ACS Sustainable Chemistry & Engineering

Research Article

pubs.acs.org/journal/ascecg

Open Access on 05/20/2015

A New Stochastic Kriging Method for Modeling Multi-Source Exposure–Response Data in Toxicology Studies

Kai Wang,[†] Xi Chen,[‡] Feng Yang,*[,†] Dale W. Porter,[§] and Nianqiang Wu[∥]

[†]Department of Industrial and Management System Engineering, West Virginia University, Morgantown, West Virginia 26506, United States

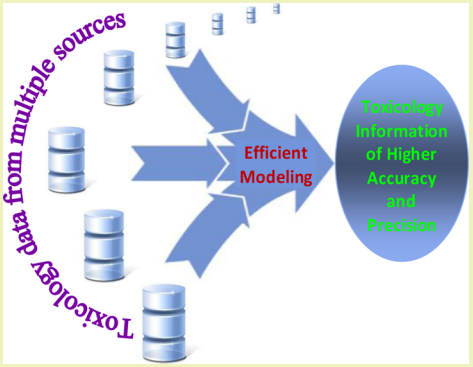
[‡]Department of Statistical Sciences and Operations Research, Richmond, Virginia 23284, United States

[§]National Institute for Occupational Safety and Health (NIOSH), Morgantown, West Virginia 26505, United States

[∥]Department of Mechanical and Aerospace Engineering, West Virginia University, Morgantown, West Virginia 26506, United States

Supporting Information

ABSTRACT: 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, 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. SKQ fits into the mosaic of efficient decision-making methods for assessing the risk of a tremendously large variety of nanomaterials and helps to alleviate safety concerns regarding the enormous amount of new nanomaterials.



KEYWORDS: *Exposure–response, Multi-source data, Stochastic kriging, Gaussian process, Nanotoxicology*

INTRODUCTION

Nanomaterials (NM) are finding wide applications in areas such as the 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 for conducting expensive and time-consuming biological experiments, there is an urgent need to develop efficient decision-making methods for NM risk assessment. Primarily motivated by such a need, this paper 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 given 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[11,12]), 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: dose level of the substance of interest and 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. On the basis of the experimental data collected, statistical methods are then used to fit exposure–response models quantifying the relationships of interest.

However, it remains a challenge to achieve the exposure–response models of the highest quality from given toxicology

Special Issue: Sustainable Nanotechnology 2013

Received: February 16, 2014
Revised: May 13, 2014
Published: May 20, 2014

   dx.doi.org/10.1021/sc500102h | ACS Sustainable Chem. Eng. 2014, 2, 1581−1591

data. Existing statistical models[13−15] 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,[16] animal providers,[17] and shapes of nanomaterials.[18] The observed response of animals depends not only on the exposure condition but also on these source factors. The subset of data collected from a different source may well reflect a different exposure−response profile. (ii) Second, the exposure−response relationship may well be nonlinear.[19−23] For two-dimensional dose−response curves, a range of nonlinear regression models (e.g., power and logistic models) has been developed[13] 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.[24] 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 difficulty in making statistical inferences for the estimated exposure−response model (i.e., quantifying the uncertainty of the model estimates).

To address these challenges, a new stochastic kriging model is developed in this work, which will be referred to as SKQ (stochastic kriging with qualitative factors). SKQ particularly aims at pooling information from all sources of data to obtain the highest-quality models. It is highly flexible and able to accurately approximate practically any continuous response surfaces,[15,25,26] without requiring a preassumed functional form (e.g., logistic model) as traditional nonlinear regression does.[27] SKQ is able to accommodate variance heterogeneity across exposure conditions as well as different data sources and is able to provide valid statistical inference.

In the literature, the majority of exposure−response modeling methods are restricted to single-source data.[28−30] These methods separately model each source of data and make no effort at all to pool information across different sources. The drawback of such no-pooling methods is that for each data source, a large sample size is required to obtain an adequate exposure−response profile. The mixed effects modeling (MEM) method[14,29,31] represents the closest existing approach for information pooling. (A brief review of MEM is given in Review of Mixed Effects Model, Supporting Information.) However, MEM is parametric regression-based and is subject to restrictive assumptions such as the identification of a specific functional form for regression and the prior-assumed common variance structure for random errors across different data sources. Thus, MEM falls short of addressing challenges (ii) and (iii) stated earlier in this section, whereas SKQ is a more general modeling method.

The remainder of this paper is organized as follows. The Statement of the Research Problem section describes in precise terms the research problem of modeling multi-source exposure−response data in toxicology studies. The new modeling method SKQ is detailed in the Stochastic Kriging with Qualitative Factors (SKQ) section. In the Empirical Studies section, through simulation-based empirical studies, SKQ's modeling efficiency via information pooling is demonstrated, and the advantages of SKQ over two existing methods are also illustrated.

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 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, provider for the experimental animals, and shape of the nanomaterials. The qualitative factors are denoted by the vector $\mathbf{z}$.

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

$$\mathcal{Y}(\mathbf{w}) = \mathrm{E}[\mathcal{Y}(\mathbf{w})] + \varepsilon(\mathbf{w}) = \mathbf{Y}(\mathbf{w}) + \varepsilon(\mathbf{w}) \quad (1)$$

where $\mathbf{Y}(\mathbf{w}) = E[\mathcal{Y}(\mathbf{w})]$ represents the true expected response, and $\varepsilon(\mathbf{w})$ is 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 Case 1: Modeling Multi-Source Dose-Time-Response Data, where the toxicity of TiO_2 nanoparticles (NPs) is investigated. In that case, the vector $\mathbf{x} = (x_1,x_2)^T$ includes two quantitative factors: dosage of NPs x_1 and post-exposure time x_2. There is one qualitative factor z, which has two category levels for the shape of TiO_2 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 z are denoted as $\{c_q; q = 1, 2,...,Q\}$. In the case of TiO_2 NPs, there are $Q = 2$ subpopulations: c_1 for short TiO_2 nanobelts and c_2 for long TiO_2 nanobelts. The population is the union of all the subpopulations and is considered as the TiO_2 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 $\mathbf{Y}(\mathbf{w}|c_q) = \mathbf{Y}(\mathbf{x},c_q)$, 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

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

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, $\mathcal{Y}_j(\mathbf{w}_i)$ is the observed response from the j^{th} replication at $\mathbf{w}_i$, and $n(\mathbf{w}_i)$ is the number of replications at $\mathbf{w}_i$.

