

2014

A New Gaussian Process Method For Modeling and Design of Multi-Source Data in Exposure-Response Toxicology Studies

Kai Wang

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Wang, Kai, "A New Gaussian Process Method For Modeling and Design of Multi-Source Data in Exposure-Response Toxicology Studies" (2014). *Graduate Theses, Dissertations, and Problem Reports*. 6902.
<https://researchrepository.wvu.edu/etd/6902>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

A New Gaussian Process Method For Modeling and Design
of Multi-Source Data in Exposure-Response Toxicology Studies

Kai Wang

Dissertation submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Industrial Engineering

Feng Yang, Ph.D., Chair
Xi Chen, Ph.D., Co-chair
Majid Jaridi, Ph.D.
Dale Porter, Ph.D.
Xiaopeng Ning, Ph.D.

Department of Industrial and Management Systems Engineering

Morgantown, West Virginia
2014

Keywords: Stochastic kriging, Qualitative factors, Multi-source data, Design of experiments

Copyright 2014 Kai Wang

ABSTRACT

A New Gaussian Process Method For Modeling and Design of Multi-Source Data in Exposure-Response Toxicology Studies

Kai Wang

One of the most fundamental steps in risk assessment is to quantify the exposure-response relationship for the material/chemical of interest. This work develops a new statistical method, referred to as SKQ (stochastic kriging with qualitative factors), to synergistically model exposure-response data, which often arise from multiple sources (e.g., laboratories, animal providers, and shapes of nanomaterials) in toxicology studies. Compared to the existing methods, SKQ has several distinct features. First of all, SKQ integrates data across multiple sources, and allows for the derivation of more accurate information from limited data. Second, SKQ is highly flexible and able to model practically any continuous response surfaces (e.g., dose-time-response surface). Third, SKQ is able to accommodate variance heterogeneity across experimental conditions, and to provide valid statistical inference (i.e., quantify uncertainties of the model estimates). Through empirical studies, we have demonstrated SKQ's ability to efficiently model exposure-response surfaces by pooling information across multiple data sources.

Based on the SKQ modeling and inference, a design of experiments (DOE) procedure is developed to guide biological experiments for the efficient quantification of exposure-response relationships. Built on SKQ, the DOE procedure inherits the advantages of SKQ and is particularly tailored for experimental data arising from multiple sources, with non-normality and variance heterogeneity, and mapping nonlinear exposure-response relationships. The design procedure is built in a sequential two-stage paradigm that allows for a

learning process: In the first stage, preliminary experiments are performed to gain information regarding the underlying exposure-response curve and variance structure; in the second stage, the prior information obtained from the previous stage is utilized to guide the second-stage experiments. Matlab's global optimization function **MultiStart** is employed to search for optimal designs that will lead to exposure-response models of the highest quality.

SKQ and SKQ-based DOE fit into the mosaic of efficient decision-making methods for assessing the risk of a tremendously large variety of nanomaterials, and helps to alleviate the sustainability concerns regarding the enormous new nanomaterials.

ACKNOWLEDGMENTS

First of all, I want to thank my advisor, Dr. Feng Yang, for her great guidance and continuous support throughout my study at West Virginia University. It has been a great honor to be her master and Ph.D. student. Without her inspiration, it wouldn't be possible for me to finish probably most challenging thing, completion of Ph.D., of the first 27 years of my life. I would like to acknowledge my co-chair, Dr. Xi Chen. She provides many great advice and ideas in stochastic kriging with qualitative variables and design of experiments. I would also like to thank Dr. Majid Jaridi, Dr. Dale Porter and Dr. Xiaopeng Ning for serving on my committee, and for their insightful comments and suggestions in preparing this dissertation.

I am especially grateful to current or past graduate students of Industrial Engineering department who work and play with me: Jingang Liu, Chihui Li, Yuping Huang, Zongyu Geng, and Minqi Li.

Finally, I would like to express thanks my family for all their love and encouragement.

Contents

List of Figures	vii
List of Tables	viii
1 Introduction	1
1.1 Challenges of the Research	2
1.2 Research Objective	3
1.3 Contribution of the Research	4
1.4 Organization of the Dissertation	4
2 Problem Statement and Literature Review	6
2.1 Statement of the Research Problem	6
2.2 The Naive Method without Pooling	8
2.3 An Information-Pooling Model: Mixed-Effects Model (MEM)	8
2.3.1 Formulation and Assumptions for Mixed-Effects Model (MEM)	8
2.3.2 Estimation and Inference by Mixed-Effects Model (MEM)	10
2.4 Design of Experiments for Learning/Inferences of Gaussian Processes	12
3 Stochastic Kriging With Qualitative Factors (SKQ)	13
3.1 Extrinsic Variance Structure	14
3.2 Intrinsic Variance Structure	17
3.3 Estimation and Inference by SKQ (Stochastic Kriging with Qualitative Factors)	17
3.4 Inverse Estimation and Inference	20

3.5	Simulation Studies	22
3.5.1	Case 1: Modeling Multi-Source Dose-Time-Response Data	23
3.5.2	Case 2: Modeling Multi-Source Dose-Response Data	27
4	Design of Multiple-Source Experiments	35
4.1	The Two-Stage Procedure	35
4.1.1	Stage I: Initial Design	37
4.1.2	Stage II: Design Augmentation	38
4.2	Simulation Studies	39
4.2.1	Case 1: Design for Multi-Source Dose-Time-Response Experiments .	39
4.2.2	Case 2: Design of Multi-Source Dose-Response Experiments	45
5	Summary and Future Studies	52
6	Appendix	54
6.1	Review of Standard Stochastic Kriging (SK)	54
6.1.1	The Extrinsic and Intrinsic Variance Structures	55
6.1.2	Estimation and Inference by Standard Stochastic Kriging	57
6.2	True Expected Exposure-Response Model for Case 1	59
6.3	An Estimation Data Set (EDS) for Case 2	60
	References	61

List of Figures

3.1	The bootstrap resampling algorithm for uncertainty quantification of BMD estimates.	21
3.2	The true exposure-response surfaces for Case 1.	24
3.3	Deign points in the EDS (estimation data set) and check points in the VDS (validation data set)	26
3.4	Comparison of the dose-response fitting results from MEM and SKQ	32
3.5	Box plots for the BMDLs resulting from the two modeling methods.	34
4.1	Overview of the two-stage design procedure.	36
4.2	The true exposure-response surfaces for Design Case 1.	40
4.3	Design points and check points in the VDS (validation data set) of Short TiO ₂ nanobelts for Case 1	41
4.4	Design points and check points in the VDS (validation data set) of long TiO ₂ nanobelts of three design approaches for Case 1	43
4.5	The box plots for the DEVs resulting from the three design approaches for Case 1.	44
4.6	The true and SKQ estimated dose-response curves for Case 2.	49
4.7	The box plots for the BMDLs resulting from the two design methods for Case 2.	51

List of Tables

3.1	Comparison of the estimation results by SK and SKQ for Case 1	28
3.2	Design points in the EDS (estimation data set) for Case 2.	29
3.3	The SKQ parameters estimated from the normalized dose-response data for Case 2.	31
3.4	Comparison of MEM and SKQ in terms of the CI coverage probabilities for the expected response $Y(\cdot)$	33
3.5	Comparison of MEM and SKQ in terms of the CI coverage probabilities for the BMD.	34
4.1	Comparison of the estimation results from three design methods for Case 1 .	45
4.2	The design of experiments of run c_1 for Case 2.	47
4.3	The simulation response data of run c_1 for Case 2.	47
4.4	The design of experiments in Stage I of run c_2 for Case 2.	48
4.5	The simulation response data in Stage I of run c_2 for Case 2.	48
4.6	The augmented design in Stage II of run c_2 for Case 2.	48
4.7	The simulation response data in Stage II of run c_2 for Case 2.	49
6.1	An estimation data set (EDS) for Case 2: modeling the multi-source dose- response data	60

Chapter 1

Introduction

Nanomaterials (NM) are finding wide applications in areas such as energy [1, 2], environment [3–5], and biomedical engineering sectors [6, 7]. The rapid introduction of engineered NM raises imperative concerns on the potential hazard/risk of NM. In comparison to traditional materials and chemicals, it is particularly challenging to fully assess the risks associated with all NM, mainly due to the tremendously large variety of NM. With limited resources (time and/or money), there is an urgent need to develop efficient decision-making methods for NM risk assessment. Primarily motivated by such a need, this dissertation proposes a new statistical method, which is able to derive more accurate hazard-related information by synergistically modeling multi-source toxicology data. By making more efficient use of limited data, the proposed method will help to reduce the experimental cost/time in toxicology studies, and contribute to the use of NM in a safe and sustainable manner.

One of the most fundamental steps in assessing the risk of a nanomaterial (or any substance) is to understand and properly characterize its exposure-response relationship [8, 9]. A exposure-response relationship describes how the adverse bioactivity effects (the responses) are functionally related to the condition of exposure to a substance [10]. With a well established exposure-response profile, prediction of hazard can be made for a given level of exposure; such a profile also allows for the estimation of exposures at responses of different severities (e.g., the benchmark dose), which assists the risk assessor to make judgments to protect a population from increasingly severe effects.

To quantify exposure-response relationships, biological experiments need to be performed under different exposure conditions to observe the corresponding bioactivity responses of animals. Herein, the exposure condition is typically specified through the settings of two quantitative factors: the dose level of the substance of interest, and the time factor involved. Depending on the time scope of the toxicology study, the time factor could be exposure time for long-term studies, or post-exposure time for acute studies. Based on the experimental data collected, statistical methods are then used to fit exposure-response models quantifying the relationships of interest.

To efficiently characterize a substance (especially a nanomaterial) by its exposure-response profile, two fundamental and important research issues need to be resolved: (i) How to most efficiently design the expensive and time-consuming biological experiments for data collection? (ii) Given the collected data, how to achieve the exposure-response model of the highest quality by performing the statistical modeling most suitable to the data?

1.1 Challenges of the Research

It remains a challenge to achieve the exposure-response models of the highest quality given a limited amount of data. Existing statistical models are not adequate to model typical exposure-response data due to the following three major reasons. (i) First, the exposure-response data for a nanomaterial often arise from multiple sources. More specifically, the data available for modeling typically consists of a number of subsets obtained from different sources such as laboratories [11], animal providers[12], and shapes of nanomaterials [13]. The observed response of animals depends not only on the exposure condition, but also on these source factors. The subset of data collected from one source may well reflect a different exposure-response profile than that from another. (ii) Second, the exposure-response relationship may well be nonlinear [14–18]. For two-dimensional dose-response curves, a range of nonlinear regression models (e.g., power and logistic models) have been developed [19] to model the sole-source data, as opposed to the multi-source data described above. However,

the three-dimensional dose-time-response surface is far from being adequately investigated at least partly because of the complex nature of the target surface. (iii) Third, typically the available/affordable toxicology data are not only relatively scarce and highly variable, but are also subject to variance heterogeneity [20]: The response variability changes with the experimental setup, which is specified by the exposure condition as well as the data-source factors. The variance heterogeneity poses significant challenges in making statistical inferences via the estimated exposure-response model (i.e., quantifying the uncertainty of the model estimates).

1.2 Research Objective

The objective of this dissertation work is two folds.

A new stochastic kriging model, which will be referred to as SKQ (stochastic kriging with qualitative factors) in this dissertation, is developed to model exposure-response data in toxicology studies. SKQ distinguishes from the existing modeling methods in the following aspects. (i) SKQ leads to exposure-response models of high quality by pooling information from multi-source data. (ii) SKQ is highly flexible and able to accurately approximate practically any continuous response surfaces [21–23]; it does not require a pre-assumed functional form (e.g., logistic model) for response surface modeling, as traditional nonlinear regression does [24]. (iii) SKQ is able to accommodate variance heterogeneity across exposure conditions as well as different data sources, and to provide valid statistical inference.

The second objective is to develop a DOE method based on the SKQ modeling. Because of costs, ethics, or other limitations, sample sizes are usually very restricted for toxicology studies, and hence efficient use of available resources is crucial. In the multi-source data context, biological experiments are naturally performed in a group sequential (or batch-by-batch) manner: Given the data collected from $Q - 1$ distinct sources (say, labs), how to design the new batch of biological experiments to be carried out in the Source Q (say, our lab)? The task of DOE is to optimize the quality of the exposure-response model

fitted from all the Q -source data, by determining at what exposure conditions the new batch of experiments should be performed. The DOE optimization fundamentally relies on the statistical modeling/ inference for the target response surface, which is provided by the developed SKQ methodology. In this dissertation, a DOE method will be developed based on the SKQ modeling/inference, and thus will inherit all the advantages of SKQ.

This research will lead to statistical methods for the efficient characterization of a substance by its high-quality exposure-response profile. These methods will substantially reduce the experimental cost and time in toxicology studies required, alleviate the rising concerns for animal ethics [25, 26], and accelerate the progress toward quantifying the risk, safety and health effects of environmental and occupational exposure to nanomaterials.

1.3 Contribution of the Research

SKQ, to the best of our knowledge, is the first attempt to develop a metamodel that can accommodate both quantitative variables and qualitative variables, can have very general variance structure, and can provide valid statistical inference. The metamodel fully use information across different sources, and allows for estimation of exposure-response surface for all sources at the same time.

Utilizing the ability of SKQ to pool information across different sources and accurately predict response surface, our DOE is able to find the experiment conditions with desirable property sequentially for a new source, given the data collected from other sources. This research work provides the first DOE frame that is able to provide design procedure for experimental data arising from multiple sources, with non-normality and variance heterogeneity, and mapping nonlinear exposure-response relationships.

1.4 Organization of the Dissertation

The remainder of this dissertation is organized as follows. Chapter 2 gives a review of the related literature. In Chapter 3, the SKQ (stochastic kriging with qualitative factors) model

is detailed and its advantages over the existing models are demonstrated through simulation studies. Chapter 4 describes the two-stage DOE procedure, and evaluates the efficiency of the DOE. Chapter 5 finishes this dissertation with Summary.

Chapter 2

Problem Statement and Literature Review

In this chapter, we describe the research problem in precise terms, and review the related literature.

2.1 Statement of the Research Problem

In exposure-response studies for a substance, biological experiments are performed in a range of experimental conditions. An experimental condition is defined by the combination of a number of factors, which can be divided into two categories, quantitative and qualitative factors.

- Quantitative factors typically include but are not limited to the toxicant dosage administered to an animal, and the post-exposure time (or the exposure duration of the toxicant). In this paper, the vector \mathbf{x} is used to represent the quantitative factors considered.
- Qualitative factors mainly include various source factors such as the laboratory where the experiments were conducted, the provider for the experimental animals, and the shape of nanomaterials. The qualitative factors are denoted by the vector \mathbf{z} .

The experimental condition is specified in terms of the factor vector $\mathbf{w} = (\mathbf{x}^\top, \mathbf{z}^\top)^\top$. The random response observed from an animal subject at a factor setting \mathbf{w} can be generally written as

$$\mathcal{Y}(\mathbf{w}) = E[\mathcal{Y}(\mathbf{w})] + \varepsilon(\mathbf{w}) = Y(\mathbf{w}) + \varepsilon(\mathbf{w}), \quad (2.1)$$

where $Y(\mathbf{w}) = E[\mathcal{Y}(\mathbf{w})]$ represents the true expected response, and $\varepsilon(\mathbf{w})$ the random zero-mean error that accounts for the variations across animal subjects.

An example of such exposure-response studies is given later in the subsection 3.5.1, where the toxicity of TiO₂ nanoparticles (NPs) is investigated. In that case, the vector $\mathbf{x} = (x_1, x_2)^\top$ includes two quantitative factors, the dosage of NPs x_1 and the post-exposure time x_2 . There is one qualitative factor z , which has two category levels for the shape of TiO₂ NPs: short and long nanobelts.

A setting of the qualitative factors \mathbf{z} corresponds to a combination category, say c_q , and defines a subpopulation or a data source. The total of Q subpopulations specified by the settings of \mathbf{z} are denoted as $\{c_q; q = 1, 2, \dots, Q\}$. In the case of TiO₂ NPs, there are $Q = 2$ subpopulations: c_1 for short TiO₂ nanobelts, and c_2 for long TiO₂ nanobelts. The population is the union of all the subpopulations, and is considered as the TiO₂ NPs of both shapes in the aforementioned example. For a given subpopulation c_q , the bioactivity response obtained from an animal subject is expected to be $Y(\mathbf{w}|c_q) = Y(\mathbf{x}, c_q)$, a continuous function of the quantitative factors \mathbf{x} , while subjecting to the cross-subject random error $\varepsilon(\mathbf{w}|c_q) = \varepsilon(\mathbf{x}, c_q)$.

The biological data collected for the toxicity study of a substance are represented as

$$\{(\mathbf{w}_i, \mathcal{Y}_j(\mathbf{w}_i)); i = 1, 2, \dots, I; j = 1, 2, \dots, n(\mathbf{w}_i)\} \quad (2.2)$$

where \mathbf{w}_i denotes the i^{th} design point (factor setting at which experiments are performed) of a total of I distinct design points, $\mathcal{Y}_j(\mathbf{w}_i)$ the observed response from the j^{th} replication at \mathbf{w}_i , and $n(\mathbf{w}_i)$ the number of replications at \mathbf{w}_i .

Based on the sample data (2.2), the objective of statistical modeling is to quantify the dependence of the response upon the quantitative and qualitative factors, and to provide valid statistical inference regarding the population being investigated (e.g., the TiO₂ NPs

of both shapes). SKQ aims at achieving this objective with high data efficiency and model generality.

2.2 The Naive Method without Pooling

The most straightforward method is to model the continuous exposure-response relationship for each subpopulation separately. Since biological data are typically collected for one subpopulation at a time, such separate modeling is most commonly performed (e.g., [27]) with no effort to pool information across different subpopulations.

The biggest drawback of separate modeling is the large sample size it takes to obtain exposure-response models of high quality: For each subpopulation, sufficient sample data has to be collected in the feasible region of \mathbf{x} for an accurate capture of the target response surface.

2.3 An Information-Pooling Model: Mixed-Effects Model (MEM)

Recognizing the importance of pooling information for modeling efficiency, the mixed-effects model has been developed and used for multi-source exposure-response data in the literature. Interested readers can refer to Davidian et al. [28] for details. In this part, a brief review is provided for the readers' convenience.

2.3.1 Formulation and Assumptions for Mixed-Effects Model (MEM)

Following the notations in Section 2.1, for a subpopulation c_q ($q = 1, 2, \dots, Q$), the response from the j th replication (animal subject) is written as

$$\mathcal{Y}_j(\mathbf{x}, c_q) = g(\mathbf{x}, \boldsymbol{\alpha}_{c_q}) + \varepsilon_j(\mathbf{x}, c_q); \quad q = 1, 2, \dots, Q, \quad (2.3)$$

In Equation (2.3), g is a regression model of a prior-assumed functional form (e.g., logistic model), which is common to all subpopulations $\{c_q; q = 1, 2, \dots, Q\}$; $\boldsymbol{\alpha}_{c_q}$ is a $u \times 1$ vector of unknown parameters, the values of which are c_q -dependent.

At a factor setting (\mathbf{x}, c_q) , the random noise $\varepsilon_1(\mathbf{x}, c_q), \varepsilon_2(\mathbf{x}, c_q), \dots$ has zero mean, and is independent and identically distributed across replications (animal subjects). The error variance $\text{Var}[\varepsilon(\mathbf{x}, c_q)]$ is assumed to be response-dependent through the following common form for any c_q :

$$\text{Var}[\varepsilon(\mathbf{x}, c_q)] = \sigma^2 h^2(g(x, \boldsymbol{\alpha}_{c_q}), \boldsymbol{\gamma}); \quad q = 1, 2, \dots, Q \quad (2.4)$$

The variance function h describes the common pattern of variability. The scalar σ and vector $\boldsymbol{\gamma}$ are unknown parameters, which take common values for any c_q .

