \frac{\beta_1}{2\beta_2} \right).$$

Since the conditional distribution of $\sqrt{\eta(X)}$ given Z, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ is normal with

$$E\left[\sqrt{\eta(X)} | Z, \boldsymbol{\theta}, \boldsymbol{\phi} \right] = \frac{1}{\tau} \left(\gamma Z + \frac{\beta_1}{2\beta_2} \right), \ \text{Var}\left[\sqrt{\eta(X)} | Z, \boldsymbol{\theta}, \boldsymbol{\phi} \right] = 1,$$

it follows that conditional on Z, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, $\eta(X)$ is itself a non-central chi-square random variable with $k = 1$ degree of freedom and non-centrality parameter

$$\lambda(Z) = \frac{1}{\tau^2} \left(\gamma Z + \frac{\beta_1}{2\beta_2} \right)^2 .$$

Thus the conditional distribution of $\mu(X)$ given Z, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ 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 the conditional expectation of Y given Z, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ is:

$$\begin{aligned}E(Y | Z, \boldsymbol{\theta}, \boldsymbol{\phi}) &= E(\mu(X) | Z, \boldsymbol{\theta}, \boldsymbol{\phi}) \\ &= \beta_2 \tau^2 (k + \lambda(Z)) + \left(\beta_0 - \frac{\beta_1^2}{4\beta_2} \right) \\ &= (\beta_0 + \beta_2 \tau^2) + (\beta_1 \gamma) Z + (\beta_2 \gamma^2) Z^2 \ ,\end{aligned} \tag{36}$$

after simplification. Also, the conditional variance is:

$$\begin{aligned} Var(Y|Z,\boldsymbol{\theta},\boldsymbol{\phi}) &= E(\sigma^2|Z,\boldsymbol{\theta},\boldsymbol{\phi}) + Var(\mu(X)|Z,\boldsymbol{\theta},\boldsymbol{\phi}) \\ &= \sigma^2 + 2\beta_2^2\tau^4(k + 2\lambda(Z)) \\ &= \sigma^2 + 2\beta_2^2\tau^2\left(\tau^2 + 2\left(\gamma Z + \frac{\beta_1}{2\beta_2}\right)^2\right) \ . \end{aligned} \tag{37}$$

Since in the simulation setting of Section 5, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ have point-mass priors, the result follows from (13).