On the basis of the sample data (eq 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., TiO_2 NPs of both shapes). SKQ aims at achieving this objective with high data efficiency and model generality.

STOCHASTIC KRIGING WITH QUALITATIVE FACTORS (SKQ)

In this section, the SKQ modeling 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 the readers' convenience, a review of SK is given in Review of Standard Stochastic Kriging, Supporting Information.

Compared to existing kriging-based methods,[15,32,33] SKQ is the first one that models all of the stochastic elements from the extrinsic variability caused by both quantitative and qualitative factors and 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 $\mathbf{w} = (\mathbf{x}^T,\mathbf{z}^T)^T$, with $\mathbf{x} = (x_1, x_2,..,x_d)^T \in \mathbb{R}^d$ and $\mathbf{z} = (z_1, z_2,..., z_L)^T$ including L qualitative factors. Each qualitative factor z_l has a number of category levels.

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

$$\mathcal{Y}_j(\mathbf{w}) = \mathbf{Y}(\mathbf{w}) + \varepsilon_j(\mathbf{w}) = \mathbf{f}(\mathbf{w})^T\beta + \mathbf{M}(\mathbf{w}) + \varepsilon_j(\mathbf{w}) \quad (3)$$

The expectation $\mathbf{Y}(\mathbf{w})$ is decomposed into the sum of two parts: $\mathbf{f}(\mathbf{w})^T\beta$ and $\mathbf{M}(\mathbf{w})$. Here, $\mathbf{f}(\mathbf{w})$ is a vector of known functions of $\mathbf{w}$, and β is a vector of unknown parameters of compatible dimension. Because it has been widely accepted that $\mathbf{f}(\mathbf{w})^T\beta = \beta_0$ (that is, just a constant term) suffices for most applications,[15] this work adopts $\mathbf{f}(\mathbf{w})^T\beta = \beta_0$ unless stated otherwise. The term $\mathbf{M}(\mathbf{w})$ 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 $\varepsilon_1(\mathbf{w})$, $\varepsilon_2(\mathbf{w})$,... 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 $\mathrm{Var}[\varepsilon(\mathbf{w})]$ is allowed to be dependent on $\mathbf{w}$.

Given the sample data (eq 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

$$\bar{\mathcal{Y}} = (\bar{\mathcal{Y}}(\mathbf{w}_1), \bar{\mathcal{Y}}(\mathbf{w}_2), ..., \bar{\mathcal{Y}}(\mathbf{w}_I))^T \quad (4)$$

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

$$\bar{\varepsilon} = (\bar{\varepsilon}(\mathbf{w}_1),\bar{\varepsilon}(\mathbf{w}_2),...,\bar{\varepsilon}(\mathbf{w}_I))^T \quad (5)$$

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

Extrinsic Variance Structure. The modeling of the extrinsic variability is performed following the framework proposed by Qian et al.[32] 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

$$\mathrm{Cov}[\mathbf{M}(\mathbf{w}),\mathbf{M}(\mathbf{w}')] = \delta^2 \times \mathrm{Corr}[\mathbf{M}(\mathbf{w}),\mathbf{M}(\mathbf{w}')] = \delta^2 \times [\prod_{l=1}^{L}\tau^{(l)}_{z_l,z_l'}] \times K(\mathbf{x},\mathbf{x}') \quad (6)$$

with δ^2 being the variance of the Gaussian process. The correlation $\mathrm{Corr}[\mathbf{M}(\mathbf{w}),\mathbf{M}(\mathbf{w}')]$ is decomposed as the product of two parts: $\prod_{l=1}^{L}\tau^{(l)}_{z_l,z_l'}$ 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 eq 6, $K(\mathbf{x},\mathbf{x}')$ represents the correlation across the quantitative settings for a given combination category of the qualitative factors $\mathbf{z}$ 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 Review of Standard Stochastic Kriging, Supporting Information, can e 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 el.,[34] Qian et al.[32]), and one of the most popular correlation functions in practice is the exponential correlation function

$$K(\mathbf{x},\mathbf{x}') = \exp\{\sum_{h=1}^{d} -\theta_h|x_h - x_h'|^p\} \quad (7)$$

In eq 7, $\theta = (\theta_1, \theta_2,...,\theta_d)$ is a vector of unknown parameters. It is required that $\theta_h > 0$ $(h = 1,2,...,d)$, and θ 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.[35]

The term $\prod_{l=1}^{L}\tau^{(l)}_{z_l,z_l'}$ in eq 6 is devoted to the correlations across different levels of qualitative factors. As noted in Qian et al.,[32] $\tau^{(l)}_{z_l,z_l'}$ 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 $\tau^{(l)}_{z_l,z_l'}$, a range of functional forms have been proposed in Qian et al.[32] and Zhou et al.[36] Below, two specific correlation functions are given as examples.

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

$$\tau^{(l)}_{z_l,z_l'} = \exp\{-\varphi^{(l)}I(z_l \neq z_l')\}; \quad l = 1,2,...,L \quad (8)$$

In eq 8, $\Phi = \{\phi^{(l)}; l = 1,2,...,L\}$ 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 lth qualitative factor are of isotropic nature; that is, a given l, $\tau^{(l)}_{z_l,z_l'}$ is a constant as long as $z_l \neq z_l'$.