In MEM, the unknown model coefficient vector $\boldsymbol{\alpha}_{c_q}$ is generally modeled as

$$\boldsymbol{\alpha}_{c_q} = \mathbf{d}(\boldsymbol{\nu}, \boldsymbol{\alpha}) + \mathbf{b}_{c_q}, \quad (2.5)$$

where \mathbf{d} is a u -dimensional vector-valued function with each component associated with the corresponding component of $\boldsymbol{\alpha}_{c_q}$. The vector $\boldsymbol{\nu}$ includes the factors (or covariates) for subpopulation attributes that affect the parameter $\boldsymbol{\alpha}_{c_q}$ for the subpopulation c_q . In the problem setting of Section 2.1, all the qualitative factors in \mathbf{z} are candidate components for $\boldsymbol{\nu}$. The vector $\boldsymbol{\alpha}$ denotes the unknown fixed parameters (or fixed effects), and \mathbf{b}_{c_q} is a random vector representing the random effects. It is assumed that

$$\mathbf{b}_{c_q} \sim \text{Norm}(\mathbf{0}, \mathbf{D}) \quad (2.6)$$

with \mathbf{D} being an unknown variance-covariance matrix of compatible dimensions.

The simplest example form of (2.5) is

$$\boldsymbol{\alpha}_{c_q} = \boldsymbol{\alpha} + \mathbf{b}_{c_q}, \quad (2.7)$$

which is commonly used.

It has been demonstrated that MEM is able to provide improved estimation/inference for multi-source data, by pooling information across different data subsets [29]. Nevertheless, the following shortcomings exist, which originate from the model formulation given above. (i) A common nonlinear functional form (e.g., logistic model) has to be assumed adequate for describing the exposure-response surface of all subpopulations. Such a common form may be difficult to obtain, especially when \mathbf{x} is high-dimensional. (ii) MEM assumes a common variance model for all the subpopulations. That is, for each subpopulation, the same variance model as in (2.4) is used to describe the dependence of variance upon the response. (iii) As given in (2.6), one of the fundamental assumptions made in MEM is: The unknown coefficient vector $\boldsymbol{\alpha}_{c_q}$ for a subpopulation follows a multivariate normal distribution, which may well not hold.

2.3.2 Estimation and Inference by Mixed-Effects Model (MEM)

To estimate MEM from given data, the global two-stage (GTS) method proposed by Steimer et al. [28–30] can be used. The fitted MEM models are denoted as follows: $\{g(\mathbf{x}, \hat{\boldsymbol{\alpha}}_{c_q}); q = 1, 2, \dots, Q\}$, the fitted response surface models; $\{\hat{\sigma}h(g(\mathbf{x}, \hat{\boldsymbol{\alpha}}_{c_q}), \hat{\gamma}); q = 1, 2, \dots, Q\}$, the estimated variance model; and $\hat{\mathbf{D}}$, the estimated covariance matrix of $\hat{\boldsymbol{\alpha}}_{c_q}$.

For an arbitrary setting $\mathbf{w}_0 = (\mathbf{x}_0, c_q)$, the expected response is estimated as

$$\hat{Y}(\mathbf{x}_0, c_q) = g(\mathbf{x}_0, \hat{\boldsymbol{\alpha}}_{c_q}).$$

The variance of $\widehat{Y}(\mathbf{x}_0, c_q)$ is estimated as:

$$\widehat{\text{Var}}[\widehat{Y}(\mathbf{x}_0, c_q)] = (\mathbf{g}\boldsymbol{\alpha}(\mathbf{x}_0, \widehat{\boldsymbol{\alpha}}_{c_q}))^\top \widehat{\mathbf{D}} g\boldsymbol{\alpha}(\mathbf{x}_0, \widehat{\boldsymbol{\alpha}}_{c_q}),$$

where $\mathbf{g}\boldsymbol{\alpha}(\mathbf{x}_0, \widehat{\boldsymbol{\alpha}}_{c_q})$ is the $u \times 1$ first derivative vector of function g w.r.t. $\boldsymbol{\alpha}_{c_q}$.

The $(1 - \alpha)100\%$ confidence interval for $\widehat{Y}(\mathbf{x}_0, c_q)$ is

$$\widehat{Y}(\mathbf{x}_0, c_q) \pm t_{1-\alpha/2, v} \times \sqrt{\widehat{\text{Var}}[\widehat{Y}(\mathbf{x}_0, c_q)]} \quad (2.8)$$

where t denotes Student's t distribution with degree of freedom v . Following the context of BMD estimation in Subsection 3.4, we consider the cases where there is only one quantitative factor x , denoting the dose level. Using the fitted MEM, the BMD associated with a subpopulation c_q is estimated as follows for a given BMR:

$$\widehat{\text{BMD}}(c_q) = g^{-1}(\text{BMR}, \widehat{\boldsymbol{\alpha}}_{c_q}). \quad (2.9)$$

For MEM, the analytic form of the inverse function g^{-1} can be easily obtained.

The variance of $\widehat{\text{BMD}}(c_q)$ is estimated as:

$$\widehat{\text{Var}}[\widehat{\text{BMD}}(c_q)] = (g_{\boldsymbol{\alpha}}^{-1})^\top \widehat{\mathbf{D}} g_{\boldsymbol{\alpha}}^{-1} \quad (2.10)$$

where $g_{\boldsymbol{\alpha}}^{-1}$ is the $u \times 1$ first derivative vector of the inverse function g^{-1} w.r.t. $\boldsymbol{\alpha}_{c_q}$. The one-sided $100(1 - \alpha)$ confidence interval for $\widehat{\text{BMD}}(c_q)$ is given as:

$$[\widehat{\text{BMD}}(c_q) - t_{1-\alpha, v} \sqrt{\widehat{\text{Var}}[\widehat{\text{BMD}}(c_q)]}, \infty) \quad (2.11)$$

where t is student's t distribution with the degree of freedom v .

2.4 Design of Experiments for Learning/Inferences of Gaussian Processes

We briefly review the design of experiments (DOE) methods for kriging.

There is a substantial literature on the DOE for deterministic kriging [31], [32],[33], which can be divided into two categories: the model-free and model-based methods. The model-free DOE includes a range of Latin Hypercube designs such as Qian [34], Sallaberry et al. [35], and Williamson [36]. These designs are space-filling in nature, and provide a good coverage of the feasible design space. The model-based DOE seeks to determine the design by optimizing some estimation/prediction criteria, such as integrated mean square error(IMSE) and maximum mean squared error (MMSE) [37, 38]. In the literature, a batch sequential framework has also been developed [39], which is initiated with a space-filling design and includes follow-up stages of model-based designs.

There is some research effort to investigate model-based sequential designs for stochastic kriging. Ankenman et al. [23] provides a two-stage strategy that minimizes the IMSE criterion. In the initial stage, pilot experiments are performed following some space-filling design; and in a follow-up stage, the augmented design is determined by minimizing the IMSE provided by the stochastic kriging. Chen et al. [40] developed a sequential DOE procedure, which is adaptive to a variety of design criteria. The difficulty of designing experiments for stochastic kriging lies in determining the design-point locations as well as the the number of replications assigned to each point.

Chapter 3

Stochastic Kriging With Qualitative Factors (SKQ)

In this part, the SKQ metamodeling and inference methods are detailed. As an extension from the standard stochastic kriging (SK), which considers quantitative factors only, SKQ models the variability arising from quantitative as well as qualitative factors. For readers' convenience, a review of SK is given in the Appendix 6.1.

Compared to existing kriging-based methods [23, 41, 42], SKQ is the first one that models all the following stochastic elements: the extrinsic variability caused by both quantitative and qualitative factors, and the intrinsic variability across random replications, as will become clearer in the remainder of this section.

SKQ models the dependence of continuous responses upon the factors $\mathbf{w} = (\mathbf{x}^\top, \mathbf{z}^\top)^\top$, with $\mathbf{x} = (x_1, x_2, \dots, x_d)^\top \in \mathbb{R}^d$ and $\mathbf{z} = (z_1, z_2, \dots, z_L)^\top$ including L qualitative factors. Each qualitative factor z_ℓ has a number of category levels.

The response at setting \mathbf{w} for the j th replication (animal subject) is modeled by SKQ as

$$\mathcal{Y}_j(\mathbf{w}) = Y(\mathbf{w}) + \varepsilon_j(\mathbf{w}) = \mathbf{f}(\mathbf{w})^\top \boldsymbol{\beta} + M(\mathbf{w}) + \varepsilon_j(\mathbf{w}), \quad (3.1)$$

The expectation $Y(\mathbf{w})$ is decomposed into the sum of two parts: $\mathbf{f}(\mathbf{w})^\top \boldsymbol{\beta}$ and $M(\mathbf{w})$. Here, $\mathbf{f}(\mathbf{w})$ is a vector of known functions of \mathbf{w} , and $\boldsymbol{\beta}$ is a vector of unknown parameters of compatible dimension. Since it has been widely accepted that $\mathbf{f}(\mathbf{w})^\top \boldsymbol{\beta} = \beta_0$ (that is, just a constant term) suffices for most applications [23], this work adopts $\mathbf{f}(\mathbf{w})^\top \boldsymbol{\beta} = \beta_0$ unless stated otherwise. The term $M(\mathbf{w})$ represents a mean-zero stationary Gaussian process, which intends to capture the extrinsic variability, i.e., the variability due to the factors \mathbf{w} .

The randomness of $\varepsilon(\mathbf{w})$ is referred to as intrinsic variability. The random noise $\varepsilon_1(\mathbf{w}), \varepsilon_2(\mathbf{w}), \dots$ at a factor setting \mathbf{w} is assumed to have mean zero, and be independent and identically distributed (i.i.d.) across replications. The error variance $\text{Var}[\varepsilon(\mathbf{w})]$ is allowed to be dependent on \mathbf{w} .

Given the sample data (2.2), the sample average of the responses at \mathbf{w}_i across the $n(\mathbf{w}_i)$ replications follows as:

$$\bar{\mathcal{Y}}(\mathbf{w}_i) = \frac{1}{n(\mathbf{w}_i)} \sum_{j=1}^{n(\mathbf{w}_i)} \mathcal{Y}_j(\mathbf{w}_i) = \beta_0 + \mathbf{M}(\mathbf{w}_i) + \frac{1}{n(\mathbf{w}_i)} \sum_{j=1}^{n(\mathbf{w}_i)} \varepsilon_j(\mathbf{w}_i);$$

Denote

$$\bar{\mathcal{Y}} = (\bar{\mathcal{Y}}(\mathbf{w}_1), \bar{\mathcal{Y}}(\mathbf{w}_2), \dots, \bar{\mathcal{Y}}(\mathbf{w}_I))^\top \quad (3.2)$$

as the $I \times 1$ vector of sample average responses at the I distinct design points.

Similarly, the vector of sample average errors is denoted as

$$\bar{\varepsilon} = (\bar{\varepsilon}(\mathbf{w}_1), \bar{\varepsilon}(\mathbf{w}_2), \dots, \bar{\varepsilon}(\mathbf{w}_I))^\top, \quad (3.3)$$

with $\bar{\varepsilon}(\mathbf{w}_i) = n(\mathbf{w}_i)^{-1} \sum_{j=1}^{n(\mathbf{w}_i)} \varepsilon_j(\mathbf{w}_i), i = 1, 2, \dots, I$.

3.1 Extrinsic Variance Structure

The modeling of the extrinsic variability is performed following the framework proposed by Qian et al. [41]. For the factor settings $\mathbf{w} = (\mathbf{x}, \mathbf{z})$ and $\mathbf{w}' = (\mathbf{x}', \mathbf{z}')$, the covariance of the stationary Gaussian process $\mathbf{M}(\cdot)$ takes the form

$$\text{Cov}[\mathbf{M}(\mathbf{w}), \mathbf{M}(\mathbf{w}')] = \delta^2 \cdot \text{Corr}[\mathbf{M}(\mathbf{w}), \mathbf{M}(\mathbf{w}')] = \delta^2 \cdot \left[\prod_{\ell=1}^L \tau_{z_\ell, z'_\ell}^{(\ell)} \right] \cdot K(\mathbf{x}, \mathbf{x}'), \quad (3.4)$$

with δ^2 being the variance of the Gaussian process. The correlation $\text{Corr}[\mathbf{M}(\mathbf{w}), \mathbf{M}(\mathbf{w}')] is decomposed as the product of two parts: $\prod_{\ell=1}^L \tau_{z_\ell, z'_\ell}^{(\ell)}$ and $K(\mathbf{x}, \mathbf{x}')$. To enable the estimation of a SKQ model, specific functional forms need to be assumed for both parts.$

In (3.4), $K(\mathbf{x}, \mathbf{x}')$ represents the correlation across the quantitative settings for a given combination category of the quantitative factors \mathbf{x} , and models the variability due to quantitative factors. Hence, $K(\mathbf{x}, \mathbf{x}')$ plays the same role in SKQ as in SK, and the discussions regarding $K(\mathbf{x}, \mathbf{x}')$ in the Appendix 6.1 can be inherited here: For specific functional structures of $K(\mathbf{x}, \mathbf{x}')$, a range of choices are available in the literature (e.g., Santner et al. [43]; Qian et al. [41]), and one of the most popular correlation functions in practice is the exponential correlation function

$$K(\mathbf{x}, \mathbf{x}') = \exp \left\{ \sum_{h=1}^d -\theta_h |x_h - x'_h|^p \right\}. \quad (3.5)$$

In (3.5), $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ is a vector of unknown parameters. It is required that $\theta_h > 0$ ($h = 1, 2, \dots, d$), and $\boldsymbol{\theta}$ determines the roughness of the response surface for a given combination category of \mathbf{z} . The parameter $p \in (0, 2]$ also needs to be estimated unless p is pre-specified as 2, which corresponds to the widely used quadratic correlation function [44].

The term $\prod_{\ell=1}^L \tau_{z_\ell, z'_\ell}^{(\ell)}$ in (3.4) is devoted to the correlations across different levels of qualitative factors. As noted in Qian et al. [41], $\tau_{z_\ell, z'_\ell}^{(\ell)}$ measures the correlation (similarity) at any two settings \mathbf{w} and \mathbf{w}' that differ only on the values of the ℓ th qualitative factor. For $\tau_{z_\ell, z'_\ell}^{(\ell)}$, a range of functional forms have been proposed in Qian et al. [41] and Zhou et al. [45]. Below, two specific correlation functions are given as examples.

- Isotropic (or exchangeable) correlation functions (**EC**):

$$\tau_{z_\ell, z'_\ell}^{(\ell)} = \exp\{-\phi^{(\ell)} I(z_\ell \neq z'_\ell)\}; \quad \ell = 1, 2, \dots, L \quad (3.6)$$

In (3.6), $\Phi = \{\phi^{(\ell)}; \ell = 1, 2, \dots, L\}$ represents the set of unknown parameters to be estimated; and $I[A]$ is an indicator function that takes 1 if event A is true and 0 otherwise. Clearly, EC assumes that all the category levels of the ℓ th qualitative factor are of isotropic nature; that is, for a given ℓ , $\tau_{z_\ell, z'_\ell}^{(\ell)}$ is a constant as long as $z_\ell \neq z'_\ell$.

- Multiplicative correlation functions (**MC**):

$$\tau_{z_\ell, z'_\ell}^{(\ell)} = \exp \left\{ - (\phi_{z_\ell}^{(\ell)} + \phi_{z'_\ell}^{(\ell)}) I(z_\ell \neq z'_\ell) \right\} \quad (3.7)$$

The unknown parameter set Φ includes the following components:

$$\phi_{c_\ell}^{(\ell)}; \quad \ell = 1, 2, \dots, L;$$

c_ℓ denotes any one of all the possible category levels for the ℓ th qualitative factor.

For a given ℓ , MC allows the correlation $\tau_{z_\ell, z'_\ell}^{(\ell)}$ to be dependent on the category levels involved (i.e., z_ℓ and z'_ℓ).

Given the data $\{(\mathbf{w}_i, \mathcal{Y}_j(\mathbf{w}_i)); i = 1, 2, \dots, I; j = 1, 2, \dots, n(\mathbf{w}_i)\}$ collected at I distinct design points, the $I \times I$ variance-covariance matrix Σ_M is constructed as

$$\Sigma_M = \delta^2 \mathbf{R}(\boldsymbol{\theta}, \Phi) = \delta^2 \begin{pmatrix} 1 & \text{Corr}[\mathbf{M}(\mathbf{w}_1), \mathbf{M}(\mathbf{w}_2)] & \dots & \text{Corr}[\mathbf{M}(\mathbf{w}_1), \mathbf{M}(\mathbf{w}_I)] \\ \text{Corr}[\mathbf{M}(\mathbf{w}_2), \mathbf{M}(\mathbf{w}_1)] & 1 & \dots & \text{Corr}[\mathbf{M}(\mathbf{w}_2), \mathbf{M}(\mathbf{w}_I)] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Corr}[\mathbf{M}(\mathbf{w}_I), \mathbf{M}(\mathbf{w}_1)] & \text{Corr}[\mathbf{M}(\mathbf{w}_I), \mathbf{M}(\mathbf{w}_2)] & \dots & 1 \end{pmatrix}. \quad (3.8)$$

In (3.8), $\mathbf{R}(\boldsymbol{\theta}, \Phi)$ denotes the correlation matrix; each component of the matrix represents a correlation, which can be decomposed into two parts as explained above and which is a

function of $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$. For an arbitrary setting \mathbf{w}_0 , the $I \times 1$ vector $\Sigma_{\mathbf{M}}(\mathbf{w}_0, \cdot)$ is defined as

$$\Sigma_{\mathbf{M}}(\mathbf{w}_0, \cdot) = \delta^2 \mathbf{v}(\mathbf{w}_0, \boldsymbol{\theta}, \boldsymbol{\Phi}) = \delta^2 \begin{pmatrix} \text{Corr}[\mathbf{M}(\mathbf{w}_0), \mathbf{M}(\mathbf{w}_1)] \\ \text{Corr}[\mathbf{M}(\mathbf{w}_0), \mathbf{M}(\mathbf{w}_2)] \\ \vdots \\ \text{Corr}[\mathbf{M}(\mathbf{w}_0), \mathbf{M}(\mathbf{w}_I)] \end{pmatrix}, \quad (3.9)$$

where $\mathbf{v}(\mathbf{w}_0, \boldsymbol{\theta}, \boldsymbol{\Phi})$ denotes the correlation vector with each component being a correlation function dependent on \mathbf{w}_0 , and the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$.

3.2 Intrinsic Variance Structure

The intrinsic variance of the random response at \mathbf{w} is denoted as $\text{Var}[\varepsilon(\mathbf{w})]$, which is dependent on the setting \mathbf{w} . Let Σ_{ε} be the $I \times I$ variance-covariance matrix of vector $\bar{\boldsymbol{\varepsilon}}$, which is defined in (3.3). Under the i.i.d. assumption for random errors, Σ_{ε} is a $I \times I$ diagonal matrix

$$\Sigma_{\varepsilon} = \text{diag}\{\text{Var}[\varepsilon(\mathbf{w}_1)]/n(\mathbf{w}_1), \text{Var}[\varepsilon(\mathbf{w}_2)]/n(\mathbf{w}_2), \dots, \text{Var}[\varepsilon(\mathbf{w}_I)]/n(\mathbf{w}_I)\}. \quad (3.10)$$

3.3 Estimation and Inference by SKQ (Stochastic Kriging with Qualitative Factors)

The SKQ-based estimation and inference requires the following assumption, which parallels Assumption 2 stipulated for SK (see the Appendix 6.1).