B MCMC Trace Plots

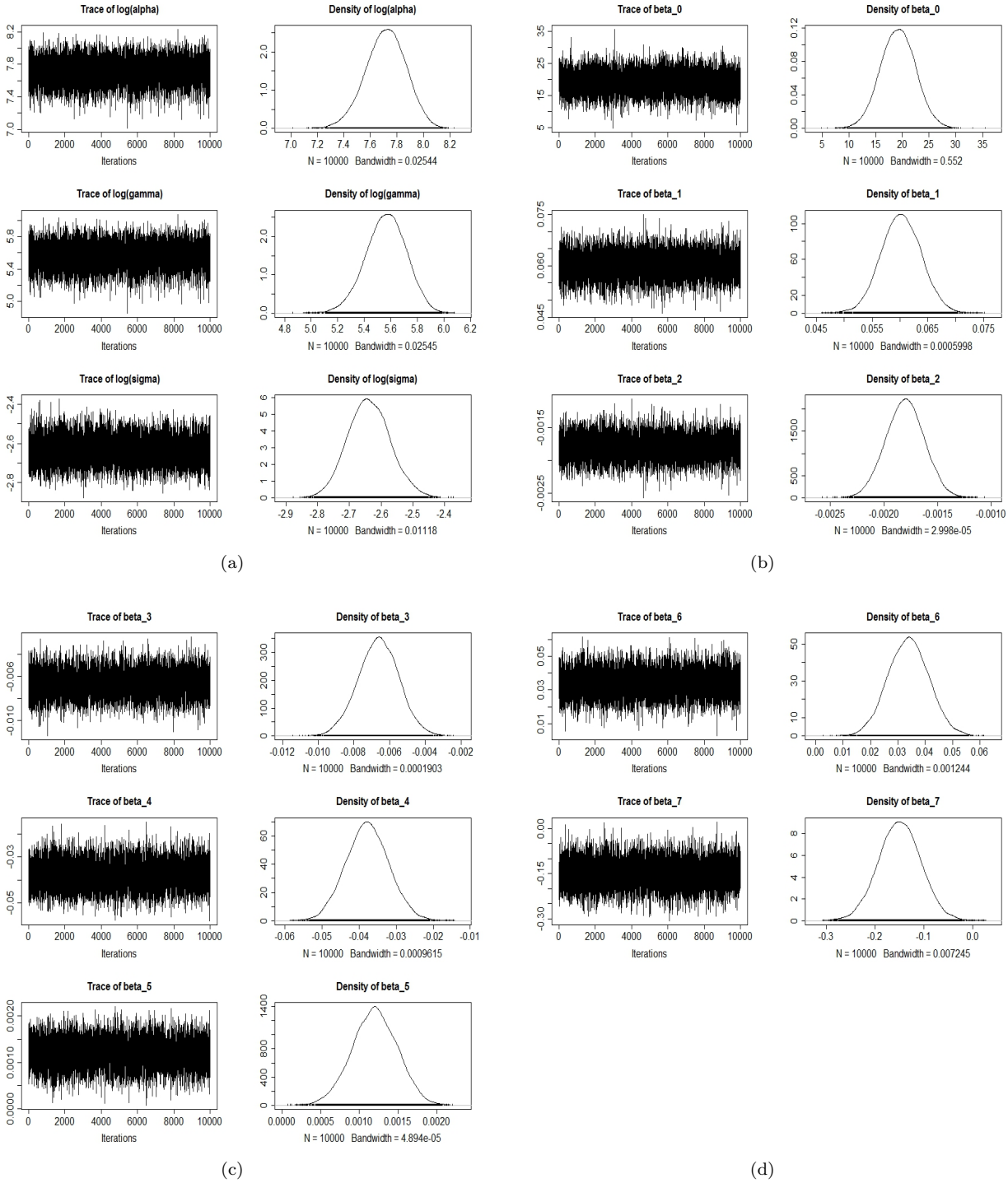


Figure 13: Posterior trace plots and distributions for parameters from (29).

References

M. K. Ardakani, D. Das, S. S. Wulff, and T. J. Robinson. Estimation in second-order models with errors in the factor levels. *Communications in Statistics - Theory and Methods*, 40:1573–1590, 2011.

J. Bérubé and C. F. J. Wu. Signal-to-noise ratio and related measures in parameter design optimization: An overview. *Sankhya - Series B*, 62:417–432, 2000.

G. E. P. Box. Signal-to-noise ratios, performance criteria, and transformations (with discussions). *Technometrics*, 30:1–40, 1988.

H. Chipman, M. Hamada, and C. F. J. Wu. A bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics*, 39:372–381, 1997.

R. Das, M. E. Ali, S. B. A. Hamid, S. Ramakrishna, and Z. Z. Chowdhury. Carbon nanotube membranes for water purification: A bright future in water desalination. *Desalination*, 336:97–109, 2014.

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

C. De Las Casas and W. Li. A review of application of carbon nanotubes for lithium ion battery anode material. *Journal of Power Sources*, 208:74–85, 2012.

X. Du and W. Chen. Sequential optimization and reliability assessment method for efficient probabilistic design. *Journal of Mechanical Design*, 126:225–233, 2004.

C. Gao, Z. Guo, J. Liu, and X. Huang. The new age of carbon nanotubes: An updated review of functionalized carbon nanotubes in electrochemical sensors. *Nanoscale*, 4:1948, 2012.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2 edition, July 2003. ISBN 158488388X.

M. Hamada and S. Steiner. Making mixtures robust to noise factors and mixing measurement errors. *Journal of Quality Technology*, 29:441–450, 1997.

M. Hamada, H. F. Martz, and S. Steiner. Accounting for mixing errors in analyzing mixture experiments. *Journal of Quality Technology*, 37:139–148, 2005.

N. S. Lee, D. S. Chung, I. T. Han, J. H. Kang, Y. S. Choi, H. Y. Kim, S. H. Park, Y. W. Jin, W. K. Yi, M. J. Yun, J. E. Jung, C. J. Lee, J. H. You, S. H. Jo, C. G. Lee, and J. M. Kim. Application of carbon nanotubes to field emission displays. *Diamond and Related Materials*, 10:265–270, 2001.

R. V. Leon, A. C. Shoemaker, and R. N. Kacker. Performance measures independent of adjustment: An explanation and extension of taguchi's signal-to-noise ratios (with discussions). *Technometrics*, 29:253–285, 1987.

W. Li and C. F. J. Wu. An integrated method of parameter design and tolerance design. *Quality Engineering*, 11:417–425, 1999.

J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer Publishing Company, Incorporated, 2008. ISBN 0387763694, 9780387763699.

P. Liu, L. Liu, Y. Wei, K. Liu, Z. Chen, K. Jiang, Q. Li, and S. Fan. Fast high-temperature response of carbon nanotube film and its application as an incandescent display. *Advanced Materials*, 21:3563–3566, 2009.

A. D. Martin, K. M. Quinn, and J. H. Park. MCMCpack: Markov chain monte carlo in R. *Journal of Statistical Software*, 42(9):1–21, 2011. URL `http://www.jstatsoft.org/v42/i09/`.

V. N. Nair. Taguchi's parameter design: A panel discussion. *Technometrics*, 34:127–161, 1992.

S. Park, M. Vosguerichian, and Z. Bao. A review of fabrication and applications of carbon nanotube film-based flexible electronics. *Nanoscale*, 5:1727–1752, 2013.

M. S. Phadke. *Quality Engineering using Robust Design*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

E. M. Remillard, Q. Zhang, S. Sosina, Z. Branson, T. Dasgupta, and C. D. Vecitis. Electric-field alignment of aqueous multi-walled carbon nanotubes on microporous substrates. *Carbon*, 100:578–589, 2016.

G. Taguchi. *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs*. White Plains, NY: UNIPUB, Kraus International Publications, 1987.

C. D. Vecitis, G. Gao, and H. Liu. Electrochemical carbon nanotube filter for adsorption, desorption, and oxidation of aqueous dyes and anions. *Journal of Physical Chemistry C*, 115:3621–3629, 2011.

I. N. Vuchkov and L. N. Boyadjieva. *Quality Improvement with Design of Experiments: A response Surface Approach*. Kluwer Academic Publishers, Boston, 2001.

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

X. Yin and W. Chen. Enhanced sequential optimization and reliability assessment method for probabilistic optimization with varying design variance. *Structure and Infrastructure Engineering*, 2:261–275, 2006.