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

$$\tau^{(l)}_{z_l,z_l'} = \exp\{-(\varphi^{(l)}_{z_l} + \varphi^{(l)}_{z_l'})I(z_l \neq z_l')\} \quad (9)$$

The unknown parameter set Φ includes the following components:

$$\varphi^{(l)}_{c_l}; \quad l = 1,2,...,L$$

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 $\tau^{(l)}_{z_l,z_l'}$ to be dependent on the category levels involved (i.e., z_l and z_l').

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

$$\Sigma_{\mathbf{M}} = \delta^2\mathbf{R}(\theta,\Phi) = \delta^2\begin{pmatrix} 1 & \mathrm{Corr}[\mathbf{M}(\mathbf{w}_1),\mathbf{M}(\mathbf{w}_2)] & \dots & \mathrm{Corr}[\mathbf{M}(\mathbf{w}_1),\mathbf{M}(\mathbf{w}_I)] \\ \mathrm{Corr}[\mathbf{M}(\mathbf{w}_2),\mathbf{M}(\mathbf{w}_1)] & 1 & \dots & \mathrm{Corr}[\mathbf{M}(\mathbf{w}_2),\mathbf{M}(\mathbf{w}_I)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Corr}[\mathbf{M}(\mathbf{w}_I),\mathbf{M}(\mathbf{w}_1)] & \mathrm{Corr}[\mathbf{M}(\mathbf{w}_I),\mathbf{M}(\mathbf{w}_2)] & \dots & 1 \end{pmatrix} \quad (10)$$

In eq 10, $\mathbf{R}(\theta,\Phi)$ 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 θ and Φ. For an arbitrary setting $\mathbf{w}_0$, the $I \times 1$ vector $\sum_{\mathbf{M}}(\mathbf{w}_0,\cdot)$ is defined as

dx.doi.org/10.1021/sc500102h | ACS Sustainable Chem. Eng. 2014, 2, 1581−1591

$$\sum_{M}(\mathbf{w}_0,\cdot) = \delta^2\mathbf{v}(\mathbf{w}_0,\theta,\Phi) = \delta^2\begin{pmatrix}\mathrm{Corr}[\mathbf{M}(\mathbf{w}_0),\mathbf{M}(\mathbf{w}_1)]\\ \mathrm{Corr}[\mathbf{M}(\mathbf{w}_0),\mathbf{M}(\mathbf{w}_2)]\\ \vdots \\ \mathrm{Corr}[\mathbf{M}(\mathbf{w}_0),\mathbf{M}(\mathbf{w}_I)]\end{pmatrix} \tag{11}$$

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

Intrinsic Variance Structure. The intrinsic variance of the random response at $\mathbf{w}$ is denoted as $\mathrm{Var}[\varepsilon(\mathbf{w})]$, which is dependent on the setting $\mathbf{w}$. Let $\sum_{\varepsilon}$ be the $I \times I$ variance−covariance matrix of vector $\bar{\varepsilon}$, which is defined in eq 5. Under the i.i.d. assumption for random errors, $\sum_{\varepsilon}$ is a $I \times I$ diagonal matrix

$$\sum_{\varepsilon} = \mathrm{diag}\{\mathrm{Var}[\varepsilon(\mathbf{w}_1)]/n(\mathbf{w}_1),\mathrm{Var}[\varepsilon(\mathbf{w}_2)]/n(\mathbf{w}_2), \ldots,\mathrm{Var}[\varepsilon(\mathbf{w}_I)]/n(\mathbf{w}_I)\} \tag{12}$$

Estimation and Inference by SKQ (Stochastic Kriging with Qualitative Factors). The SKQ-based estimation and inference requires the following assumption, which parallels Assumption 1 for SK (see Review of Standard Stochastic Kriging, Supporting Information).

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

Given a data set $\{(\mathbf{w}_i,\mathcal{Y}_j(\mathbf{w}_i));\ i = 1,2,\ldots,I;\quad j = 1,2,\ldots,n(\mathbf{w}_i)\}$ and under Assumption 1, $\bar{Y} = (\bar{Y}(\mathbf{w}_1), \bar{Y}(\mathbf{w}_2),\ldots,\bar{Y}(\mathbf{w}_I))^T$ as defined in eq 4 follows a multivariate normal distribution with constant mean vector $(\beta_0,\beta_0,\ldots,\beta_0)^{\mathrm{T}}$ and variance−covariance matrix

$$\sum(\delta^2,\theta,\Phi) = \sum_{\mathbf{M}} + \sum_{\varepsilon} = \delta^2\mathbf{R}(\theta,\Phi) + \sum_{\varepsilon} \tag{13}$$

Recall that $\sum_{\mathbf{M}}$ and $\sum_{\varepsilon}$ are defined in eqs 10 and 12, respectively. Thus, the log-likelihood function of $\bar{Y}$ in terms of the unknown parameters $(\beta_0,\delta^2,\theta,\Phi)$ can be written as

$$\ln \mathcal{L}(\beta_0,\delta^2,\theta,\Phi) = -\ln[(2\pi)^{I/2}] - \frac{1}{2}\ln[|\delta^2\mathbf{R}(\theta,\Phi) + \sum_{\varepsilon}|] - \frac{1}{2}(\bar{\mathcal{Y}} - \beta_0\mathbf{1}_I)^{\mathrm{T}} \times [\delta^2\mathbf{R}(\theta,\Phi) + \sum_{\varepsilon}]^{-1}(\bar{\mathcal{Y}} - \beta_0\mathbf{1}_I) \tag{14}$$

where $\mathbf{1}_I$ is a $(I \times 1)$ vector of ones. Following the framework of SK estimation (see Supporting Information), the procedure is developed as follows to obtain the SKQ parameter estimates for $(\beta_0,\delta^2,\theta,\Phi)$ that maximize eq 14.