Assumption 1 *The random field \mathbf{M} is a stationary Gaussian random field; and $\varepsilon_1(\mathbf{w}), \varepsilon_2(\mathbf{w}), \dots$ are i.i.d. $N(0, \text{Var}[\varepsilon(\mathbf{w})])$, independent of $\varepsilon_j(\mathbf{w}')$ for all j and $\mathbf{w} \neq \mathbf{w}'$, and independent of \mathbf{M} .*

Given a data set $\{(\mathbf{w}_i, \mathcal{Y}_j(\mathbf{w}_i)); i = 1, 2, \dots, I; j = 1, 2, \dots, n(\mathbf{w}_i)\}$ and under Assumption 1, $\bar{\mathcal{Y}} = (\bar{\mathcal{Y}}(\mathbf{w}_1), \bar{\mathcal{Y}}(\mathbf{w}_2), \dots, \bar{\mathcal{Y}}(\mathbf{w}_I))^{\top}$ as defined in (3.2) follows a multivariate normal dis-

tribution with constant mean vector $\beta_0 \times \mathbf{1}_I$, where $\mathbf{1}_I$ is a $(I \times 1)$ vector of ones, and variance-covariance matrix

$$\Sigma(\delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi}) = \Sigma_{\mathbf{M}} + \Sigma_{\varepsilon} = \delta^2 \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\Phi}) + \Sigma_{\varepsilon}. \quad (3.11)$$

Recall that $\Sigma_{\mathbf{M}}$ and Σ_{ε} are defined in (3.8) and (3.10) respectively. Thus, the log-likelihood function of $\bar{\mathcal{Y}}$ in terms of the unknown parameters $(\beta_0, \delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi})$ can be written as

$$\ln \mathcal{L}(\beta_0, \delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi}) = -\ln[(2\pi)^{I/2}] - \frac{1}{2} \ln[|\delta^2 \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\Phi}) + \Sigma_{\varepsilon}|] - \frac{1}{2} (\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_I)^\top [\delta^2 \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\Phi}) + \Sigma_{\varepsilon}]^{-1} (\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_I). \quad (3.12)$$

The procedure to obtain the parameter estimates for $(\beta_0, \delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi})$ that maximize (3.12) is given as follows.

1. Obtain the estimated Σ_{ε} :

$$\widehat{\Sigma}_{\varepsilon} = \text{diag}\{\widehat{\text{Var}}[\varepsilon(\mathbf{w}_1)]/n(\mathbf{w}_1), \widehat{\text{Var}}[\varepsilon(\mathbf{w}_2)]/n(\mathbf{w}_2), \dots, \widehat{\text{Var}}[\varepsilon(\mathbf{w}_I)]/n(\mathbf{w}_I)\} \quad (3.13)$$

where

$$\widehat{\text{Var}}[\varepsilon(\mathbf{w}_i)] = \frac{1}{n(\mathbf{w}_i) - 1} \sum_{j=1}^{n(\mathbf{w}_i)} (\mathcal{Y}_j(\mathbf{w}_i) - \bar{\mathcal{Y}}(\mathbf{w}_i))^2, \quad i = 1, 2, \dots, I. \quad (3.14)$$

2. Replace Σ_{ε} by $\widehat{\Sigma}_{\varepsilon}$, and maximize the log-likelihood function (3.12) with respect to (w.r.t.) $(\beta_0, \delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi})$. Specifically, two steps can be taken to solve the maximum likelihood problem. (i) Given δ^2 , $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$, the maximum likelihood estimate (MLE) of β_0 is

$$\widehat{\beta}_0(\delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi}) = (\mathbf{1}_I^\top [\delta^2 \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\Phi}) + \widehat{\Sigma}_{\varepsilon}]^{-1} \mathbf{1}_I)^{-1} \mathbf{1}_I^\top [\delta^2 \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\Phi}) + \widehat{\Sigma}_{\varepsilon}]^{-1} \bar{\mathcal{Y}}.$$

(ii) Substituting $\widehat{\beta}_0(\delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi})$ into (3.12), the problem reduces to maximizing

$$\begin{aligned} \ln \mathcal{L}(\delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi}) &= -\ln[(2\pi)^{I/2}] - \frac{1}{2} \ln[|\delta^2 \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\Phi}) + \widehat{\Sigma}_\varepsilon|] \\ &\quad - \frac{1}{2} (\bar{\mathcal{Y}} - \widehat{\beta}_0(\delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi}) \mathbf{1}_I)^\top [\delta^2 \mathbf{R}(\boldsymbol{\theta}, \boldsymbol{\Phi}) + \widehat{\Sigma}_\varepsilon]^{-1} (\bar{\mathcal{Y}} - \widehat{\beta}_0(\delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi}) \mathbf{1}_I), \end{aligned} \quad (3.15)$$

w.r.t. $(\delta^2, \boldsymbol{\theta}, \boldsymbol{\Phi})$, which can be solved by a nonlinear optimization algorithm such as the Matlab `fmincon` function.

3. For an arbitrary setting \mathbf{w}_0 , estimate the expected response $\mathbf{Y}(\mathbf{w}_0)$ by

$$\widehat{\mathbf{Y}}(\mathbf{w}_0) = \widehat{\beta}_0 + \mathbf{v}(\mathbf{w}_0, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}})^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}}) + \widehat{\Sigma}_\varepsilon]^{-1} (\bar{\mathcal{Y}} - \widehat{\beta}_0 \mathbf{1}_I), \quad (3.16)$$

where $(\widehat{\beta}_0, \widehat{\delta}^2, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}})$ are the maximum likelihood estimates obtained from the previous step. Recall that $\mathbf{v}(\mathbf{w}_0, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}})$ is defined in (3.9). Following the proof scheme of Ankenman et al. [23], it can be shown that (3.16) is the unbiased estimator for $\mathbf{Y}(\mathbf{w}_0)$.

The mean squared error (MSE) is obtained as:

$$\widehat{\text{MSE}}[\widehat{\mathbf{Y}}(\mathbf{w}_0)] = \widehat{\delta}^2 - \widehat{\delta}^4 \mathbf{v}(\mathbf{w}_0, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}})^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}}) + \widehat{\Sigma}_\varepsilon]^{-1} \mathbf{v}(\mathbf{w}_0, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}}) + \eta^2 (\mathbf{1}_I^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}}) + \widehat{\Sigma}_\varepsilon]^{-1} \mathbf{1}_I)^{-1}, \quad (3.17)$$

where $\eta = 1 - \mathbf{1}_I^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}}) + \widehat{\Sigma}_\varepsilon]^{-1} \mathbf{v}(\mathbf{x}_0, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Phi}}) \widehat{\delta}^2$.

The two-sided $100(1-\alpha)\%$ confidence interval for $\mathbf{Y}(\mathbf{w}_0)$ can be constructed as:

$$\widehat{\mathbf{Y}}(\mathbf{w}_0) \pm z_{1-\alpha/2} \sqrt{\widehat{\text{MSE}}[\widehat{\mathbf{Y}}(\mathbf{w}_0)]} \quad (3.18)$$

where $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)^{th}$ percentile of standard normal distribution.

3.4 Inverse Estimation and Inference

In this part, we consider the SKQ-based inverse estimation/inference for a special stream of cases, which are of particular interest in toxicology studies. For such cases, attention is focused on modeling two-dimensional dose-response data, and the subsequent derivation of the BMD (benchmark dose) for the substance of interest. The BMD is the dose that corresponds to a specified level of adverse response called the benchmark response (BMR), and plays an important role in setting safety standard.

Following the notation adopted earlier, multi-source dose-response data are collected at factor setting $\mathbf{w} = (x, \mathbf{z})$, which includes one quantitative factor x representing the dose level, and a number of qualitative factors \mathbf{z} with all the possible combination categories being $\{c_q; q = 1, 2, \dots, Q\}$. The expected dose-response curve is denoted as $Y(x, c_q)$ for a subpopulation specified by c_q .

The BMR can be defined as a relative change in the mean response from the control mean, or as an absolute level [19, 46]. Either definition can be selected based on the knowledge available regarding the substance's adverse effects, and the BMR defined in one way can be easily converted to that defined in the other. For illustration, we let the BMR be a pre-selected absolute response in this work, and the BMD for a subpopulation c_q is written as

$$\text{BMD}(c_q) = Y^{-1}(\text{BMR}, c_q). \quad (3.19)$$

Here, we abuse the notation a little and use Y^{-1} to represent the functional dependence of $\text{BMD}(c_q)$ upon BMR, assuming that the inverse mapping exists.

The collection of dose-response curves $\{Y(x, c_q); q = 1, 2, \dots, Q\}$ are modeled by SKQ. To perform the inverse calculation as given in (3.19), numerical interpolation needs to be employed based on the fitted SKQ model. In this work, the cubic spline interpolation recommended by Hastie et al. [47] is used to perform the inverse computation for BMD estimation.

Denote $\widehat{\text{BMD}}(c_q)$ as the estimated BMD from the SKQ model for the subpopulation c_q . The uncertainty of $\widehat{\text{BMD}}(c_q)$, represented by $\text{Var}[\widehat{\text{BMD}}(c_q)]$, directly relates to the safety standard of the substance being investigated. To estimate $\text{Var}[\widehat{\text{BMD}}(c_q)]$, numerical methods have to be employed again, because explicit formula is not available. The bootstrap resampling method developed by Kirk et al. [48] is adapted to quantify the uncertainty of $\{\widehat{\text{BMD}}(c_q); q = 1, 2, \dots, Q\}$ based on the SKQ modeling of given data $\{(\mathbf{w}_i, \mathcal{Y}_j(\mathbf{w}_i)); i = 1, 2, \dots, I; j = 1, 2, \dots, n(\mathbf{w}_i)\}$. The adapted bootstrapping algorithm is described as follows.

Bootstrap Resampling Algorithm

Input: (a) The I design points $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_I\}$, at which real data have been collected; (b) The sample variance $\widehat{\text{Var}}[\varepsilon(\mathbf{w}_i)]; i = 1, 2, \dots, I$ obtained as in (3.14) from the real sample data; (c) the fitted SKQ model $\widehat{\mathbf{Y}}(\mathbf{w})$ and $\widehat{\text{MSE}}[\widehat{\mathbf{Y}}(\mathbf{w})]$ as given in (3.16) and (3.17), both obtained from the real data; (d) The bootstrap sample size B .

Do for $b = 1, 2, \dots, B$

(i) Do for $i = 1, 2, \dots, I$

- Set $\mathbf{w}_{i,b}^* = \mathbf{w}_i$
- Do for $j = 1, 2, \dots, n(\mathbf{w}_i)$
 - Randomly sample $e_{i,j,b}^*$ from the normal distribution $N(0, \text{MSE}(\widehat{\mathbf{Y}}(\mathbf{w}_{i,b}^*)) + \widehat{\text{Var}}[\varepsilon(\mathbf{w}_{i,b}^*)])$
 - Set $\mathcal{Y}_j^*(\mathbf{w}_{i,b}^*) = \widehat{\mathbf{Y}}(\mathbf{w}_{i,b}^*) + e_{i,j,b}^*$

(ii) Based on the bootstrap data

$$\{(\mathbf{w}_{i,b}^*, \mathcal{Y}_j^*(\mathbf{w}_{i,b}^*)); i = 1, 2, \dots, I; j = 1, 2, \dots, n(\mathbf{w}_{i,b}^*)\},$$

build the SKQ model and denote the resulting model as $\widehat{\mathbf{Y}}_b^*(\mathbf{w})$

Output B SKQ models $\{\widehat{\mathbf{Y}}_b^*(\mathbf{w}); b = 1, 2, \dots, B\}$, respectively fitted from the B bootstrap sample data sets

Figure 3.1: The bootstrap resampling algorithm for uncertainty quantification of BMD estimates.

Empirical studies suggest that $B = 999$ typically suffices as a bootstrap sample size for the construction of confidence intervals (CIs). With the B fitted SKQ metamodels $\{\widehat{Y}_b^*(\mathbf{w}); b = 1, 2, \dots, B\}$, B BMD estimates can be obtained for a specified BMR and a given subpopulation c_q :

$$\{\widehat{\text{BMD}}_b^*(c_q); b = 1, 2, \dots, B\} \quad (3.20)$$

Based on the bootstrap estimates given in (3.20), the nonparametric method suggested by Davison and Hinkley [49] can be easily employed to estimate the $100\alpha^{\text{th}}$ ($\alpha \in (0, 1)$) percentile of $\widehat{\text{BMD}}(c_q)$. In this context, the percentile estimate is referred to as BMDL, which serves as the lower bound of the one-sided $100\alpha\%$ CI for the BMD. For the subpopulation c_q , the resulting one-sided CI of $\widehat{\text{BMD}}(c_q)$ can be written as $[\widehat{\text{BMDL}}(c_q), \infty)$.

3.5 Simulation Studies

Empirical case studies were designed and performed to demonstrate SKQ's advantages to model multi-source exposure-response data over the existing approach, SK (Appendix 6.1) and MEM (Section 2.3) method.

Case 1: A multi-source dose-time-response case is developed to show SKQ's modeling efficiency by pooling information across multiple data sources. The SKQ results are compared to those provided by SK, which models each source of data separately without information pooling.

Case 2: A multi-source dose-response case is developed to show that compared to MEM (the existing information-pooling method), SKQ is a more general method which is free of the restrictive assumptions stipulated by MEM.

The empirical studies are based on simulation experiments, i.e., sampling through computer experiments whose outputs mimic real lab experiment data. Simulation, rather than real lab experiments, is employed for the following reasons. First, only in a simulation-based study, the true response surfaces (i.e., the simulation models) are available to evaluate

the approximation delivered by methods proposed. Second, the performance of a modeling approach can be evaluated in a rigorous statistical manner by applying it to a large number of randomly-sampled data sets, which is impossible to achieve in real lab experiments. These advantages of simulation studies will become clear later.

3.5.1 Case 1: Modeling Multi-Source Dose-Time-Response Data

This case is constructed based on the dose-time-response study of TiO₂ nanoparticles (NPs) performed by Porter et al. [50]. There are two quantitative factors $\mathbf{x} = (x_1, x_2)$, with $x_1 \in [0, 15]$ μg representing the TiO₂ dosage, and $x_2 \in [1, 112]$ days representing the post-exposure time. There is one qualitative factor for the shape of NPs, which is denoted as z . The variable z has two category levels $\{c_1, c_2\}$: c_1 denotes short TiO₂ nanobelts and c_2 long TiO₂ nanobelts. Each category corresponds to a different subpopulation/data source. The vector of all the factors is denoted as $\mathbf{w} = (\mathbf{x}, z)$. The response of interest is BAL (bronchoalveolar lavage) PMNs measured in the units of $10^3/\text{mouse}$.

Simulation Model

The simulation model, which is used to generate simulation data that mimic real experimental data, is described as follows. The true expected responses for the two subpopulations (short and long nanobelts) are represented as $\{Y(\mathbf{x}, c_1), Y(\mathbf{x}, c_2)\}$, with specific expressions given by Model (6.16-6.17) in the Appendix 6.2. Both Models (6.16) and (6.17) take the form of a single-hidden layer feedforward neural network, and are estimated from real biological data [51]. The true dose-time-response surfaces are plotted in Figure 3.2.

The true variance models used in the simulation are given as:

$$\text{Var}[\varepsilon(\mathbf{x}, c_1)] = (0.2Y(\mathbf{x}, c_1)^{0.7})^2 \quad (3.21)$$

$$\text{Var}[\varepsilon(\mathbf{x}, c_2)] = (0.3 \exp(Y(\mathbf{x}, c_2) \times 0.005))^2 \quad (3.22)$$

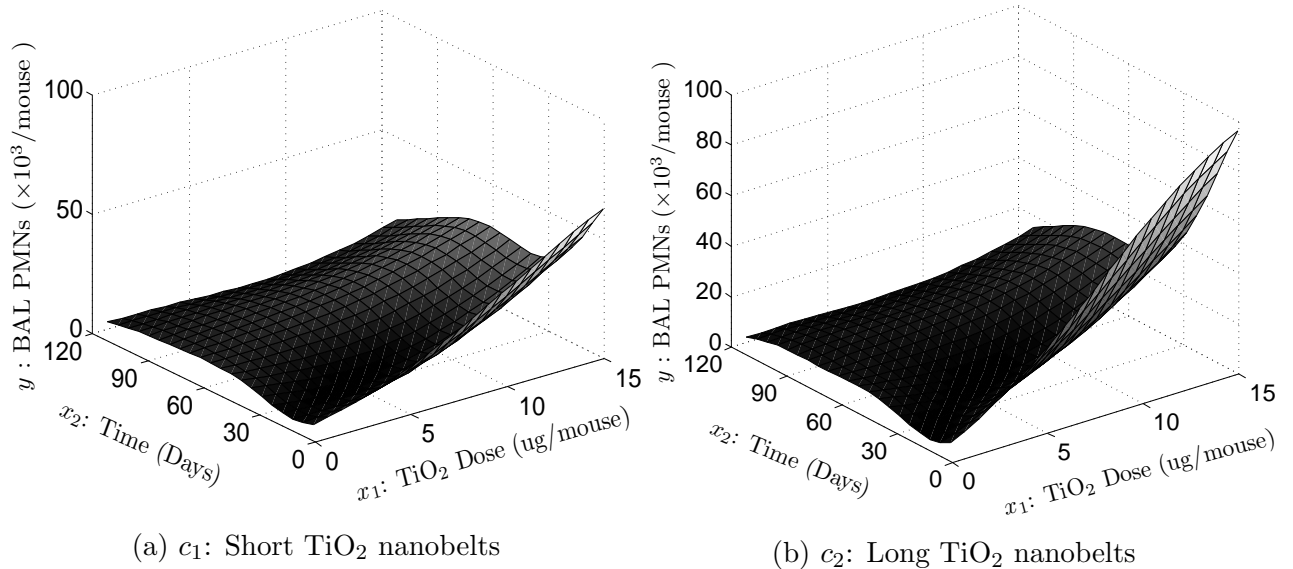


Figure 3.2: The true exposure-response surfaces for Case 1.

For a subpopulation c_q ($q = 1, 2$) and at an exposure level \mathbf{x}_0 , a random response y_0 is simulated as

$$y_0 = Y(\mathbf{x}_0, c_q) + \sqrt{\text{Var}[\varepsilon(\mathbf{x}_0, c_q)]} \cdot \epsilon; \quad q = 1, 2, \quad (3.23)$$

where ϵ is a random error generated by a standard normal random generator [52].

Case 1 is designed to compare the modeling efficiency of SKQ and SK. MEM has not been applied to this case for the following two reasons. First, MEM relies on a common nonlinear functional form adequate to model the underlying dose-time-response surface for each subpopulation (data source) c_q , which is very difficult, if not impossible, to identify for the three-dimensional complex surfaces (Figure 3.2) in this case. Second, the variance structures of (3.21-3.22) are different across the two categories, which violates the assumption of common variance structure required by MEM. For details regarding the related MEM assumptions, please refer to Section 2.3.

Sampling via Simulation

Two types of data sets are obtained via simulation: an estimation data set (EDS) for model estimation/inference, and an validation data set (VDS) which is used to evaluate the quality of the model estimated from an EDS.

For Case 1, an EDS includes a total of 32 distinct design points: 16 points for subpopulation c_1 (short nanobelts) depicted as stars in Figure 3.3(a), and 16 points for subpopulation c_2 (long nanobelts) depicted as stars in Figure 3.3(b). At each design point, Model (4.13) is used to generate 8 i.i.d. random responses; that is, 8 replications are assigned to each distinct design point.