1. Obtain the estimated $\sum_{\varepsilon}$

$$\hat{\Sigma}_{\varepsilon} = \mathrm{diag}\{\widehat{\mathrm{Var}}[\varepsilon(\mathbf{w}_1)]/n(\mathbf{w}_1), \widehat{\mathrm{Var}}[\varepsilon(\mathbf{w}_2)]/n(\mathbf{w}_2), \ldots, \widehat{\mathrm{Var}}[\varepsilon(\mathbf{w}_I)]/n(\mathbf{w}_I)\} \tag{15}$$

where

$$\widehat{\mathrm{Var}}[\varepsilon(\mathbf{w}_i)] = \frac{1}{n(\mathbf{w}_i)-1}\sum_{j=1}^{n(\mathbf{w}_i)}\left(\mathcal{Y}_j(\mathbf{w}_i) - \bar{\mathcal{Y}}(\mathbf{w}_i)\right)^2,\ i = 1,2,\ldots,I. \tag{16}$$

1. Replace $\sum_{\varepsilon}$ by $\hat{\sum}_{\varepsilon}$ and maximize the log-likelihood function (eq 14) with respect to (w.r.t.) $(\beta_0,\delta^2,\theta,\Phi)$. Specifically, two steps can be taken to solve the maximum likelihood problem. (i) Given δ^2, θ, and Φ, the maximum likelihood estimate (MLE) of β_0 is

$$\hat{\beta}_0(\delta^2,\theta,\Phi) = (\mathbf{1}_I^{\mathrm{T}}[\delta^2\mathbf{R}(\theta,\Phi) + \hat{\sum}_{\varepsilon}]^{-1}\mathbf{1}_I)^{-1} \times \mathbf{1}_I^{\mathrm{T}}[\delta^2\mathbf{R}(\theta,\Phi) + \hat{\sum}_{\varepsilon}]^{-1}\bar{\mathcal{Y}}$$

(ii) Substituting $\hat{\beta}_0(\delta^2,\theta,\Phi)$ into eq 14, the problem reduces to maximizing

$$\ln \mathcal{L}(\delta^2,\boldsymbol{\theta},\boldsymbol{\Phi}) = -\ln[(2\pi)^{I/2}] - \frac{1}{2}\ln[|\delta^2\mathbf{R}(\boldsymbol{\theta},\boldsymbol{\Phi}) + \hat{\Sigma}_{\varepsilon}|] - \frac{1}{2}(\bar{\mathcal{Y}} - \hat{\beta}_0(\delta^2,\boldsymbol{\theta},\boldsymbol{\Phi})\mathbf{1}_I)^{\top}[\delta^2\mathbf{R}(\boldsymbol{\theta},\boldsymbol{\Phi}) + \hat{\Sigma}_{\varepsilon}]^{-1}(\bar{\mathcal{Y}} - \hat{\beta}_0(\delta^2,\boldsymbol{\theta},\boldsymbol{\Phi})\mathbf{1}_I), \tag{17}$$

w.r.t. (δ^2,θ,Φ), which can be solved by a nonlinear optimization algorithm such as the Matlab fmincon function

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

$$\hat{\mathbf{Y}}(\mathbf{w}_0) = \hat{\beta}_0 + \mathbf{v}(\mathbf{w}_0,\hat{\theta},\hat{\Phi})^{\mathrm{T}}[\hat{\delta}^2\mathbf{R}(\hat{\theta},\hat{\Phi}) + \hat{\sum}_{\varepsilon}]^{-1}(\bar{\mathcal{Y}} - \hat{\beta}_0\mathbf{1}_I) \tag{18}$$

where $(\hat{\beta}_0,\hat{\delta}^2,\hat{\theta},\hat{\Phi})$ are the maximum likelihood estimates obtained from the previous step. Recall that $\mathbf{v}(\mathbf{w}_0,\hat{\theta},\hat{\Phi})$ is defined in eq 11. Following the proof scheme of Ankenman et al.,[15] it can be shown that eq 18 is the best linear unbiased estimator for $\mathbf{Y}(\mathbf{w}_0)$.

The mean squared error (MSE) is obtained as

$$\widehat{\mathrm{MSE}}[\hat{\mathrm{Y}}(\mathbf{w}_0)] = \hat{\delta}^2 - \hat{\delta}^4\mathbf{v}(\mathbf{w}_0,\hat{\theta},\hat{\Phi})^{\top}[\hat{\delta}^2\mathbf{R}(\hat{\theta},\hat{\Phi}) + \hat{\Sigma}_{\varepsilon}]^{-1}\mathbf{v}(\mathbf{w}_0,\hat{\theta},\hat{\Phi}) + \eta^2(\mathbf{1}_I^{\top}[\hat{\delta}^2\mathbf{R}(\hat{\theta},\hat{\Phi}) + \hat{\Sigma}_{\varepsilon}]^{-1}\mathbf{1}_I)^{-1}, \tag{19}$$

where $\eta = 1 - \mathbf{1}_I^{\mathrm{T}}[\hat{\delta}^2\,\mathbf{R}(\hat{\theta},\hat{\Phi}) + \hat{\sum}_{\varepsilon}]^{-1}\mathbf{v}(x_0,\hat{\theta},\hat{\Phi})\hat{\delta}^2$.

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

$$\hat{\mathrm{Y}}(\mathbf{w}_0) \pm z_{1-\alpha/2}\sqrt{\widehat{\mathrm{MSE}}[\hat{\mathrm{Y}}(\mathbf{w}_0)]} \tag{20}$$

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

Inverse Estimation and Inference. In this section, we consider in particular the SKQ-based modeling/inference for 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), plays an important role in setting safety standard, and is of particular interest in toxicology studies.

Following the notations 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 $\{c_q;q = 1,2\cdots,Q\}$. The expected dose−response curve is denoted as $\mathbf{Y}(x,c_q)$ 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.[13,37] 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 preselected absolute response in this work, and the BMD for a subpopulation c_q is written as

$$\mathrm{BMD}(c_q) = \mathbf{Y}^{-1}(\mathrm{BMR},c_q) \tag{21}$$

Here, $\mathbf{Y}^{-1}$ represents the functional dependence of $\mathrm{BMD}(c_q)$ upon BMR, assuming that the inverse mapping exists.[28]

The collection of dose−response curves $\{\mathbf{Y}(x,c_q);q = 1,2,\ldots,Q\}$ are modeled by SKQ. To perform the inverse calculation as given

dx.doi.org/10.1021/sc500102h | ACS Sustainable Chem. Eng. 2014, 2, 1581−1591

in eq 21, numerical interpolation needs to be employed based on the fitted SKQ model. In this work, the cubic spline interpolation recommended by Hastie et al.[38] is used to perform the inverse computation for BMD estimation.

Denote $\widehat{\mathrm{BMD}}(c_q)$ as the estimated BMD from the SKQ model for the subpopulation c_q. The uncertainty of $\widehat{\mathrm{BMD}}(c_q)$, represented by $\mathrm{Var}[\widehat{\mathrm{BMD}}(c_q)]$, directly relates to the safety standard of the substance being investigated. To estimate $\mathrm{Var}[\widehat{\mathrm{BMD}}(c_q)]$, numerical methods have to be resorted to again. The bootstrap resampling method developed by Kirk et al.[39] is adapted to quantify the uncertainty of $\{\widehat{\mathrm{BMD}}(c_q); q = 1,2,...,Q\}$ based on the SKQ modeling of given data $\{(\mathbf{w}_i, \mathcal{Y}_j(\mathbf{w}_i)); \quad i = 1,2,...,I; \quad j = 1,2,...,n(\mathbf{w}_i)\}$, and the adapted bootstrapping algorithm is described in Figure 1.

Bootstrap Resampling Algorithm

Input: (a) The I design points $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_I\}$, at which real data have been collected; (b) The sample variance $\widehat{\mathrm{Var}}[\varepsilon(\mathbf{w}_i)]; i = 1, 2, \ldots, I$ estimated as in (16) from the real sample data; (c) the fitted SKQ model $\widehat{\mathrm{Y}}(\mathbf{w})$ and $\widehat{\mathrm{MSE}}[\widehat{\mathrm{Y}}(\mathbf{w})]$ as given in (18) and (19), both estimated from the real data; (d) The resample size B.

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

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

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

(ii) Based on the bootstrap data

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

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

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

Figure 1. Bootstrap resampling algorithm for uncertainty quantification of BMD estimates.

This algorithm falls into the category of parametric bootstrap resampling method. There are also nonparametric and semiparametric counterparts,[40−42] and kriging-based bootstrapping inference remains a research topic under investigation.

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

$$\{\widehat{\mathrm{BMD}}^*_b(c_q); b = 1, 2, \ldots, B\} \tag{22}$$

On the basis of bootstrapping estimates in eq 22, the nonparametric method suggested by Davison and Hinkley[43] can be easily employed to estimate the $100\alpha^{\mathrm{th}}$ ($\alpha \in (0,1)$) percentile of $\widehat{\mathrm{BMD}}(c_q)$. In this context, the percentile estimate is referred to as BMDL, which serves as the lower bound of the one-sided 100α% CI for the BMD. For the subpopulation c_q, the one-sided CI of $\widehat{\mathrm{BMD}}(c_q)$ is written as $[\widehat{\mathrm{BMD}}(c_q), \infty)$. It is worth noting that the percentile method, which is employed here for the estimation of the lower bound BMDL, is one of the various existing methods for estimating the CI boundaries from a given bootstrap sample, and we refer the interested readers to Efron et al.[44] for more information.

EMPIRICAL 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 (see Review of Standard Stochastic Kriging, Supporting Information) and MEM (see Review of Mixed Effects Model, Supporting Information) 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 with no information pooling.

Case 2: A multi-source dose−response case is developed to show that compared to MEM (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 experiment data. Simulation, rather than real experiments, is employed for the following reasons. First, in only a simulation-based study, the true expected response surfaces (i.e., the simulation models) are available to evaluate the modeling results. Second, the resulting estimated model is a random outcome, which depends on the random sample data; hence, a modeling method needs to be evaluated in a statistical manner based on the outcomes of applying it on a large number of randomly sampled data sets, which is impossible with real experiments. These advantages of simulation will become clear in the case studies below.

Case 1: Modeling Multi-Source Dose−Time−Response Data.

This case is constructed based on the dose−time−response study of TiO_2 nanoparticles (NPs) performed by Porter et al.[45] There are two quantitative factors, $\mathbf{x} = (x_1, x_2)$, with $x_1 \in [0,15]$ μg 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. 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/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 $\{\mathbf{Y}(\mathbf{x}, c_1), \mathbf{Y}(\mathbf{x}, c_2)\}$, with specific expressions given by Models S25 and S26 in True Expected Exposure−Response Model for Case 1, Supporting Information. Both Models S25 and S26 take the form of a single hidden layer feed-forward neural network and are estimated from real biological data.[46] The true dose−time−response surfaces are plotted in Figure 2.