The VDS includes a dense grid of 16,912 check points, which are depicted as dots in Figure 3.3. The collection of all the check points is denoted as \mathcal{C} , with $\mathcal{C} = \mathcal{C}_{c_1} \cup \mathcal{C}_{c_2}$; \mathcal{C}_{c_1} denotes the collection of dots in Figure 3.3(a), and \mathcal{C}_{c_2} the collection of dots in Figure 3.3(b). Further, \mathcal{C}_{c_q} ($q = 1, 2$) is divided into a number of subsets: $\mathcal{C}_{c_q} = \mathcal{C}_{c_q,1} \cup \mathcal{C}_{c_q,2} \cup \dots \cup \mathcal{C}_{c_q,15}$, where $\mathcal{C}_{c_q,k}$ represents the collection of check points within the subregion specified by the dose range $[k - 1, k)$; the subregions are shown in Figure 3.3(a) and (b) by the alternating white and gray rectangles. At each check point, the true expected response $Y(\cdot)$ is available (Figure 3.2) to evaluate the models fitted from the EDS.

Applying the Modeling Methods

On an EDS generated following the design as given in Figure 3.3, both SK and SKQ were applied to model the target response surfaces.

SK has no information-pooling ability, and fits a separate SK model for each subpopulation solely based on the corresponding subset of data. When applying SK, the exponential correlation function (3.5) is adopted to capture the extrinsic variability for both subpopulations. Two separate SK models are fitted for the short-nanobelt and long-nanobelt data subsets, respectively. At an arbitrary setting $\mathbf{w} = (\mathbf{x}, c_q)$, Formula (6.13) and (6.14) were

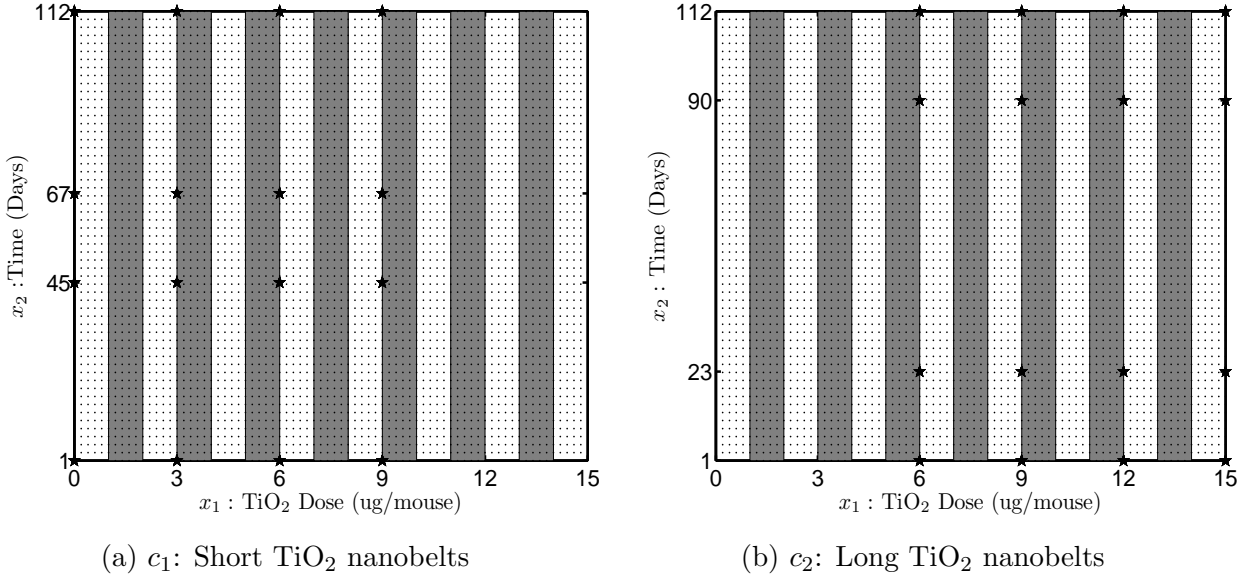


Figure 3.3: Design points in the EDS (estimation data set) and check points in the VDS (validation data set)

used, based on the fitted SK model associated with c_q , to obtain respectively the point estimate and CI of the expected response $Y(\mathbf{w})$.

To apply SKQ, the correlation functions for the extrinsic variability (3.4) are specified as follows: the exponential correlation function (3.5) is adopted for $K(\mathbf{x}, \mathbf{x}')$, and the EC (isotropic) correlation function (3.6) is used for $\tau_{z_\ell, z'_\ell}^{(\ell)}$. Fitted from an EDS consisting of two sources of data subsets, the resulting SKQ model can be used for estimation and inference. At an arbitrary factor setting $\mathbf{w} = (\mathbf{x}, c_q)$, Formula (3.16) provides the point estimate for $Y(\mathbf{w})$; and (3.18) gives the CI for $Y(\mathbf{w})$.

The Pooling Effects of SKQ

SKQ's strength lies in its ability to pool information across multiple subpopulations/data sources. The estimation quality w.r.t a subpopulation can be substantially improved over SK because SKQ allows for the borrowing information (or data) from all the other subpopulations. To demonstrate SKQ's estimation efficiency, herein we compare the performance of SK and SKQ in terms of predicting w.r.t $Y(\cdot)$.

The estimated root mean squared error (ERMSE) defined in (3.24) is adopted as the performance measure. Recall that $\mathcal{C}_{c_q,k}$ is the collection of the checkpoints for subpopulation c_q included in the rectangle specified by dose range $[k-1, k)$ as shown in Figure 3.3. We define

$$\text{ERMSE}(\mathcal{C}_{c_q,k}) = \sqrt{\frac{1}{\#\mathcal{C}_{c_q,k}} \sum_{\mathbf{w} \in \mathcal{C}_{c_q,k}} \left(\widehat{Y}(\mathbf{w}) - Y(\mathbf{w}) \right)^2}; \quad q = 1, 2; k = 1, 2, \dots, 15, \quad (3.24)$$

where $\#\mathcal{C}_{c_q,k}$ represents the total number of check points in the set $\mathcal{C}_{c_q,k}$. Clearly, (3.24) measures the average deviation of $\widehat{Y}(\cdot)$ from the true value $Y(\cdot)$ at the check points in $\mathcal{C}_{c_q,k}$.

Applying a modeling method (SK or SKQ) to one EDS, denoted as $\text{EDS}^{(r)}$ is considered as one macro-replication, and leads to a set of performance statistics

$$\{\text{ERMSE}^{(r)}(\mathcal{C}_{c_q,k}); q = 1, 2; k = 1, 2, \dots, 15\}, \quad r = 1, 2, \dots, R.$$

For empirical evaluation, a total of $R = 1000$ independent EDS are generated by simulation experiments. The average ERMSE across the macro-replications are calculated as

$$\overline{\text{ERMSE}}(\mathcal{C}_{c_q,k}) = \frac{1}{R} \sum_{r=1}^R \text{ERMSE}^{(r)}(\mathcal{C}_{c_q,k}); q = 1, 2; k = 1, 2, \dots, 15\},$$

, and are summarized in Table 3.1. Evidently, in each subset of check points $\mathcal{C}_{c_q,k}$, the point estimates given by SKQ are much more accurate than those provided by SK. For 22 out of the 30 check-point subsets in Table 3.1, the average ERMSE given by SKQ is less than half of that given by SK. Clearly, by synergistically modeling multi-source data, SKQ leads to substantially improved prediction results.

3.5.2 Case 2: Modeling Multi-Source Dose-Response Data

This case is derived from the dose-response study of TiO_2 nanoparticles (NPs) after 3 days of exposure performed by Porter et al. [50]. There is one quantitative factor x representing

Table 3.1: Comparison of the estimation results by SK and SKQ for Case 1

Subset of checkpoints	$\overline{\text{ERMSE}}(\mathcal{C}_{c_q,k})$		Subset of checkpoints	$\overline{\text{ERMSE}}(\mathcal{C}_{c_q,k})$	
	SK	SKQ		SK	SKQ
$\mathcal{C}_{c_1,1}$	1.6110	1.5876	$\mathcal{C}_{c_2,1}$	5.1988	2.9626
$\mathcal{C}_{c_1,2}$	2.0351	1.4277	$\mathcal{C}_{c_2,2}$	4.1851	2.5193
$\mathcal{C}_{c_1,3}$	2.4907	1.2814	$\mathcal{C}_{c_2,3}$	3.4493	2.1025
$\mathcal{C}_{c_1,4}$	2.9825	1.1738	$\mathcal{C}_{c_2,4}$	2.8623	1.6950
$\mathcal{C}_{c_1,5}$	3.4742	1.1290	$\mathcal{C}_{c_2,5}$	2.4246	1.3363
$\mathcal{C}_{c_1,6}$	3.9355	1.1518	$\mathcal{C}_{c_2,6}$	2.1431	1.0676
$\mathcal{C}_{c_1,7}$	4.3396	1.2258	$\mathcal{C}_{c_2,7}$	1.9998	0.9487
$\mathcal{C}_{c_1,8}$	4.6247	1.3294	$\mathcal{C}_{c_2,8}$	1.9694	1.0057
$\mathcal{C}_{c_1,9}$	4.7185	1.4512	$\mathcal{C}_{c_2,9}$	2.0524	1.1582
$\mathcal{C}_{c_1,10}$	4.5792	1.5929	$\mathcal{C}_{c_2,10}$	2.2103	1.3311
$\mathcal{C}_{c_1,11}$	4.2379	1.7578	$\mathcal{C}_{c_2,11}$	2.3942	1.4815
$\mathcal{C}_{c_1,12}$	3.8478	1.9497	$\mathcal{C}_{c_2,12}$	2.5792	1.5902
$\mathcal{C}_{c_1,13}$	3.7447	2.1753	$\mathcal{C}_{c_2,13}$	2.7650	1.6665
$\mathcal{C}_{c_1,14}$	4.3577	2.4427	$\mathcal{C}_{c_2,14}$	2.9780	1.7489
$\mathcal{C}_{c_1,15}$	5.8143	2.7547	$\mathcal{C}_{c_2,15}$	3.2707	1.9007

the dose level, and $x \in [0, 20]$ μg . The one qualitative factor is denoted as z with three categories $\{c_1, c_2, c_3\}$. Each category is used to represent a different batch of animals. The vector of factors is written as $\mathbf{w} = (x, z)$. The response of interest is BAL (bronchoalveolar lavage) PMNs measured in the units of $10^3/\text{mouse}$.

Simulation Model

Simulation data that mimic the real-experiment data are generated according to the following true dose-response model

$$\begin{aligned}
 Y(x, c_1) &= 20 \exp(x/11) \\
 Y(x, c_2) &= 19.5 \exp(x/10.5) \\
 Y(x, c_3) &= 26 \exp(x/13.2),
 \end{aligned} \tag{3.25}$$

and the variance model

$$\text{Var}[\varepsilon(x, c_q)] = 0.2^2 Y(x, c_q)^{2 \times 0.6}; \quad q = 1, 2, 3. \tag{3.26}$$

For a certain category c_q and at a dose level x_0 , a random response y_0 is simulated as:

$$y_0 = Y(x_0, c_q) + 0.2Y(x_0, c_q)^{0.6} \cdot \epsilon; \quad q = 1, 2, 3, \quad (3.27)$$

where ϵ denotes a random error generated by a standard normal random generator. [52].

Case 2 is used to compare SKQ and MEM, and is designed in such a way that the two basic assumptions required by MEM are met: (i) A nonlinear functional form can be easily identified and employed to model the target dose-response curves; (ii) There is a common variance structure (3.26) across different data sources. The third assumption made by MEM is the multivariate normality of the model coefficient vector (Section 2.3). According to (3.25), the three true coefficient vectors in this case are: $(20, 11)^\top$, $(19.5, 10.5)^\top$ and $(26, 13.2)^\top$, based on which it is hardly possible to judge whether the normality assumption holds or not; this is quite typical of multi-source data.

Simulation-Based Sampling

As in Case 1, both EDS and VDS are generated in this study for model estimation and evaluation respectively. To generate an EDS, the design as shown in Table 3.2 is used: 8 replications are carried out at 5 evenly-spaced dose levels for each of the three categories. An example EDS is given in Table 6.1 (the Appendix 6.3).

Table 3.2: Design points in the EDS (estimation data set) for Case 2.

x : dose	0	5	10	15	20	0	5	10	15	20	0	5	10	15	20
c_q : subpopulation	c_1					c_2					c_3				

Applying the Modeling Methods

The two alternative information-pooling methods, MEM and SKQ, were applied respectively to the EDS given in Table 6.1.

To perform MEM, the common nonlinear functional form of the dose-response curve for each of the three subpopulations is assumed to be

$$Y(x, c_q) = \alpha_{c_q,1} \exp(x/\alpha_{c_q,2}); \quad q = 1, 2, 3, \quad (3.28)$$

with $\alpha_{c_q} = (\alpha_{c_q,1}, \alpha_{c_q,2})$ being the unknown parameters for subpopulation c_q ($q = 1, 2, 3$).

The variance model is assumed to follow

$$\text{Var}[\varepsilon(x, c_q)] = \sigma^2 Y(x, c_q)^{2\gamma}; \quad q = 1, 2, 3, \quad (3.29)$$

where σ and γ are unknown parameters common to different subpopulations. Note that by using (3.28) and (3.29), there is no model misspecification problem for MEM and only unknown parameters need to be estimated. This is meant to the advantage of MEM.

With the assumed forms (3.28-3.29), MEM is performed on the EDS in Table 6.1. The fitted dose-response models are:

$$\begin{aligned} \widehat{Y}(x, c_1) &= 20.04 \exp(x/11.09) \\ \widehat{Y}(x, c_2) &= 19.60 \exp(x/10.54) \\ \widehat{Y}(x, c_3) &= 26.24 \exp(x/13.38), \end{aligned} \quad (3.30)$$

and the fitted variance model is

$$\widehat{\text{Var}}[\varepsilon(x, c_q)] = 0.214^2 \widehat{Y}(x, c_q)^{2 \times 0.597}; \quad q = 1, 2, 3. \quad (3.31)$$

With the fitted MEM, the expected response can be estimated at any dose level, and the BMD estimate can be obtained for a given BMR; the confidence intervals can also be constructed for these quantities of interest (see Appendix 2.3).

When applying SKQ to the same EDS (Table 6.1), the correlation (3.4) is constructed as follows: the exponential correlation function (3.5) is used to model the correlations be-

tween quantitative variables, and the MC correlation function (3.7) is used to model the correlations across different levels of qualitative factors. Normalization of the original data (Table 6.1) was also performed so that both the quantitative factors and responses range over [0,1]. Applying the maximum likelihood estimation procedure (see the subsection 3.3) on the normalized data leads to the fitted parameters for the SKQ model as displayed in Table 3.3.

Table 3.3: The SKQ parameters estimated from the normalized dose-response data for Case 2.

$\hat{\beta}_0$	$\hat{\delta}^2$	$\hat{\theta}_1$	\hat{p}	$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\phi}_3$
0.5426	0.1272	1.3588	1.9168	0.01	0.0144	0.022

With the fitted SKQ model, the expected response can be estimated at any dose level (the subsection 3.3), and the inverse BMD can be obtain numerically (the subsection 3.4); the confidence intervals for these quantities can also be obtained accordingly. Note that when utilizing the SKQ specified by Table 3.3 for the estimation/inference, a simple conversion calculation is needed to ensure that the estimates are given on the original scale, since those parameters are obtained from the normalized data.

Comparison of the Two Modeling Methods

The estimation/inference abilities of MEM and SKQ are compared in terms of (I) the expected responses as well as (II) the BMD values.

(I) *Estimation/Inference of the expected response* $Y(\cdot)$

From the one EDS given in Table 6.1, both MEM and SKQ were applied as described earlier, and the estimation results for the two methods are displayed in Figure 3.4. The circles denote the EDS, and are plotted in both Figure 3.4(a) and (b). The dashed curves represent the estimated expected responses, and the solid curves are the lower and upper 95% CI bands for the true expected responses. As shown in Figure 3.4, over the dose range, the widths of the CIs (that is, the vertical distances between the lower and upper CI bands) provided

by SKQ are typically narrower than those given by MEM; this is illustrated for one EDS in Figure 3.4, which holds consistently for all 1000 macro-replications carried out in this study.

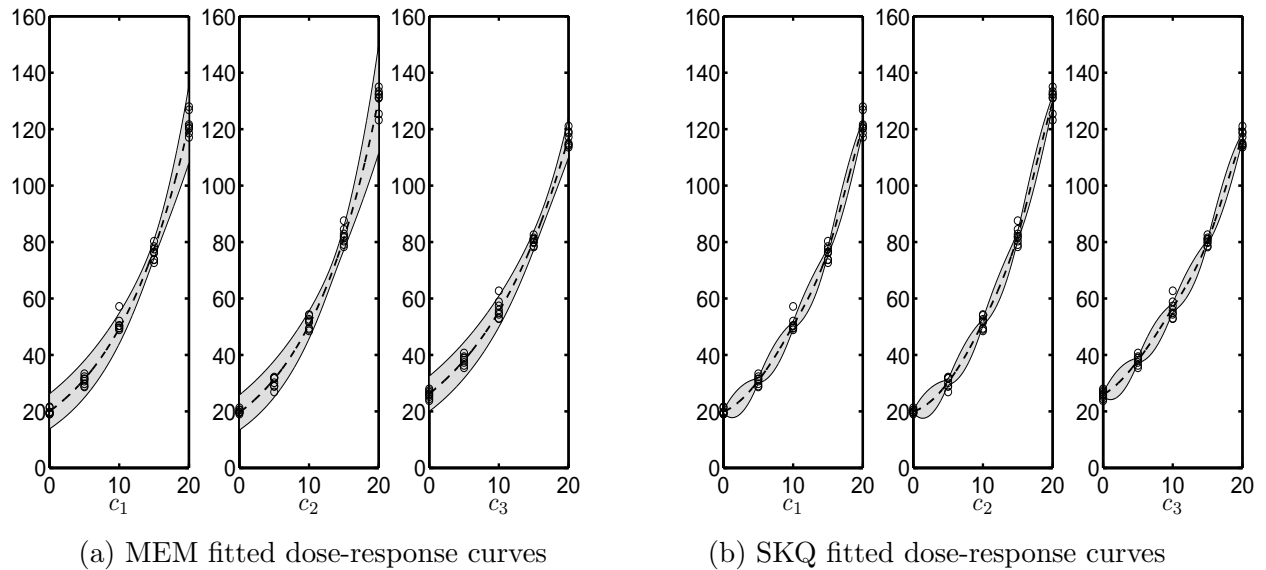


Figure 3.4: Comparison of the dose-response fitting results from MEM and SKQ

As explained in Case 1, one macro-replication refers to the process of applying a method on one randomly-generated EDS. Using a modeling method (MEM or SKQ), $R = 1000$ macro-replications lead to 1000 CIs of the true expected response $Y(\cdot)$ for any check point specified in terms of (x, c_q) . Hence, the coverage probability of the CIs can be estimated as the percentage of the 1000 CIs that include the true expectation $Y(\cdot)$. In our simulation study, $Y(\cdot)$ is available from the simulation model (3.25) for the purpose of evaluating the CIs. Ideally, among these 1000 CIs, the percentage of the CIs that actually contain $Y(\cdot)$ should be very close to 95%, the nominal coverage level.

Table 3.4 presents the coverage probabilities of the 95% CIs given by MEM and SKQ respectively, based on each method's 1000 macro-replications. The first two rows of Table 3.4 specify a number of check points. The estimated coverage probabilities of the MEM CIs are given in the row marked as "MEM", and are all 1.000 at the check points, which is way higher than the nominal 95%. The estimated coverage probabilities resulting from SKQ are given in the row labeled as "SKQ", and are much closer to the nominal percentage 95%.

Therefore, as can be seen from Figure 3.4 and Table 3.4, SKQ is able to provide tighter CIs (i.e., CIs that are narrower and with more on-target coverage probabilities) for the true expected responses, while MEM overshoots the nominal coverage percentage by providing overly conservative CIs.