The true variance models used in the simulation are given as

$$\mathrm{Var}[\varepsilon(\mathbf{x}, c_1)] = (0.2\mathbf{Y}(\mathbf{x}, c_1)^{0.7})^2 \tag{23}$$

$$\mathrm{Var}[\varepsilon(\mathbf{x}, c_2)] = (0.3 \exp(\mathbf{Y}(\mathbf{x}, c_2) \times 0.005))^2 \tag{24}$$

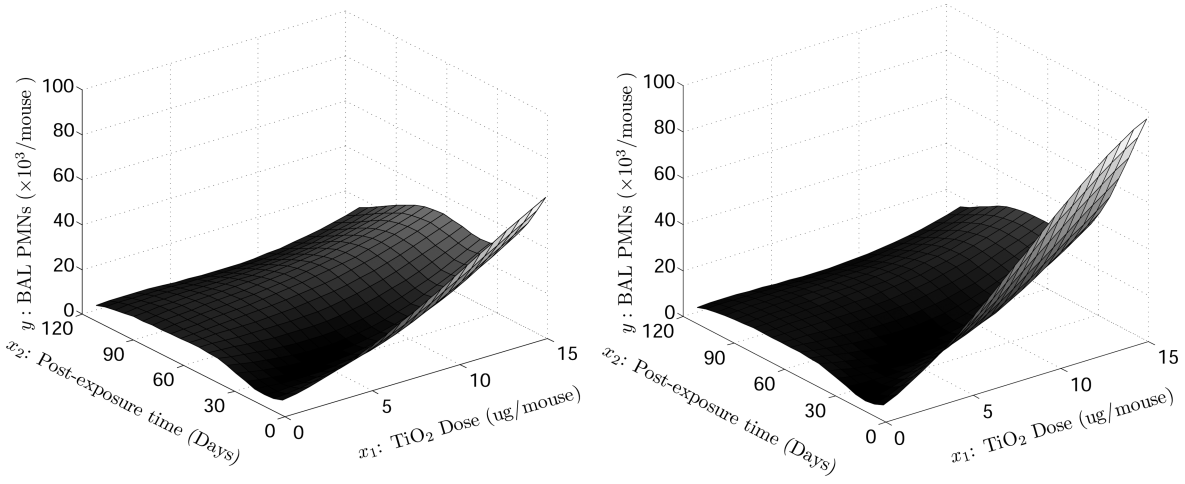
For a subpopulation c_q ($q = 1,2$) and at an exposure level $\mathbf{x}_0$, a random response y_0 is simulated as

$$y_0 = \mathbf{Y}(\mathbf{x}_0, c_q) + \sqrt{\mathrm{Var}[\varepsilon(\mathbf{x}_0, c_q)]} \times \epsilon; \quad q = 1,2 \tag{25}$$

where ϵ is a random error provided by a standard normal random generator.[47]

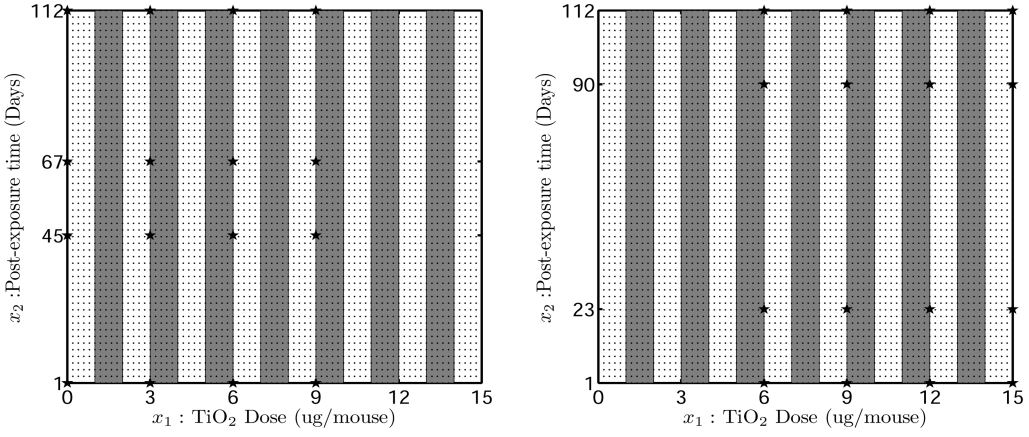
Case 1 is designed to compare the modeling efficiency of SKQ and SK. MEM has not been applied to this case for the following

dx.doi.org/10.1021/sc500102h | ACS Sustainable Chem. Eng. 2014, 2, 1581−1591



(a) c_1: Short TiO_2 nanobelts (b) c_2: Long TiO_2 nanobelts

Figure 2. True exposure−response surfaces for Case 1.



(a) c_1: Short TiO_2 nanobelts (b) c_2: Long TiO_2 nanobelts

Figure 3. Design points in the EDS (estimation data set) and check points in the VDS (validation data set).

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 2) in this case. Second, the variance structures of eqs 23 and 24 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 Review of Mixed Effects Model of the Supporting Information.

Sampling via Simulation. Two types of data sets are obtained via simulation: an estimation data set (EDS) for model estimation/inference and a validation data set (VDS) that 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 3a and 16 points for subpopulation c_2 (long nanobelts) depicted as stars in Figure 3b. At each design point, Model S25 is used to generate eight i.i.d. random responses; that is, eight 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. The collection of all the check points is denoted as C, with $C = C_{c_1} \cup C_{c_2}$, C_{c_1} denotes the collection of dots in Figure 3a, and C_{c_2} is the collection of dots in Figure 3b. Further, C_{c_q} $(q = 1,2)$ is divided into a number of subsets: $C_{c_q} = C_{c_q,1} \cup C_{c_q,2} \cup \cdots \cup C_{c_q,15}$, where $C_{c_q,k}$ represents the collection of check points within the subregion specified by the dose range $[k-1,k)$. The subregions are shown in Figure 3a and b as alternating white and gray rectangles. At each check point, the true expected response $\mathbf{Y}(\cdot)$ is available (Figure 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, 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 (eq 7) is adopted to capture the extrinsic variability for both subpopulations. Two separate SK models are estimated from the short-nanobelt and long-nanobelt data subsets, respectively. At an arbitrary setting $\mathbf{w} = (\mathbf{x},c_q)$, formulas S13 and S14 were used, based on the fitted SK model associated with c_q, to obtain, respectively, the point estimate and CI of the expected response $\mathbf{Y}(\mathbf{w})$.

To apply SKQ, the correlation functions for the extrinsic variability (eq 6) are specified as follows: the exponential

correlation function (eq 7) is adopted for $K(\mathbf{x},\mathbf{x}')$, and the EC (isotropic) correlation function (eq 8) is used for $\tau^{(l)}_{z_l,z_l'}$. 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 (eq 18) provides the point estimate for $\mathbf{Y}(\mathbf{w})$, and eq 20 gives the CI for $\mathbf{Y}(\mathbf{w})$.

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 because SKQ allows for the borrowing of information (or data) from all the other subpopulations. To demonstrate SKQ's estimation efficiency, herein we compare the SK and SKQ results in terms of the quality of their point estimates for $\mathbf{Y}(\cdot)$.

The estimated root mean squared error (ERMSE) defined as follows is used to evaluate the goodness of point estimate $\hat{\mathbf{Y}}(\cdot)$. Recall that $C_{c_q,k}$ 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. We define

$$\mathrm{ERMSE}(C_{c_q,k}) = \sqrt{\frac{1}{\#[C_{c_q,k}]}\sum_{\mathbf{w}\in C_{c_q,k}}(\hat{\mathbf{Y}}(\mathbf{w}) - \mathbf{Y}(\mathbf{w}))^2}\,; \quad q = 1,\ 2; \quad k = 1,2,...,15 \tag{26}$$

where $\#[C_{c_q,k}]$ represents the total number of check points in the set $C_{c_q,k}$. Clearly, eq 26 measures the average deviation of $\hat{\mathbf{Y}}(\cdot)$ from the true value $\mathbf{Y}(\cdot)$ at the check points in $C_{c_q,k}$.

Applying a modeling method (SK or SKQ) on one EDS, denoted as $\mathrm{EDS}^{(r)}$ $(r = 1,2,...,R)$, is considered as one macro-replication and leads to a set of performance statistics

$$\{\mathrm{ERMSE}^{(r)}(C_{c_q,k});\quad q = 1,2;\quad k = 1,2,\ ...,\ 15\}$$

For empirical evaluation, a total of $R = 1000$ independent EDS has been generated by simulation experiments, and the average ERMSE across the macro-replications are denoted as

$$\left\{\overline{\mathrm{ERMSE}}(C_{c_q,k}) = \frac{1}{R}\sum_{r=1}^{R}\mathrm{ERMSE}^{(r)}(C_{c_q,k});\quad q = 1,2;\quad k = 1,2,...,15\right\}$$

and are summarized in Table 1. The statistic $\overline{\mathrm{ERMSE}}(C_{c_q,k})$ measures the expected overall deviation between $\mathbf{Y}(\cdot)$ and $\hat{\mathbf{Y}}(\cdot)$ within the subset $C_{c_q,k}$. As shown in Table 1, in each subset of check points $C_{c_q,k}$, 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 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 statistical models of substantially improved quality.

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.[45] There is one quantitative factor x representing 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/mouse.

Table 1. Comparison of Estimation Results from SK and SKQ for Case 1

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

Simulation Model. Simulation data that mimic the real experiment data are generated using the following true dose–response model

$$\begin{aligned}\mathbf{Y}(x,c_1) &= 20\exp(x/11)\\ \mathbf{Y}(x,c_2) &= 19.5\exp(x/10.5)\\ \mathbf{Y}(x,c_3) &= 26\exp(x/13.2)\end{aligned} \tag{27}$$

and variance model

$$\mathrm{Var}[\varepsilon(x,c_q)] = 0.2^2\mathbf{Y}(x,c_q)^{2\times 0.6};\quad q = 1,2,3 \tag{28}$$

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

$$y_0 = \mathbf{Y}(x_0,c_q) + 0.2\mathbf{Y}(x_0,c_q)^{0.6}\times\varepsilon;\quad q = 1,2,3 \tag{29}$$

where ϵ is a random error provided by a standard normal random generator.[47]

This case 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 (eq 28) across different data sources. The third assumption made by MEM is the multivariate normality of the model coefficient vector (see Review of Mixed Effects Model, Supporting Information). According to eq 27, the three true coefficient vectors in this case are $(20,11)^{\mathrm{T}}$, $(19.5,10.5)^{\mathrm{T}}$, and $(26,13.2)^{\mathrm{T}}$; on the basis of 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

dx.doi.org/10.1021/sc500102h | ACS Sustainable Chem. Eng. 2014, 2, 1581–1591

Table 2. Design Points in 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_1	c_1	c_1	c_1	c_2	c_2	c_2	c_2	c_2	c_3	c_3	c_3	c_3	c_3

evaluation, respectively. To generate an EDS, the design as shown in Table 2 is used; eight replications are carried out at 5 evenly spaced dose levels for each of the three categories. An example EDS is given in Table S1 of An Estimation Data Set for Case 2 of the Supporting Information.

Applying the Modeling Methods. The two alternative information pooling methods, MEM and SKQ, were applied on the EDS given in Table S1 of the Supporting Information.