Table 3.4: Comparison of MEM and SKQ in terms of the CI coverage probabilities for the expected response $Y(\cdot)$.

	subpopulation c_1				subpopulation c_2				subpopulation c_3			
x :dose	2.5	7.5	12.5	17.5	2.5	7.5	12.5	17.5	2.5	7.5	12.5	17.5
MEM	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SKQ	0.957	0.966	0.974	0.933	0.952	0.955	0.970	0.926	0.955	0.970	0.964	0.968

(II) Estimation/Inference of the BMD

Herein, the two methods, MEM and SKQ, are compared in terms of their inverse estimates for the BMD associated with a pre-specified BMR. For demonstration, the BMR is set as 42 10^3 /mouse in this case.

As already explained, $R = 1000$ macro-replications were performed using each of the two methods based on the 1000 data sets $\{\text{EDS}^{(r)}; r = 1, 2, \dots, R\}$. From the r^{th} ($r = 1, 2, \dots, R$) macro-replication, a one-sided 95% CI was constructed for $\text{BMD}(c_q)$, $q = 1, 2, 3$, following the bootstrapping resampling method (Section 3.4); the lower bound of the one-sided CI is called BMDL. The BMDLs estimated by MEM are denoted as

$$\{\widehat{\text{BMDL}}_{\text{MEM}}^{(r)}(c_q); q = 1, 2, 3; r = 1, 2, \dots, 1000\}, \quad (3.32)$$

and the BMDLs obtained by SKQ are represented as

$$\{\widehat{\text{BMDL}}_{\text{SKQ}}^{(r)}(c_q); q = 1, 2, 3; r = 1, 2, \dots, 1000\}, \quad (3.33)$$

The true values for $\text{BMD}(c_q)$, $q = 1, 2, 3$, can be easily obtained for $\text{BMR} = 42$ based on the simulation model (3.25), and are represented by the horizontal lines in Figure 3.5 (a), (b),

and (c) corresponding to the three subpopulations. Figure 3.5 (a) is devoted to the subpopulation c_1 , and the two box plots are respectively for $\{\widehat{\text{BMDL}}_{\text{MEM}}^{(r)}(c_1); r = 1, 2, \dots, 1000\}$ and $\{\widehat{\text{BMDL}}_{\text{SKQ}}^{(r)}(c_1); r = 1, 2, \dots, 1000\}$; Figure 3.5 (b) and (c) are plotted for the other two subpopulations. Clearly, the BMDLs given by SKQ are much closer to the true BMD than those provided by MEM.

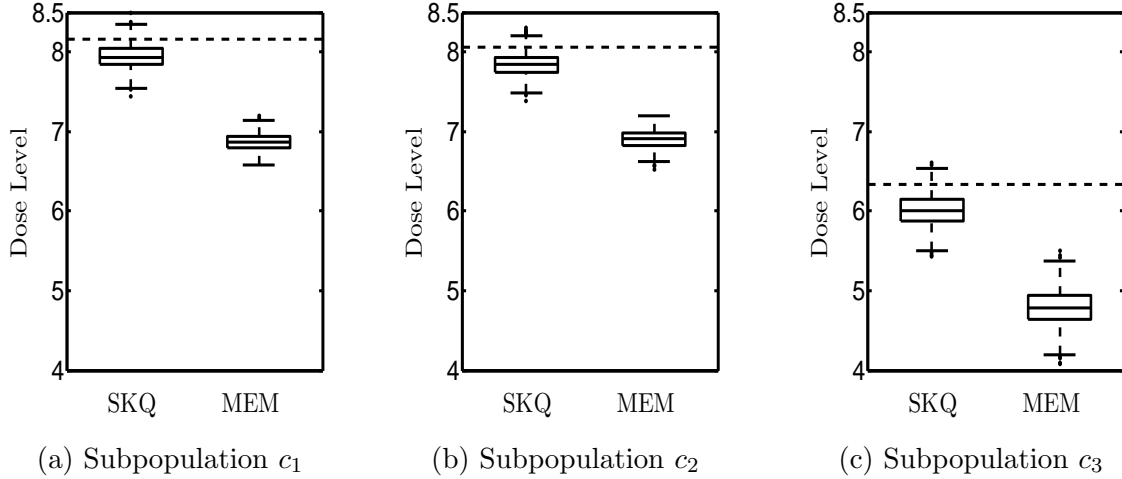


Figure 3.5: Box plots for the BMDLs resulting from the two modeling methods.

Each BMDL estimate in (3.32) and (3.33) corresponds to a one-sided 95% CI: $[\text{BMDL}, \infty]$. Table 3.5 compares the coverage probabilities of the CIs obtained from MEM and SKQ. It can be seen that the coverage probabilities of SKQ is close to the nominal level 95%, whereas MEM's estimated probabilities are all 1.000. Therefore, as shown in Figure 3.5 and Table 3.5, SKQ is able to give more informative CIs of the BMD as compared to MEM.

Table 3.5: Comparison of MEM and SKQ in terms of the CI coverage probabilities for the BMD.

	subpopulation c_1	subpopulation c_2	subpopulation c_3
Pre-specified BMR	42	42	42
MEM	1.000	1.000	1.000
SKQ	0.9349	0.933	0.959

Chapter 4

Design of Multiple-Source Experiments

In this chapter, we develop a two-stage DOE procedure for multi-source exposure-response modeling. Compared to standard once-and-for-all designs, sequential (multi-stage) designs typically lead to (i) savings in sample size if model estimates of desired accuracy (measured by the uncertainty/variability of estimates) are to be achieved, or (ii) model estimates of improved accuracy given a fixed total sample size [15, 53]. Since biological studies are typically performed with restricted sample sizes, the two-stage design procedure in this chapter is tailored to achieve models of the highest quality with a pre-specified total sample size. It is worth pointing out that the two-stage design can be easily adapted to a design procedure which is driven by a desired model accuracy (such as the variance of the BMD estimate), which is not constrained by a pre-specified total sample size; such a procedure may involve multiple (more than two) stages of experiments.

4.1 The Two-Stage Procedure

Given data involving $Q - 1$ sources, the inputs of the two-stage DOE procedure for Source Q are given as

Input 1: \mathfrak{R} , the design region of interest of Source Q .

Input 2: N , the total sample size available for the Source Q .

Input 3: For Source Q , N_I , the batch size of the experiments (i.e., the number of samples) performed at Stage I (Section 4.1.1), which implies that the sample size is $N - N_I$ at Stage II.

Input 4: BMR (optional), the pre-specified benchmark response level, which is not necessary unless the design criterion is chosen as $\text{Var}[\widehat{\text{BMD}}(c_Q)]$.

The outputs of the procedure include: the design of experiments for the N samples for the Source Q ; the SKQ model, which is fitted on data from all Q sources; and the estimated uncertainty of the fitted SKQ

Suppose that some existing exposure-response data have been obtained from $Q - 1$ sources. The DOE task is to find the design for Source- Q experiments to optimize the model estimation quality. Figure 4.1 provides an overview of the two-stage design for the Source Q . In Stage I, pilot experiments with a sample size of N_I are carried out following the initial design, and an initial statistical modeling/inference is performed. The design augmentation in Stage II is performed utilizing the information obtained in Stage I, and $N - N_I$ additional experiments are carried out following the augmented design.

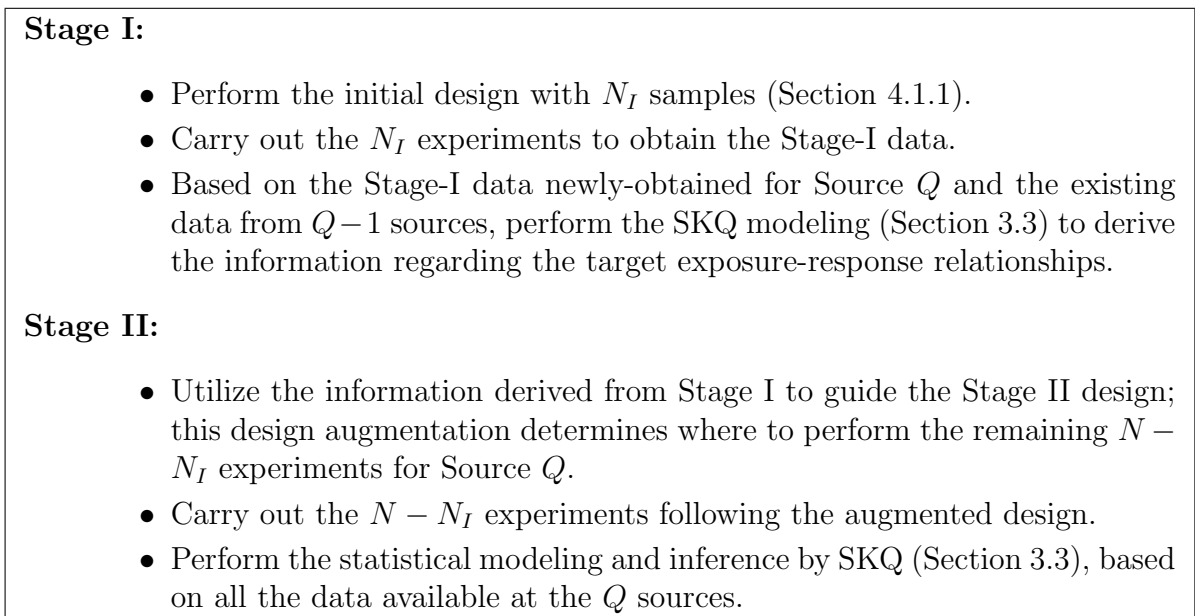


Figure 4.1: Overview of the two-stage design procedure.

Both stages of the design in Figure 4.1 are restricted to Source Q only, and intends to address the following question: Given the data collected from $Q - 1$ distinct sources, at what experiment conditions of the Source Q should the total N samples be allocated?

Two alternative sets of notations are used in this dissertation to represent a design.

One is given as:

$$\tilde{\mathfrak{D}} = \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_D \\ n_1 & n_2 & \cdots & n_D \end{pmatrix} \quad (4.1)$$

with $\mathbf{w}_d = (\mathbf{x}_d^\top, c_d)^\top$ being the experiment condition, n_d the number of replications assigned to \mathbf{w}_d , and D the total number of distinct design points in $\tilde{\mathfrak{D}}$. Clearly, the total sample size given by $\tilde{\mathfrak{D}}$ is $\sum_{d=1}^D n_d$. Alternatively, a design (4.1) can also be represented by

$$\mathfrak{D} = \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_M \end{pmatrix} \quad (4.2)$$

with $M = \sum_{d=1}^D n_d$ being the number of samples. Note that the array \mathfrak{D} may well include the same design points for multiple times which corresponds to multiple replications. In the remainder of this dissertation, both formulas will be used to refer to a design.

4.1.1 Stage I: Initial Design

At Stage I, N_I preliminary samples are to be performed for Source Q . Given N , the total number of samples available for Source Q , we recommend to set N_I as $\frac{1}{4}N$ to $\frac{1}{2}N$. Guidelines for specifying N_I can also be found in Santner et al. [15]. Having selected a value for N_I , the next question is how to allocate these N_I samples. Following the notation in (4.1), the initial design is represented as

$$\tilde{\mathfrak{D}}^{(I)} = \begin{pmatrix} \mathbf{w}_1^{(I)} & \mathbf{w}_2^{(I)} & \cdots & \mathbf{w}_{D_I}^{(I)} \\ n_1^{(I)} & n_2^{(I)} & \cdots & n_{D_I}^{(I)} \end{pmatrix} \quad (4.3)$$

where $\{\mathbf{w}_d^{(I)} = (\mathbf{x}_d^{(I)\top}, c_Q)^\top; d = 1, 2, \dots, D_I\}$ and D_I is the number of distinct experiment conditions in Stage I for the Source Q .

At Stage I, a total of N_I samples are to be allocated within the feasible region of Source Q , out of the following considerations. First, the initial design points should provide a fair

coverage of the feasible region, and a space-filling design such as Latin hypercube designs [54] and orthogonal Latin hypercube designs [55] could be adopted. Second, endpoints of the design region have to be included in the design to avoid potential extrapolation [56]. Third, we have $n_d^{(I)} \geq 3$ for any d , to allow for the detection of variance heterogeneity. Lastly, in the absence of any other concerns, we can simply set $n_1^{(I)} = n_2^{(I)} = \dots = n_I^{(I)}$. Examples of such initial designs can be found in Sections 4.2.1 and 4.2.2.

Following the initial design, a total of N_I samples are collected, and are denoted as $\{(\mathbf{x}_d^{(I)}, c_Q), \mathcal{Y}(\mathbf{x}_d^{(I)}, c_Q); d = 1, 2, \dots, D_I\}$. Based on the initial data for Source Q , and the existing data from $Q - 1$ sources, the SKQ model is fitted and will be used to guide the DOE in Stage II.

4.1.2 Stage II: Design Augmentation

The task of the Stage-II design augmentation is to find out how to allocate the rest $N - N_I$ samples in the design region of the Source Q . The stage-II design is denoted by

$$\tilde{\mathfrak{D}}^{(II)} = \begin{pmatrix} \mathbf{w}_1^{(II)} & \mathbf{w}_2^{(II)} & \dots & \mathbf{w}_M^{(II)} \\ n & n & \dots & n \end{pmatrix}, \quad (4.4)$$

where the same design point may appear multiple times, and n represents the number of replications performed at a design point every time experiments are to be performed at that point.

Hence, $M = (N - N_I)/n$ design points are to be determined by solving the optimization problem (4.5).

$$\begin{aligned} & \min_{\tilde{\mathfrak{D}}^{(II)}} C(\cdot) \\ \text{Subject to} & \quad \text{the existence of } N_I \text{ Source-}Q \text{ samples } \{(\mathbf{x}_d^{(I)}, c_Q), \mathcal{Y}(\mathbf{x}_d^{(I)}, c_Q); d = 1, 2, \dots, N_I\} \\ & \quad \text{and the existing data from } Q - 1 \text{ sources.} \end{aligned} \quad (4.5)$$

Solving (4.5) is challenging. (i) First, there may well not be closed-form formulation for the objective criterion $C(\cdot)$. The design criterion in (4.5) quantifies the quality of the fitted SKQ model from all the data available (that is, the existing data from $Q-1$ sources, the initial data for Source Q , and the data to be collected in Stage II for Source Q), and a range of criteria can be adopted [39] such as IMSE, MMSE and entropy. For instance, in Section 4.2.2, the criterion is set to be $\text{Var}[\widehat{\text{BMD}}(c_Q)]$, the variability of the estimated BMD rendered by the fitted SKQ, and $\text{Var}[\widehat{\text{BMD}}(c_Q)]$ cannot be evaluated analytically. (ii) Second, determining $\tilde{\mathfrak{D}}^{(II)}$ simultaneously is difficult due to the high dimension (which is $M = N_{II}/n$) of the decision variables. Hence, numerical methods are developed to evaluate the design criterion if necessary, and heuristics could be adopted to add a portion of the M design points at a time until all the M points have been located. In addition, global optimization algorithms, such as the Matlab function `MultiStart` which allows for parallel computing, could be used for optimization search.

Following the augmented design $\tilde{\mathfrak{D}}^{(II)}$, Stage-II experiments will be performed. Based on all the available data across the Q sources, SKQ model fitting and inference will be performed to obtain the exposure-response models, and to quantify the uncertainty of the estimates of interest.

4.2 Simulation Studies

The proposed two-stage design procedure has been evaluated via simulation studies.

4.2.1 Case 1: Design for Multi-Source Dose-Time-Response Experiments

This case is derived based on the dose-time-response study of TiO_2 nanoparticles (NPs), the one in Subsection 3.5.1. There are two quantitative factors $\mathbf{x} = (x_1, x_2)$, with $x_1 \in [0, 15] \mu\text{g}$ representing the TiO_2 dosage, and $x_2 \in [1, 112]$ days representing the post-exposure time. There is one qualitative factor for the shape of NPs, which is denoted as z . The variable z has two category levels $\{c_1, c_2\}$: c_1 denotes short TiO_2 nanobelts and c_2 long TiO_2 nanobelts.

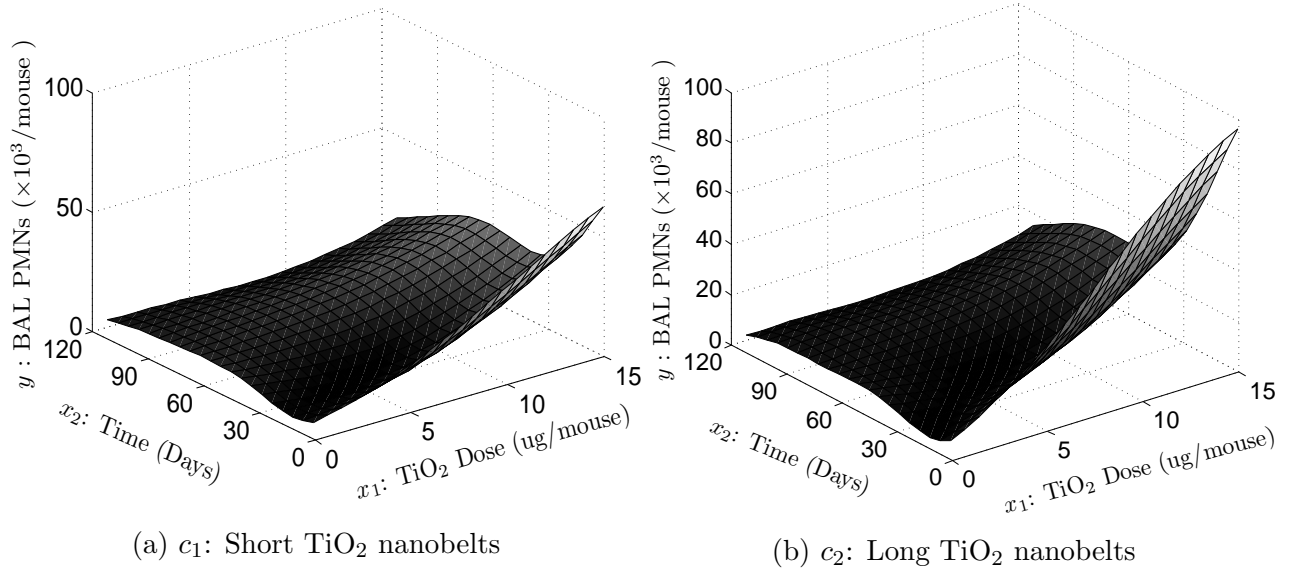


Figure 4.2: The true exposure-response surfaces for Design Case 1.

Each category corresponds to a different subpopulation/data source. The vector of all the factors is given as $\mathbf{w} = (\mathbf{x}, z)$. The response of interest is BAL (bronchoalveolar lavage) PMNs measured in the units of $10^3/\text{mouse}$.

Simulation Models

The simulation model, which is used to generate simulation data that mimic real experimental data, is described as follows. The true expected responses for the two subpopulations (short and long nanobelts) are represented as $\{Y(\mathbf{x}, c_1), Y(\mathbf{x}, c_2)\}$, with specific expressions given by Model (6.16-6.17) in Appendix 6.2. Both Models (6.16) and (6.17) take the form of a single-hidden layer feedforward neural network, and are estimated from real biological data [51]. The true dose-time-response surfaces are plotted in Figure 4.2.

The true variance models used in the simulation are given as:

$$\text{Var}[\varepsilon(\mathbf{x}, c_1)] = (0.02Y(\mathbf{x}, c_1)^{0.4})^2 \quad (4.6)$$

$$\text{Var}[\varepsilon(\mathbf{x}, c_2)] = (0.03 \exp(Y(\mathbf{x}, c_2) \times 0.05))^2 \quad (4.7)$$

For a subpopulation c_q ($q = 1, 2$) and at an exposure level \mathbf{x}_0 , a random response y_0 is simulated as

$$y_0 = Y(\mathbf{x}_0, c_q) + \sqrt{\text{Var}[\varepsilon(\mathbf{x}_0, c_q)]} \cdot \epsilon; \quad q = 1, 2, \quad (4.8)$$

where ϵ is a random error generated by a standard normal random generator [52].

The true models (6.16-6.17) and (4.6-4.7) are blind to the two-stage design, and are only used for two purposes in this study: (i) to generate simulation data that mimic real biological data, and (ii) to serve as the true benchmark to evaluate the model estimates obtained by the proposed DOE procedure.

Existing data for $Q - 1$ sources

The existing data obtained for short TiO_2 nanobelts (Source 1) include a total of 128 samples, which are allocated to the design points depicted as solid stars in Figure 4.3. At each of the 16 design points, there are 8 replications. The response subject to random errors is generated by plugging the value of \mathbf{x}_i and c_1 into (4.8).

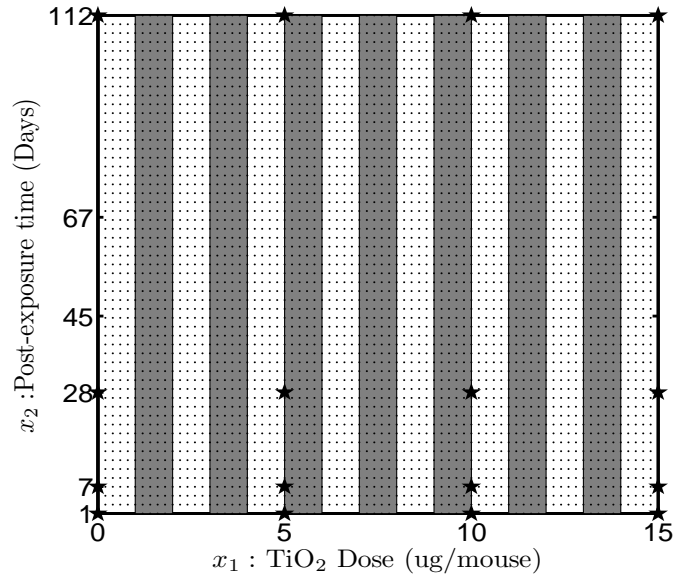


Figure 4.3: Design points and check points in the VDS (validation data set) of Short TiO_2 nanobelts for Case 1

Applying the two-stage design

The two-stage DOE procedure (Figure 4.1) is applied to obtain the DOE for long TiO_2 nanobelts (Source 2). The inputs of the procedure are given as follows:

Input 1: $\mathfrak{R} = [0, 15] \times [1, 112]$, the dose and time ranges of interest of long TiO_2 nanobelts.

Input 2: $N = 64$, the total sample size available for long TiO_2 nanobelts.

Input 3: $N_I = 32$: the sample size available to Stage I, which implies that the sample size is $N_{II} = 32$ at Stage II.

The criterion in the design optimization (4.5) is set as $\max_{\mathbf{w}_0 \in \mathfrak{R}} \widehat{\text{MSE}}[\widehat{Y}(\mathbf{w}_0)]$ for long TiO_2 nanobelt.

Stage I. Following Section 4.1.1, a sample of $N_I = 32$ samples was obtained in Stage I as solid stars in Figure 4.4 (a): four distinct design points are selected over the design region $\mathfrak{R} = [0, 15] \times [1, 112]$ with 8 replications at each point. At each design point \mathbf{x}_i ($i = 1, 2, \dots, 32$), the response subject to random errors is generated by plugging the value of \mathbf{x}_i and c_2 into (4.8).

Stage II. For Stage II, $n = 8$, and the augmented design points obtained from the design optimization are depicted as diamonds in Figure 4.4 (a). At each point, 8 replications are allocated.

Comparison with the traditional design

The results presented above in Section 4.2.1 represent the outcome of applying the two-stage design for one time. Due to the random nature of responses, re-applying the procedure will lead to different responses of Stage I, and different fitted SKQ model; and consequently, the augmented design in Stage II will turn out to be different; the final 64-sample data will be different, and the final modeling/inference results will be different.

Also because of the randomness in responses, every time the same traditional design is applied, a different set of data will be obtained, and hence different modeling/inference

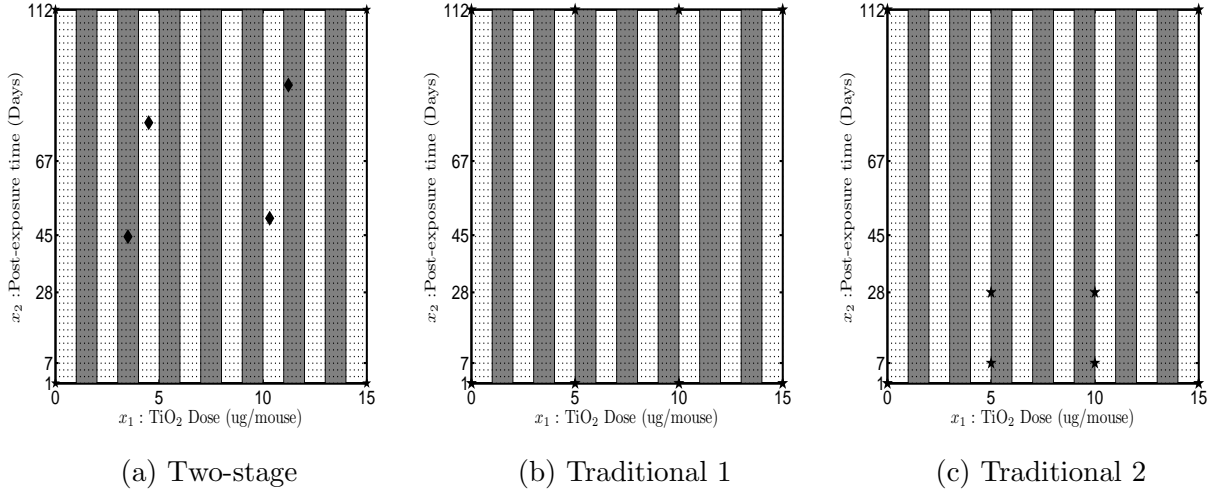


Figure 4.4: Design points and check points in the VDS (validation data set) of long TiO_2 nanobelts of three design approaches for Case 1

results will be obtained. Two traditional designs for this case are given in Figures 4.4 (b) and (c) respectively.

Each of the three design approaches was applied for 100 times with a given sample size of $N = 64$. As a result, 100 sets of results were obtained for each approach respectively. Since the primary goal of this study is to minimize maximum of MSE over the design region, \mathfrak{R} , for the long TiO_2 nanobelts, the two approaches are compared in terms of $\text{DEV} = \max_{\mathbf{w}_0 \in \mathfrak{R}} |\hat{Y}(\mathbf{w}_0) - Y(\mathbf{w}_0)|$, where \mathbf{w}_0 represents a check point of the long TiO_2 nanobelts.

Specifically, applying the two-stage design 100 times leads to 100 data sets; from each data set, SKQ in Section 3.3 were performed to obtain the model. From the 100 data sets, the 100 DEVs obtained are denoted as

$$\{\text{DEV}_{\text{Two-stage}}^{(r)}; r = 1, 2, \dots, 100\} \quad (4.9)$$

In the same way, applying each traditional design (Figures 4.4 (b) and (c)) 100 times also results in 100 data sets respectively. Performing the statistical modeling and inference on

these data sets respectively leads to 100 DEVs, which are denoted as

$$\{\text{DEV}_{\text{Trad1}}^{(r)}; r = 1, 2, \dots, 100\} \quad \text{and} \quad \{\text{DEV}_{\text{Trad2}}^{(r)}; r = 1, 2, \dots, 100\}. \quad (4.10)$$

The box plots are generated in Figure 4.5 summarize the DEVs (4.9) obtained from our two-stage design (Figure 4.4 (a)) and the DEVs (4.10) obtained from the two traditional designs (Figures 4.4 (b) and (c)). A box plot provides the basic information regarding the distribution of a data set (say, the DEVs in (4.10)), with the lower hinge being the 25th percentile, and the upper hinge being the 75th percentile of the data. For details of box plots, please refer to McGill et al. [57].

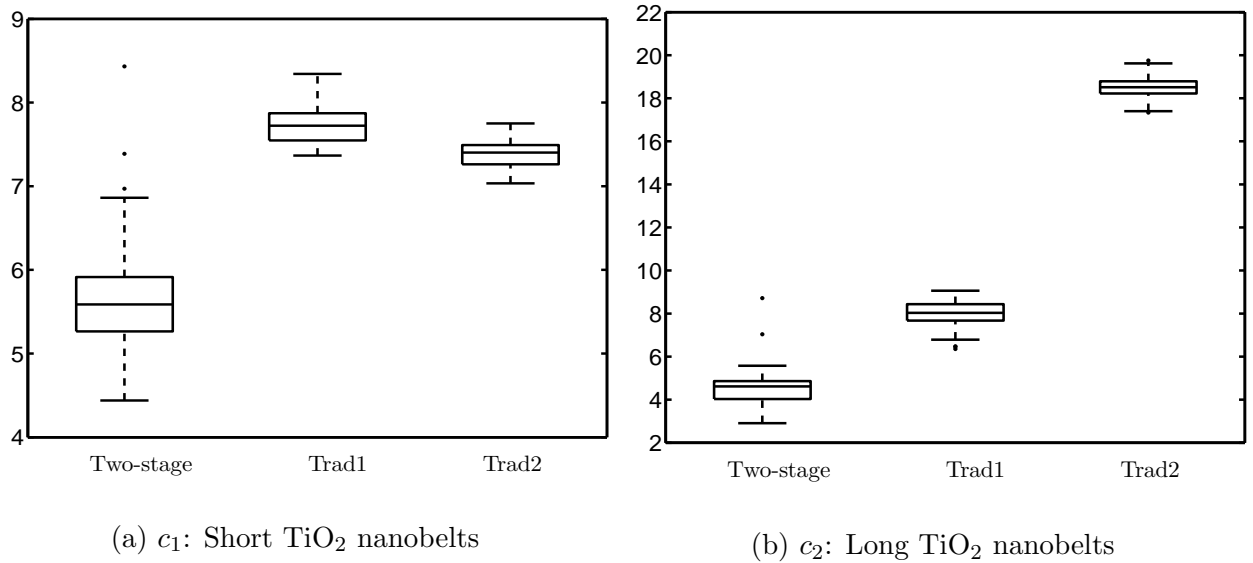


Figure 4.5: The box plots for the DEVs resulting from the three design approaches for Case 1.

Comparing the boxplots in Figure 4.5, it is evident that the DEVs resulting from our two-stage procedure (Figures 4.4 (a)) are smaller than that of two traditional designs (Figures 4.4 (b) and (c)). That is, with the same sample size, the two-stage design leads to fitted SKQ models with smaller DEVs from the true response surface for long nanobelts (Source 2).

For comparison of the three design approaches, a table as the one in Section 3.5 has been generated to compare these approaches in terms of $\overline{\text{ERMSE}}(\mathcal{C}_{c_q,k})$ (Subsection 3.5.1). It can be seen from Table 4.1 that the two-stage design excels for the long nanobelts, the modeling quality of which is used to drive the two-stage design; and performs about the same as the other two traditional designs in terms of short nanobelts.

Table 4.1: Comparison of the estimation results from three design methods for Case 1

Subset of checkpoints	$\overline{\text{ERMSE}}(\mathcal{C}_{c_q,k})$			Subset of checkpoints	$\overline{\text{ERMSE}}(\mathcal{C}_{c_q,k})$		
	Two-stage	Trad1	Trad2		Two-stage	Trad1	Trad2
$\mathcal{C}_{c_1,1}$	3.5140	3.6303	3.6702	$\mathcal{C}_{c_2,1}$	1.0107	2.5880	4.2095
$\mathcal{C}_{c_1,2}$	3.1735	3.2465	3.2129	$\mathcal{C}_{c_2,2}$	0.6562	2.2948	5.1801
$\mathcal{C}_{c_1,3}$	2.8757	2.9045	2.7788	$\mathcal{C}_{c_2,3}$	0.4181	1.9398	6.0649
$\mathcal{C}_{c_1,4}$	2.5929	2.5608	2.3235	$\mathcal{C}_{c_2,4}$	0.4635	1.5435	6.9346
$\mathcal{C}_{c_1,5}$	2.3338	2.2150	1.8506	$\mathcal{C}_{c_2,5}$	0.7519	1.2260	7.7804
$\mathcal{C}_{c_1,6}$	2.0895	1.8663	1.3593	$\mathcal{C}_{c_2,6}$	1.0803	1.1551	8.5845
$\mathcal{C}_{c_1,7}$	1.8676	1.5327	0.8743	$\mathcal{C}_{c_2,7}$	1.3825	1.3577	9.3195
$\mathcal{C}_{c_1,8}$	1.6925	1.2530	0.5006	$\mathcal{C}_{c_2,8}$	1.6263	1.7050	9.9525
$\mathcal{C}_{c_1,9}$	1.5871	1.0773	0.5442	$\mathcal{C}_{c_2,9}$	1.7887	2.1068	10.4443
$\mathcal{C}_{c_1,10}$	1.5568	1.0363	0.8887	$\mathcal{C}_{c_2,10}$	1.8625	2.5569	10.7505
$\mathcal{C}_{c_1,11}$	1.5857	1.1022	1.2434	$\mathcal{C}_{c_2,11}$	1.8955	3.1089	10.8459
$\mathcal{C}_{c_1,12}$	1.6434	1.2150	1.5344	$\mathcal{C}_{c_2,12}$	1.9785	3.7922	10.7804
$\mathcal{C}_{c_1,13}$	1.7028	1.3306	1.7409	$\mathcal{C}_{c_2,13}$	2.0840	4.5050	10.6734
$\mathcal{C}_{c_1,14}$	1.7463	1.4337	1.8592	$\mathcal{C}_{c_2,14}$	2.0936	5.1003	10.5820
$\mathcal{C}_{c_1,15}$	1.7606	1.5301	1.8964	$\mathcal{C}_{c_2,15}$	2.0218	5.5590	10.4671

4.2.2 Case 2: Design of Multi-Source Dose-Response Experiments

Simulation models

In this case, the simulation models (4.11) and (4.12) are derived from the real experimental bioassay data for therapeutic protein relaxin [58]: There is only one quantitative factor x representing the relaxin dosage in terms of ng/ml, and $x \in [0, 22]$. The one qualitative factor is denoted as z with two categories $\{c_1, c_2\}$. Each category is used to represent a different run of doing the experiment. The vector of factors is written as $\mathbf{w} = (x, z)$. The response y

is intracellular Cyclic adenosine monophosphate (cAMP) measurements by normal human uterine endometrial cells in the presence of relaxin determined by Radioimmunoassay (RIA). The response is measured in the units of pmoles/ml. In this study, it is of particular interest to establish the BMD corresponding to a BMR of 40 (pmoles/ml) for the relaxin endpoint.

Simulation data are generated according to the following true four-parameter logistic models

$$\begin{aligned} Y(x, c_1) &= 1.76 + \frac{127.32 - 1.76}{1 + \exp(1.43 \times (\log(x) - 1.57))} \\ Y(x, c_2) &= 1.90 + \frac{105.56 - 1.90}{1 + \exp(1.59 \times (\log(x) - 1.23))} \end{aligned} \quad (4.11)$$

and the true variance models

$$\text{Var}[\varepsilon(x, c_q)] = 0.2^2 Y(x, c_q)^{2 \times 0.4}; \quad q = 1, 2. \quad (4.12)$$

For a subpopulation c_q ($q = 1, 2$) and at an exposure level x_0 , a random response y_0 is simulated as

$$y_0 = Y(x_0, c_q) + \sqrt{\text{Var}[\varepsilon(x_0, c_q)]} \cdot \epsilon; \quad q = 1, 2, \quad (4.13)$$

where ϵ is a random error generated by a standard normal random generator [52].

Existing data from $Q - 1$ sources

The existing data for run c_1 (Source 1) includes a total of 40 samples, which are allocated following the design in Table 4.2: five evenly-spaced design points are selected over the dose range $\mathfrak{R} = [0, 22]$ with 8 replications at each point. At each dose level x_i ($i = 1, 2, \dots, 40$), the response subject to random errors is generated by plugging the value of x_i and c_1 into (4.11), and the response data are given in Table 4.3.

Table 4.2: The design of experiments of run c_1 for Case 2.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.5	5.5
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
5.5	5.5	5.5	5.5	5.5	5.5	11.0	11.0	11.0	11.0
x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}	x_{27}	x_{28}	x_{29}	x_{30}
11.0	11.0	11.0	11.0	16.5	16.5	16.5	16.5	16.5	16.5
x_{31}	x_{32}	x_{33}	x_{34}	x_{35}	x_{36}	x_{37}	x_{38}	x_{39}	x_{40}
16.5	16.5	22.0	22.0	22.0	22.0	22.0	22.0	22.0	22.0

Table 4.3: The simulation response data of run c_1 for Case 2.

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
1.72	2.04	1.48	1.34	1.38	1.85	1.69	1.61	69.59	70.98
y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	y_{18}	y_{19}	y_{20}
69.94	71.13	70.29	68.65	71.48	70.92	99.22	99.41	97.76	97.09
y_{21}	y_{22}	y_{23}	y_{24}	y_{25}	y_{26}	y_{27}	y_{28}	y_{29}	y_{30}
98.30	99.53	97.12	98.58	112.33	106.81	109.03	110.91	108.18	112.38
y_{31}	y_{32}	y_{33}	y_{34}	y_{35}	y_{36}	y_{37}	y_{38}	y_{39}	y_{40}
106.41	108.38	115.66	115.05	116.05	113.60	114.88	115.92	116.02	112.54

Applying the two-stage design

The two-stage DOE procedure (Figure 4.1) is applied for the efficient collection of dose-response data, aiming at obtaining a good BMD estimate corresponding to the target BMR=40. The inputs of the procedure are given as follows:

Input 1: $\mathfrak{R} = [0, 22]$, the dose range of interest of run c_2 (Source 2).

Input 2: $N = 40$, the total sample size available for run c_2 .

Input 3: $N_I = 24$: the sample size available at Stage I, which implies that the sample size is $N - N_I = 16$ at Stage II.

Input 4: BMR= 40: the pre-specified benchmark response.

The criterion in the design optimization (4.5) is set as $\text{Var}[\widehat{\text{BMD}}(c_2)]$ in this case.

Stage I Following Subsection 4.1.1, in stage I, a sample of $N_I = 40$ observations was obtained for run c_2 (Source 2) and given in Table 4.4: three evenly-spaced distinct design points are selected over the dose range $\mathfrak{R} = [0, 22]$ with 8 replications at each point. At each dose level x_i ($i = 1, 2, \dots, 24$), the response subject to random errors is generated by plugging the value of x_i and c_2 into (4.11), and the response data are obtained as given in Table 4.5.

Table 4.4: The design of experiments in Stage I of run c_2 for Case 2.

x_{41}	x_{42}	x_{43}	x_{44}	x_{45}	x_{46}	x_{47}	x_{48}	x_{49}	x_{50}	x_{51}	x_{52}
0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0	11.0	11.0	11.0
x_{53}	x_{54}	x_{55}	x_{56}	x_{57}	x_{58}	x_{59}	x_{60}	x_{61}	x_{62}	x_{63}	x_{64}
11.0	11.0	11.0	11.0	22.0	22.0	22.0	22.0	22.0	22.0	22.0	22.0

Table 4.5: The simulation response data in Stage I of run c_2 for Case 2.

y_{41}	y_{42}	y_{43}	y_{44}	y_{45}	y_{46}	y_{47}	y_{48}	y_{49}	y_{50}	y_{51}	y_{52}
1.74	2.13	1.93	1.98	2.37	2.06	2.08	1.63	91.99	92.22	92.19	92.17
y_{53}	y_{54}	y_{55}	y_{56}	y_{57}	y_{58}	y_{59}	y_{60}	y_{61}	y_{62}	y_{63}	y_{64}
91.52	91.07	93.52	90.29	100.55	99.69	100.53	99.42	99.58	99.09	101.59	99.13

Stage II The augmented design was determined with the following givens: the existing data of run c_1 (Source 1), Stage-I data of run c_2 (Source 2), the SKQ model fitted from the existing and Stage-I data, and the target BMR= 40.

The resulting augmented design is provided in Table 4.6.

Table 4.6: The augmented design in Stage II of run c_2 for Case 2.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
2.4	2.4	2.4	2.4	2.4	2.4	2.4	2.4
x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
3.7	3.7	3.7	3.7	3.7	3.7	3.7	3.7

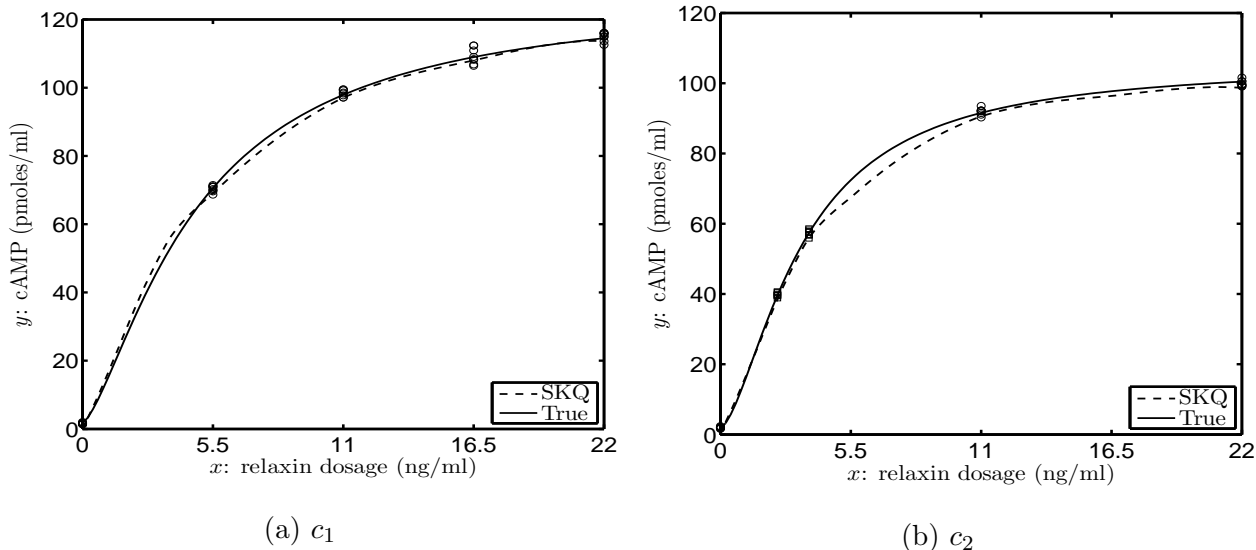


Figure 4.6: The true and SKQ estimated dose-response curves for Case 2.

Following the augmented design (Table 4.6), simulation was performed to obtain the responses, which are given in Table 4.7.

Table 4.7: The simulation response data in Stage II of run c_2 for Case 2.

y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
40.66	38.87	40.44	39.47	39.98	39.61	39.77	39.59
y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}
56.77	56.86	55.88	56.83	57.55	57.82	58.59	58.23

From all the data available for runs c_1 and c_2 (Sources 1 and 2), SKQ modeling was performed (Section 3.3). In Figure 4.6, the fitted dose-responses are depicted as the dashed curves, and the true dose-responses as the solid curves; the existing data for c_1 are represented as circles in Figure 4.6 (a), Stage-I data for c_2 as circles in Figure 4.6 (b), and Stage-II data for c_2 as squares in Figure 4.6 (b).

Comparison with the traditional design

The traditional design of run c_2 for this case is given in (4.14) below.

$$\tilde{\mathfrak{D}}^{\text{Trad}} = \begin{pmatrix} (0, c_2)^\top & (5.5, c_2)^\top & (11, c_2)^\top & (16.5, c_2)^\top & (22, c_2)^\top \\ 8 & 8 & 8 & 8 & 8 \end{pmatrix}. \quad (4.14)$$

The two-stage design and the traditional design are compared based on the results of applying both approaches for 100 times. As a result, 100 sets of data and model fitting results were obtained for each approach. Since the primary goal of this study is to estimate BMD for BMR= 40, the two approaches are compared in terms of their delivered BMDL, the lower bound of the one-sided confidence interval for the BMD.

Specifically, applying the two-stage design 100 times leads to 100 data sets; from each data set, a one-sided 95% CI was constructed following bootstrapping resampling method (Section 3.4). From the 100 data sets, the 100 BMDLs obtained are denoted as

$$\{\text{BMDL}_{\text{Two-stage}}^{(r)}(c_Q); r = 1, 2, \dots, 100\}. \quad (4.15)$$

In the same way, applying the traditional design (4.14) 100 times also results in 100 data sets. Performing the statistical modeling and inference on these data sets respectively leads to 100 BMDLs, which are denoted as

$$\{\text{BMDL}_{\text{Trad}}^{(r)}(c_Q); r = 1, 2, \dots, 100\}. \quad (4.16)$$

The true BMD can be easily calculated from the true dose-response simulation model (4.11), and it is found out to be 2.43, which is represented by the horizontal line in Figure 4.7. In Figure 4.7, the box plots are generated for the BMDLs (4.15) obtained from our two-stage design (on the left) and the BMDLs (4.16) obtained from the traditional design (on the right). Comparing the two boxes in Figure 4.7, it is evident that the BMDLs resulting from

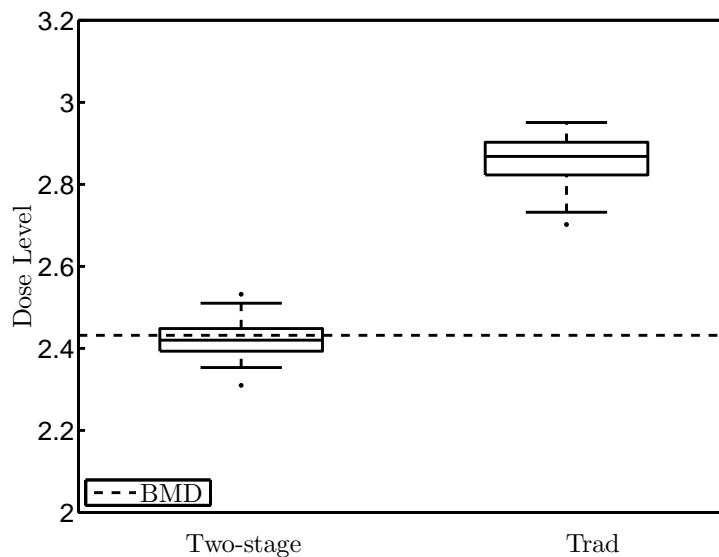


Figure 4.7: The box plots for the BMDLs resulting from the two design methods for Case 2.

our procedure are closer to the true BMD and vary over a narrower dose range. In other words, with the same sample size, collecting a data set following the traditional design leads to a BMDL that may be anywhere between the lower and upper adjacent of the right box plot, whereas collecting a data set following the two-stage design leads to a BMDL that falls within the much narrower and more accurate range of the left box.

Chapter 5

Summary and Future Studies

The focus of this dissertation is on efficient modeling and design of multi-source biological experiments.

First, a new semi-parametric statistical model, SKQ (stochastic kriging with qualitative factors), has been developed. SKQ is the first kriging-based model that is able to take into account the following three types of variability: the variability that stems from quantitative factors, qualitative factors, and uncontrollable sources (random errors). Compared to the parametric MEM (mixed effects modeling) method, the closest regression-based counterpart, SKQ represents a more general modeling approach and is free of the various restrictive assumptions stipulated by MEM.

Through the empirical simulation studies, the modeling efficiency of SKQ is demonstrated to be superior to the existing methods, i.e., SK (standard stochastic kriging) and MEM. SKQ is able to pool information across multiple data sources, to accommodate general data features, and to provide more informative estimation/inference from given data. For clarity and succinctness of the presentation, the two cases for this dissertation were designed to involve a relatively small number of data sources (subpopulations). It is worth noting that when applying SKQ to data from a large, as opposed to a small, number of sources, there is hardly any additional theoretical or implementation hurdles. In our empirical experience, SKQ's modeling efficiency (thanks to its information pooling ability) is more pronounced with more sources of data.

Second, SKQ's ability to provide more informative statistical inference from limited data certainly opens up opportunities for efficient design of experiments (DOE): Given some subsets of data already collected and likely to be from different sources, how to efficiently design the next-stage biological experiments so that all the integrated data is most informative? DOE is directly associated with model estimation and inference, a SKQ model-based DOE method is developed to achieve experimental efficiency. The experimental design procedure has been evaluated through two simulation cases designed to mimic real toxicity data. It has been shown that the design resulting from our procedure is superior to the "naive" traditional designs.

The SKQ developed assumes stationarity of the extrinsic variability. An extension of the SQK could be to allow the extrinsic variance to be non-stationary [59–62], and to accommodate "global" Gaussian processes. Based on the extended global SKQ, the design of experiments methods can be extended accordingly.

Chapter 6

Appendix

6.1 Review of Standard Stochastic Kriging (SK)

Standard stochastic kriging (SK), which was first introduced by Ankenman et al.[23], models the dependence of a continuous response upon the quantitative factors \mathbf{x} only. With no qualitative factors involved, the response from the j^{th} replication (animal subject) is written in terms of \mathbf{x} as follows

$$\mathcal{Y}_j(\mathbf{x}) = Y(\mathbf{x}) + \varepsilon_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + M(\mathbf{x}) + \varepsilon_j(\mathbf{x}), \quad (6.1)$$

where $Y(\mathbf{x})$ represents the true expected response at the factor setting $\mathbf{x} = (x_1, x_2, \dots, x_d)$, with $\mathbf{x} \in \mathbb{R}^d$.

The expectation $Y(\mathbf{x})$ is decomposed into two parts: $Y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + M(\mathbf{x})$. $\mathbf{f}(\mathbf{x})$ is a vector of known functions of \mathbf{x} , and $\boldsymbol{\beta}$ is a vector of unknown parameters of compatible dimension. Since it has been reported that $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta_0$ (that is, just a constant term) suffices for most applications [23], this work adopts $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta_0$ unless stated otherwise. The term M represents a realization of a mean-zero stationary Gaussian random field, and can be considered as being randomly sampled from a space of functions mapping $\mathbb{R}^d \rightarrow \mathbb{R}$; the functions in this space are assumed to exhibit spatial correlation, and thus $M(\mathbf{x})$ and $M(\mathbf{x}')$ will tend to be similar if \mathbf{x} and \mathbf{x}' are close to each other in the space. As in Ankenman et al. (2010)[23], the stochastic nature of $M(\mathbf{x})$ is referred to as extrinsic variability.

The random noise $\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \dots$ at a factor setting \mathbf{x} has zero mean, and is independent and identically distributed across replications (animal subjects). The error variance $\text{Var}[\varepsilon(\mathbf{x})]$ is allowed to be dependent on \mathbf{x} . The randomness of $\varepsilon(\mathbf{x})$ is referred to as intrinsic variability.

A data set, on which SK is to be applied, consists of $n(\mathbf{x}_i)$ replications taken at design point \mathbf{x}_i ($i = 1, 2, \dots, I$), with I representing the number of distinct design points (factor settings) in the data. The paired data can be represented as $\{(\mathbf{x}_i, \mathcal{Y}_j(\mathbf{x}_i)); i = 1, 2, \dots, I; j = 1, 2, \dots, n(\mathbf{x}_i)\}$. The sample average of the responses at \mathbf{x}_i across the $n(\mathbf{x}_i)$ replications is given by:

$$\bar{\mathcal{Y}}(\mathbf{x}_i) = \frac{1}{n(\mathbf{x}_i)} \sum_{j=1}^{n(\mathbf{x}_i)} \mathcal{Y}_j(\mathbf{x}_i) = \beta_0 + \mathbf{M}(\mathbf{x}_i) + \frac{1}{n(\mathbf{x}_i)} \sum_{j=1}^{n(\mathbf{x}_i)} \varepsilon_j(\mathbf{x}_i).$$

Denote

$$\bar{\mathcal{Y}} = (\bar{\mathcal{Y}}(\mathbf{x}_1), \bar{\mathcal{Y}}(\mathbf{x}_2), \dots, \bar{\mathcal{Y}}(\mathbf{x}_I))^{\top} \quad (6.2)$$

as the $I \times 1$ vector of sample average responses at the I distinct design points.

Similarly, the vector of sample average errors is denoted as

$$\bar{\varepsilon} = (\bar{\varepsilon}(\mathbf{x}_1), \bar{\varepsilon}(\mathbf{x}_2), \dots, \bar{\varepsilon}(\mathbf{x}_I))^{\top}, \quad (6.3)$$

with $\bar{\varepsilon}(\mathbf{x}_i) = n(\mathbf{x}_i)^{-1} \sum_{j=1}^{n(\mathbf{x}_i)} \varepsilon_j(\mathbf{x}_i), i = 1, 2, \dots, I$.

6.1.1 The Extrinsic and Intrinsic Variance Structures

The key of SK lies in the modeling of extrinsic as well as intrinsic variability, which are presented respectively as follows.

Denote $\mathbf{x} = (x_1, x_2, \dots, x_d)^{\top}$ and $\mathbf{x}' = (x'_1, x'_2, \dots, x'_d)^{\top}$ as two vectors of the quantitative factors. For a stationary Gaussian process $\mathbf{M}(\mathbf{x})$, the covariance function can be represented as

$$\text{Cov}[\mathbf{M}(\mathbf{x}), \mathbf{M}(\mathbf{x}')] = \delta^2 \cdot K(\mathbf{x}, \mathbf{x}') \quad (6.4)$$

where $\delta^2 > 0$ denotes the variance of the Gaussian process, and $K(\mathbf{x}, \mathbf{x}')$ the correlation between $\mathbf{M}(\mathbf{x})$ and $\mathbf{M}(\mathbf{x}')$.

For the estimation of a SK model, a certain functional structure needs to be imposed on the correlation $K(\mathbf{x}, \mathbf{x}')$. A range of choices are available in the literature (e.g., Santner et al. 2003 [43]; Qian et al. 2008 [41]), and one of the most popular structures in practice is the exponential correlation function

$$K(\mathbf{x}, \mathbf{x}') = \exp \left\{ \sum_{h=1}^d -\theta_h |x_h - x'_h|^p \right\}. \quad (6.5)$$

In (6.5), $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ is a vector of unknown parameters. It is required that $\theta_h > 0$ ($h = 1, 2, \dots, d$), and $\boldsymbol{\theta}$ determines the roughness of the response surface. The parameter $p \in (0, 2]$ also needs to be estimated unless p is prespecified as 2, which leads to the widely-used quadratic correlation function [44].

With a selected correlation function such as (6.5), the $I \times I$ variance-covariance matrix $\Sigma_{\mathbf{M}}$ is constructed as follows for a data set including I distinct design points

$$\Sigma_{\mathbf{M}} = \delta^2 \mathbf{R}(\boldsymbol{\theta}) = \delta^2 \begin{pmatrix} 1 & K(\mathbf{x}_1, \mathbf{x}_2) & \dots & K(\mathbf{x}_1, \mathbf{x}_I) \\ K(\mathbf{x}_2, \mathbf{x}_1) & 1 & \dots & K(\mathbf{x}_2, \mathbf{x}_I) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{x}_I, \mathbf{x}_1) & K(\mathbf{x}_I, \mathbf{x}_2) & \dots & 1 \end{pmatrix}. \quad (6.6)$$

In (6.6), $\mathbf{R}(\boldsymbol{\theta})$ represents the correlation matrix, with each component being a correlation function of unknown parameters $\boldsymbol{\theta}$. For an arbitrary setting \mathbf{x}_0 , the $I \times 1$ vector $\Sigma_{\mathbf{M}}(\mathbf{x}_0, \cdot)$ is defined as

$$\Sigma_{\mathbf{M}}(\mathbf{x}_0, \cdot) = \delta^2 \mathbf{v}(\mathbf{x}_0, \boldsymbol{\theta}) = \delta^2 \begin{pmatrix} K(\mathbf{x}_0, \mathbf{x}_1) \\ K(\mathbf{x}_0, \mathbf{x}_2) \\ \vdots \\ K(\mathbf{x}_0, \mathbf{x}_I) \end{pmatrix}, \quad (6.7)$$

where $\mathbf{v}(\mathbf{x}_0, \boldsymbol{\theta})$ is a correlation vector with each component dependent on \mathbf{x}_0 and the unknown parameter $\boldsymbol{\theta}$.

The intrinsic variance of the random response at \mathbf{x}_i ($i = 1, 2, \dots, I$) is denoted as $\text{Var}[\varepsilon(\mathbf{x}_i)]$. Let Σ_ε be the $I \times I$ variance-covariance matrix of vector $\bar{\varepsilon}$, which is defined in (6.3). Under the i.i.d assumption for random errors, Σ_ε is a diagonal matrix

$$\Sigma_\varepsilon = \text{diag}\{\text{Var}[\varepsilon(\mathbf{x}_1)]/n(\mathbf{x}_1), \text{Var}[\varepsilon(\mathbf{x}_2)]/n(\mathbf{x}_2), \dots, \text{Var}[\varepsilon(\mathbf{x}_I)]/n(\mathbf{x}_I)\}. \quad (6.8)$$

6.1.2 Estimation and Inference by Standard Stochastic Kriging

Recall that the random response can be written as

$$\mathcal{Y}_j(\mathbf{x}) = \beta_0 + \mathbf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x}), \quad (6.9)$$

As stated in Ankenman et al. (2010) [23], the SK-based modeling and inference requires the following assumption.

Assumption 2 *The random field \mathbf{M} is a stationary Gaussian random field; and $\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \dots$ are i.i.d. $N(0, \text{Var}[\varepsilon(\mathbf{x})])$, independent of $\varepsilon_j(\mathbf{x}')$ for all j and $\mathbf{x} \neq \mathbf{x}'$, and independent of \mathbf{M} .*

The assumption of \mathbf{M} being a stationary Gaussian random field is standard for kriging (Santner et al. 2003, [43]), and it implies that the random vector $(\mathbf{M}(\mathbf{x}_1), \mathbf{M}(\mathbf{x}_2), \dots, \mathbf{M}(\mathbf{x}_I))^\top$ follows a multivariate normal distribution with constant marginal mean 0, variance δ^2 , and correlation matrix \mathbf{R} as in (6.6). The assumption for the random errors allows the variance $\text{Var}[\varepsilon(\mathbf{x})]$ to be dependent on \mathbf{x} .

For a data set $\{(\mathbf{x}_i, \mathcal{Y}_j(\mathbf{x}_i)); i = 1, 2, \dots, I; j = 1, 2, \dots, n(\mathbf{x}_i)\}$, and under Assumption 2, the log-likelihood function of the unknown parameters $(\beta_0, \delta^2, \boldsymbol{\theta})$ is

$$\ln \mathcal{L}(\beta_0, \delta^2, \boldsymbol{\theta}) = -\ln[(2\pi)^{I/2}] - \frac{1}{2} \ln[|\delta^2 \mathbf{R}(\boldsymbol{\theta}) + \Sigma_\varepsilon|] - \frac{1}{2} (\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_I)^\top [\delta^2 \mathbf{R}(\boldsymbol{\theta}) + \Sigma_\varepsilon]^{-1} (\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_I), \quad (6.10)$$

where $\mathbf{1}_I$ is a $(I \times 1)$ vector of ones.

Ankenman et al. (2010) [23] summarizes the SK modeling steps as follows.

1. Obtain the estimated Σ_ε :

$$\widehat{\Sigma}_\varepsilon = \text{diag}\{\widehat{\text{Var}}[\varepsilon(\mathbf{x}_1)]/n(\mathbf{x}_1), \widehat{\text{Var}}[\varepsilon(\mathbf{x}_2)]/n(\mathbf{x}_2), \dots, \widehat{\text{Var}}[\varepsilon(\mathbf{x}_I)]/n(\mathbf{x}_I)\} \quad (6.11)$$

where

$$\widehat{\text{Var}}[\varepsilon(\mathbf{x}_i)] = \frac{1}{n(\mathbf{x}_i) - 1} \sum_{j=1}^{n(\mathbf{x}_i)} (\mathcal{Y}_j(\mathbf{x}_i) - \bar{\mathcal{Y}}(\mathbf{x}_i))^2, \quad i = 1, 2, \dots, I. \quad (6.12)$$

2. Using $\widehat{\Sigma}_\varepsilon$ instead of Σ_ε , maximize the log-likelihood (6.10) over $(\widehat{\beta}_0, \widehat{\delta}^2, \widehat{\boldsymbol{\theta}})$.
3. Estimate the expected response $\Upsilon(\mathbf{x}_0)$ by

$$\widehat{\Upsilon}(\mathbf{x}_0) = \widehat{\beta}_0 + \widehat{\delta}^2 \mathbf{v}(\mathbf{x}_0, \widehat{\boldsymbol{\theta}})^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}) + \widehat{\Sigma}_\varepsilon]^{-1} (\bar{\mathcal{Y}} - \widehat{\beta}_0 \mathbf{1}_I), \quad (6.13)$$

where $(\widehat{\beta}_0, \widehat{\delta}^2, \widehat{\boldsymbol{\theta}})$ are obtained from the previous step. The mean squared error (MSE) is estimated as

$$\widehat{\text{MSE}}[\widehat{\Upsilon}(\mathbf{x}_0)] = \widehat{\delta}^2 - \widehat{\delta}^4 \mathbf{v}(\mathbf{x}_0, \widehat{\boldsymbol{\theta}})^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}) + \widehat{\Sigma}_\varepsilon]^{-1} \mathbf{v}(\mathbf{x}_0, \widehat{\boldsymbol{\theta}}) + \eta^\top \eta (\mathbf{1}_I^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}) + \widehat{\Sigma}_\varepsilon]^{-1} \mathbf{1}_I)^{-1} \quad (6.14)$$

where $\eta = 1 - \mathbf{1}_I^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}) + \widehat{\Sigma}_\varepsilon]^{-1} \mathbf{v}(\mathbf{x}_0, \widehat{\boldsymbol{\theta}}) \widehat{\delta}^2$.

The $(1 - \alpha) \times 100\%$ confidence interval for $Y(\mathbf{x}_0)$ is

$$\hat{Y}(\mathbf{x}_0) \pm z_{(1-\alpha)/2} \sqrt{\widehat{\text{MSE}}[\hat{Y}(\mathbf{x}_0)]} \quad (6.15)$$

where $z_{(1-\alpha)/2}$ is the upper $(1 - \alpha)/2$ critical value for standard normal distribution.

6.2 True Expected Exposure-Response Model for Case 1

As part of the simulation model for Case 1 (Subsection 3.5.1), the following two models (6.16-6.17) provide the true expected response surfaces for the two subpopulations (short and long TiO₂ nanobelts) respectively. Both of the models take the form of a single-hidden layer feedforward network (SLFN)[51, 63], and are estimated from the real dose-time-response data in NIOSH's in-vivo study of TiO₂ nanoparticles [50].

$$\begin{aligned} Y(\mathbf{x}, c_1) = & 67.55 + \frac{0.14}{\exp(-12.85 + 0.41x_1 + 0.06x_2)} + \frac{-33.23}{\exp(-1.39 + 0.12x_1 - 0.02x_2)} \quad (6.16) \\ & + \frac{-0.08}{\exp(-8.57 + 1.22x_1 - 0.2x_2)} + \frac{-76.00}{\exp(-0.21 + 0.06x_1 - 0.01x_2)} \\ & + \frac{-3.49}{\exp(-0.40 + 0.20x_1 - 0.04x_2)} + \frac{-76.99}{\exp(-0.51 - 0.05x_1 + 0.03x_2)} \\ & + \frac{5.15}{\exp(-5.52 + 0.16x_1 + 0.04x_2)} + \frac{-0.03}{\exp(-25.89 + 2.99x_1 + 0.28x_2)} \\ & + \frac{32.77}{\exp(-0.62 + 0.05x_1 + 0.03x_2)} + \frac{1.09}{\exp(5.76 + 0.26x_1 - 0.07x_2)} \end{aligned}$$

$$\begin{aligned} Y(\mathbf{x}, c_2) = & 71.74 + \frac{-104.84}{\exp(-0.66 + 0.09x_1 - 0.02x_2)} + \frac{-1.18}{\exp(7.66 - 0.45x_1 - 0.01x_2)} \quad (6.17) \\ & + \frac{0.04}{\exp(28.01 - 3.62 + 0.24x_2)} + \frac{-4.50}{\exp(5.02 - 0.16x_1 - 0.04x_2)} \\ & + \frac{82.50}{\exp(0.74 + 0.04x_1 - 0.03x_2)} + \frac{-27.73}{\exp(0.83 + 0.13x_1 - 0.02x_2)} \\ & + \frac{35.84}{\exp(-1.18 + 0.04x_1 + 0.03x_2)} + \frac{-13.58}{\exp(-0.85 - 0.18x_1 + 0.03x_2)} \\ & + \frac{-2.61}{\exp(0.32 - 0.27x_1 + 0.01x_2)} + \frac{2.59}{\exp(-8.86 - 0.05x_1 + 0.08x_2)}. \end{aligned}$$

6.3 An Estimation Data Set (EDS) for Case 2

Table 6.1 shows an estimation data set (EDS) randomly generated from the simulation models of Case 2 (Subsection 3.5.2). The first column provides the 15 distinct design points, with each one specified in terms of (x, c_q) , the dosage x and subpopulation c_q ($q = 1, 2, 3$). At each design point, 8 replications were obtained corresponding to the 8 columns of the responses in the table.

Table 6.1: An estimation data set (EDS) for Case 2: modeling the multi-source dose-response data

$\mathbf{w} = (x, c_q)$	Response							
$(0, c_1)$	19.606	19.630	19.059	19.119	19.799	19.197	21.341	21.641
$(5, c_1)$	30.715	29.792	28.982	33.335	31.580	31.415	28.696	32.305
$(10, c_1)$	52.016	49.630	50.397	49.462	57.212	48.896	50.396	50.329
$(15, c_1)$	76.507	76.149	80.295	77.499	73.711	72.691	78.404	73.824
$(20, c_1)$	121.624	121.009	120.366	120.168	126.850	118.587	117.079	127.984
$(0, c_2)$	20.476	19.672	19.071	19.832	21.288	19.604	20.582	19.601
$(5, c_2)$	28.907	26.911	31.777	28.914	31.867	30.150	31.717	32.173
$(10, c_2)$	48.531	52.631	51.625	54.031	54.321	49.066	49.427	52.142
$(15, c_2)$	81.604	82.863	79.106	80.348	84.498	87.584	82.051	78.335
$(20, c_2)$	134.977	123.213	133.363	125.405	131.264	130.982	132.469	132.215
$(0, c_3)$	23.819	27.994	26.262	26.843	25.228	27.381	25.827	24.452
$(5, c_3)$	38.927	35.478	36.313	40.699	39.396	38.104	37.377	38.829
$(10, c_3)$	57.377	54.406	54.933	55.961	62.745	52.841	58.847	53.044
$(15, c_3)$	82.621	81.036	78.257	81.493	78.597	80.067	80.918	79.754
$(20, c_3)$	114.988	113.588	118.589	117.208	121.077	114.184	114.511	119.074

References

- [1] Mingjia Zhi, Ayyakkannu Manivannan, Fanke Meng, and Nianqiang Wu. Highly conductive electrospun carbon nanofiber/mno₂ coaxial nano-cables for high energy and power density supercapacitors. *Journal of Power Sources*, 208:345–353, 2012.
- [2] Chengcheng Xiang, Ming Li, Mingjia Zhi, Ayyakkannu Manivannan, and Nianqiang Wu. Reduced graphene oxide/titanium dioxide composites for supercapacitor electrodes: shape and coupling effects. *Journal of Materials Chemistry*, 22:19161–19167, 2012.
- [3] Ming Li, Qiaoyi Wang, Xiaodong Shi, Lawrence A. Hornak, and Nianqiang Wu. Detection of mercury(ii) by quantum dot-dna-gold nanoparticle ensemble based nanosensor via nanometal surface energy transfer. *Analytical Chemistry*, 83:7061–7065, 2011.
- [4] Nianqiang Wu, Minhua Zhao, Jian-Guo Zheng, Chuanbin Jiang, Ben Myers, Shuoxin Li, Minking Chyu, and Scott X. Mao. Porous cuo–zno nanocomposite for sensing electrode of high-temperature co solid-state electrochemical sensor. *Nanotechnology*, 16:2878–2881, 2005.
- [5] Nianqiang Wu, Zheng Chen, Jianhui Xu, Minking Chyu, and Scott X. Mao. Impedance-metric pt/ysz/au-ga₂o₃ sensors for co detection at high-temperature. *Sensors & Actuators B*, 110:49–53, 2005.
- [6] Ming Li, Jianming Zhang, Savan Suri, Letha J. Sooter, Dongling Ma, and Nianqiang Wu. Detection of adenosine triphosphate with an aptamer biosensor based on surface-enhanced raman scattering. *Analytical Chemistry*, 84:2837–2842, 2012.
- [7] Na Ren, Rui Li, Limei Chen, Guancong Wang, Duo Liu, Yingjun Wang, Lin Zheng, Wei Tang, Xiaoqiang Yu, Huaidong Jiang, Hong Liu, and Nianqiang Wu. In-situ construction of titanate/silver nanoparticle/titanate sandwich nanostructure on metallic titanium surface for bacteriostatic and biocompatible implants. *Journal of Materials Chemistry*, 22:19151–19160, 2012.
- [8] L. Edler, K. Poirier, M. Dourson, J. Kleiner, B. Mileson, H. Nordmann, A. Renwickf, W. Slobg, K. Waltonf, and G. Wurtzenh. Mathematical modelling and quantitative methods. *Food and Chemical Toxicology*, 40:283–326, 2002.

- [9] Frank Bretz, Jason Hsu, Jose Pinheiro, and Yi Liu. Dose finding—a challenge in statistics. *Biometrical Journal*, 50(4):480–504, 2008.
- [10] US EPA. *Benchmark dose technical guidance*. Washington, DC: U.S. Environmental Protection Agency, 2012.
- [11] John A. Bailer, Michael R. Hughes, Debra L. Denton, and James T. Oris. An empirical comparison of effective concentration estimators for evaluating aquatic toxicity test responses. *Environmental Toxicology and Chemistry*, 19(1):141–150, 2000.
- [12] Robert E. Kass and Duane Steffey. Approximate bayesian inference in conditionally independent hierarchical models (parametric empirical bayes models). *Journal of the American Statistical Association*, 84(407):717–726, 1989.
- [13] NIOSH. *Current Intelligence Bulletin 65 Occupational Exposure to Carbon Nanotubes and Nanofibers*. Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, April 2013.
- [14] Wout Slob. Dose-response modeling of continuous endpoints. *Toxicological Sciences*, 66:298–312, 2002.
- [15] Salomon Sand, Dietrich von Rosen, Per Eriksson, Anders Fredriksson, Henrik Viberg, Katarina Victorin, and Agneta Falk Filipsson. Dose-response modeling and benchmark calculations from spontaneous behavior data on mice neonatally exposed to 2,2',4,4',5-pentabromodiphenyl ether. *Toxicological Sciences*, 81:491–501, 2004.
- [16] Günter Oberdörster, Eva Oberdörster, and Jan Oberdörster. Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles. *Environmental Health Perspectives*, 113(7):823–839, 2005.
- [17] George M. Woodall, Jeffrey S. Gift, and Gary L. Foureman. Empirical methods and default approaches in consideration of exposure duration in dose–response relationships. *General, Applied and Systems Toxicology*.
- [18] Kai Sun, Gary F. Krause, Foster L. Mayer, Mark R. Ellersieck, and Asit P. Basu. Estimation of acute toxicity by fitting a dose-time-response surface. *Risk Analysis*, 15(2):247–252, 1995.
- [19] U.S.EPA. Benchmark dose software, 2012. Version 2.3.1.
- [20] G. Dunn. Optimal designs for drug, neurotransmitter and hormone receptor assays. *Statistics in Medicine*, 7(7):805–815, 1988.

- [21] Xi Chen, Bruce E. Ankenman, and Barry L. Nelson. Enhancing stochastic kriging metamodels with gradient estimators. *Operations Research*, 61(2):512–528, 2013.
- [22] Xi Chen, Bruce E. Ankenman, and Barry L. Nelson. The effects of common random numbers on stochastic kriging metamodels. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(2):7, 2012.
- [23] Bruce Ankenman, Barry L. Nelson, and Jeremy Staum. Stochastic kriging for simulation metamodeling. *Operations research*, 58(2):371–382, 2010.
- [24] GAF Seber and CJ Wild. *Nonlinear regression*. Wiley, New York, 2005.
- [25] Bernard E. Rollin. The regulation of animal research and the emergence of animal ethics: A conceptual history. *Theoretical Medicine and Bioethics*, 27:285–304, 2006.
- [26] Simon Festing and Robin Wilkinson. The ethics of animal research. talking point on the use of animals in scientific research. *The European Molecular Biology Organization Report*, 8:526–530, 2007.
- [27] Qi Zeng and Marie Davidian. Bootstrap-adjusted calibration confidence intervals for immunoassay. *Journal of the American Statistical Association*, 92(437):278–290, 1997.
- [28] Marie Davidian and David M. Giltinan. *Nonlinear models for repeated measurement data*. Chapman & Hall, 1995.
- [29] David M. Giltinan and Marie Davidian. Assays for recombinant proteins: A problem in non-linear calibration. *Statistics in Medicine*, 13(11):1165–1179, 1994.
- [30] Jean-Louis Steimer, Alain Mallet, Jean-Louis Golmard, and Jean-François Boisvieux. Alternative approaches to estimation of population pharmacokinetic parameters: comparison with the nonlinear mixed-effect model. *Drug metabolism reviews*, 15(1-2):265–292, 1984.
- [31] Jerome Sacks and Susannah Schiller. Spatial designs. *Statistical decision theory and related topics IV*, 2(S S):385–399, 1988.
- [32] Jeremy Oakley and Anthony O’Hagan. Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4):769–784, 2002.
- [33] Jerome Sacks, Susannah B. Schiller, and William J. Welch. Designs for computer experiments. *Technometrics*, 31(1):41–47, 1989.

- [34] Peter ZG Qian. Sliced latin hypercube designs. *Journal of the American Statistical Association*, 107(497):393–399, 2012.
- [35] Cedric J. Sallaberry, Jon C. Helton, and Stephen C. Hora. Extension of latin hypercube samples with correlated variables. *Reliability Engineering & System Safety*, 93(7):1047–1059, 2008.
- [36] Williamson Daniel. Exploratory ensemble designs for environmental models using k-extended latin hypercubes.
- [37] Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. Design and analysis of computer experiments. *Statistical science*, 4(4):409–423, 1989.
- [38] J.W. Van Groenigen. The influence of variogram parameters on optimal sampling schemes for mapping by kriging. *Geoderma*, 97(3):223–236, 2000.
- [39] Jason L. Loepky, Leslie M. Moore, and Brian J. Williams. Batch sequential designs for computer experiments. *Journal of Statistical Planning and Inference*, 140(6):1452–1464, 2010.
- [40] Xi Chen and Qiang Zhou. Sequential experimental designs for stochastic kriging. *Proceedings of the 2014 Winter Simulation Conference*, 2014.
- [41] Peter Z.G. Qian, Huaiqing Wu, and C.F. Jeff Wu. Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics*, 50(3), 2008.
- [42] Gang Han, Thomas J. Santner, William I. Notz, and Donald L. Bartel. Prediction for computer experiments having quantitative and qualitative input variables. *Technometrics*, 51(3), 2009.
- [43] Thomas J. Santner, Brian J. Williams, and William Notz. *The design and analysis of computer experiments*. Springer, 2003.
- [44] Carl Edward Rasmussen and K. I. Williams Christopher. Gaussian processes for machine learning. 2006.
- [45] Qiang Zhou, Peter Z.G. Qian, and Shiyu Zhou. A simple approach to emulation for computer models with qualitative and quantitative factors. *Technometrics*, 53(3), 2011.
- [46] Agneta Falk Filipsson, Salomon Sand, John Nilsson, and Katarina Victorin. The benchmark dose method-review of available models and recommendations for application in health risk assessment. *Critical Reviews in Toxicology*, 33(5):505–542, 2003.

- [47] Hastie Trevor, Tibshirani Robert, and Jerome Friedman. *The elements of statistical learning*, volume 1. Springer New York, 2001.
- [48] Paul D.W. Kirk and Michael P.H. Stumpf. Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data. *Bioinformatics*, 25(10):1300–1306, 2009.
- [49] Anthony Christopher Davison and David V. Hinkley. *Bootstrap methods and their application*, volume 1. Cambridge university press, 1997.
- [50] Dale W. Porter, Nianqiang Wu, Ann F. Hubbs, Robert R. Mercer, Kathleen Funk, Fanke Meng, Jiangtian Li, Michael G. Wolfarth, Lori Battelli, Sherri Friend, Andrew Michael, Hamilton Jr Raymond, Sriram Krishnan, Yang Feng, Castranova Vincent, and Holian Andrij. Differential mouse pulmonary dose and time course responses to titanium dioxide nanospheres and nanobelts. *Toxicological Sciences*, 131(1):179–193, 2013.
- [51] Feng Yang. Neural network metamodeling for cycle time-throughput profiles in manufacturing. *European Journal of Operational Research*, 205(1):172–185, 2010.
- [52] David Kelton and Averill Law. *Simulation modeling and analysis*. McGraw Hill Boston, MA, 2000.
- [53] Christopher Jennison and Bruce W. Turnbull. *Group sequential methods: applications to clinical trials*. Chapman & Hall/CRC, 1999.
- [54] Michael D. McKay, Richard J. Beckman, and William J. Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.
- [55] Kenny Q. Ye. Orthogonal column latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association*, 93(444):1430–1439, 1998.
- [56] Rudolf J. Freund, Donna Mohr, and William Wilson. *Statistical methods*. Academic Press, 2010.
- [57] Robert McGill, John W. Tukey, and Wayne A. Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, 1978.
- [58] Marie Davidian and David M. Giltinan. Some general estimation methods for nonlinear mixed-effects model. *Journal of Biopharmaceutical Statistics*, 3(1):23–55, 1993.

- [59] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H. Liu. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998.
- [60] Ying Xiong, Wei Chen, Daniel Apley, and Xuru Ding. A non-stationary covariance-based kriging method for metamodelling in engineering design. *International Journal for Numerical Methods in Engineering*, 71(6):733–756, 2007.
- [61] Alexandra M. Schmidt and Anthony O’Hagan. Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(3):743–758, 2003.
- [62] Hyoungh-Moon Kim, Bani K. Mallick, and CC Holmes. Analyzing nonstationary spatial data using piecewise gaussian processes. *Journal of the American Statistical Association*, 100(470):653–668, 2005.
- [63] Guangbin Huang, Yanqiu Chen, and Haroon A. Babri. Classification ability of single hidden layer feedforward neural networks. *Neural Networks, IEEE Transactions on*, 11(3):799–801, 2000.