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

$$\mathbf{Y}(x,c_q) = \alpha_{c_{q,1}} \exp(x/\alpha_{c_{q,2}}); \quad q = 1,2,3 \tag{30}$$

with $\boldsymbol{\alpha}_{cq} = (\alpha_{cq,1},\alpha_{cq,2})$ being the unknown parameters for subpopulation c_q (q = 1,2,3). The variance model is assumed to follow

$$\mathrm{Var}[\varepsilon(x,c_q)] = \sigma^2 \mathrm{Y}(x,c_q)^{2\gamma}; \quad q = 1,2,3 \tag{31}$$

where σ and γ are unknown parameters common to different subpopulations. Note that the assumed forms (eqs 30 and 31) are the ones that the true models (eqs 27 and 28) fall into, respectively; the adoption of functional forms (eqs 30 and 31) is meant to allow MEM to achieve its ideal estimation results.

With the assumed forms (eqs 30 and 31), MEM is performed on the EDS in Table S1 of the Supporting Information. The fitted dose−response models are

$$\begin{aligned} \hat{\mathbf{Y}}(x,c_1) &= 20.04 \exp(x/11.09) \\ \hat{\mathbf{Y}}(x,c_2) &= 19.60 \exp(x/10.54) \\ \hat{\mathbf{Y}}(x,c_3) &= 26.24 \exp(x/13.38) \end{aligned} \tag{32}$$

The fitted variance model is

$$\widehat{\mathrm{Var}}[\varepsilon(x,c_q)] = 0.214^2 \hat{\mathrm{Y}}(x,c_q)^{2\times 0.597}; \; q = 1,2,3. \tag{33}$$

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 Review of Mixed Effects Model, Supporting Information).

When applying SKQ to the same EDS (Table S1, Supporting Information), the correlation (eq 6) is constructed as follows: The exponential correlation function (eq 7) is used to model the correlations between quantitative variables, and the MC correlation function (eq 9) is used to model the correlations across different levels of qualitative factors. Normalization of the original data (Table S1, Supporting Information) was also performed so that both the quantitative factors and responses range over [0,1]. Applying the maximum likelihood estimation procedure (see Estimation and Inference by SKQ) on the normalized data leads to the fitted parameters for the SKQ model as displayed in Table 3.

Table 3. SKQ Parameters Estimated from Normalized Dose−Response Data for Case 2

$\hat{\beta}_0$	$\hat{\delta}^2$	$\hat{\theta}_1$	$\hat{p}$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$
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 (see Estimation and Inference by SKQ), and the inverse BMD can be obtain numerically (see Inverse Estimation and Inference); the confidence intervals for these quantities can also be obtained accordingly. Note that when utilizing the SKQ specified by Table 3 for the estimation/inference, a simple conversion calculation is needed to ensure that the estimates are given on the original scale because those parameters are obtained from the normalized data.

Comparison of the Two Modeling Methods. The modeling results of MEM and SKQ are compared in terms of their estimation/inference abilities for (I) the expected responses as well as (II) the BMD values.

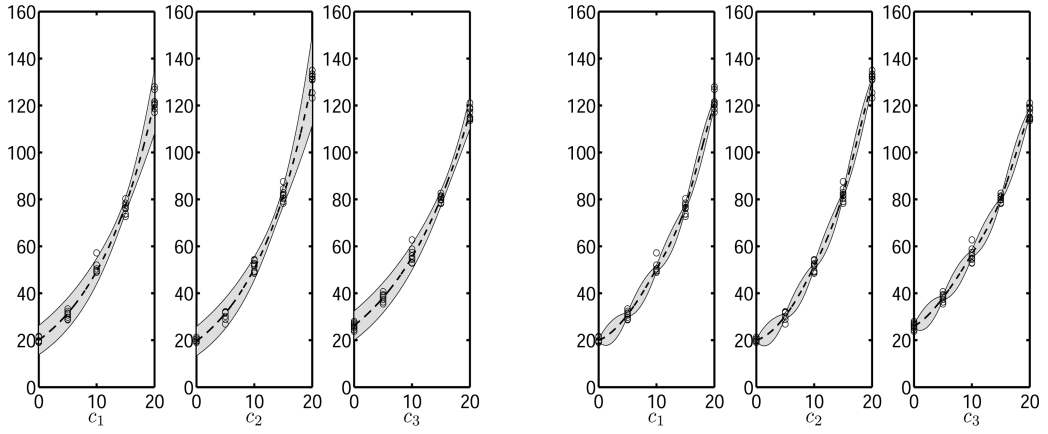
(I) Estimation/Inference of the Expected Response $\hat{\mathbf{Y}}(\cdot)$. From the one EDS given in Table S1 of the Supporting Information, both MEM and SKQ were applied as described earlier, and the estimation results for the two methods are displayed in Figure 4. The circles denote the EDS and are plotted in both Figure 4a and b. The dashed curves represent the estimated expected responses, and the solid curves are the lower and upper 95% CI curves for the true expected responses. As shown in Figure 4, over the dose range, the widths of the CIs (that is, the vertical distances between the lower and upper CI curves) provided by SKQ are generally narrower than those given by MEM; this pattern, which is graphically illustrated for the one EDS plotted as circles in Figure 4, holds consistently for all of the R = 1000 macro-replications carried out in this study.

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 $\hat{\mathbf{Y}}(\cdot)$ 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 $\hat{\mathbf{Y}}(\cdot)$. In our simulation study, $\hat{\mathbf{Y}}(\cdot)$ is available from the simulation model (eq 27) for the purpose of evaluating the CIs. Ideally, among these 1000 CIs, the percentage of the CIs that actually contain $\hat{\mathbf{Y}}(\cdot)$ should be very close to 95%, the nominal coverage level.

Table 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 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 much 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 shown in Figure 4 and Table 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 wider CIs.

(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

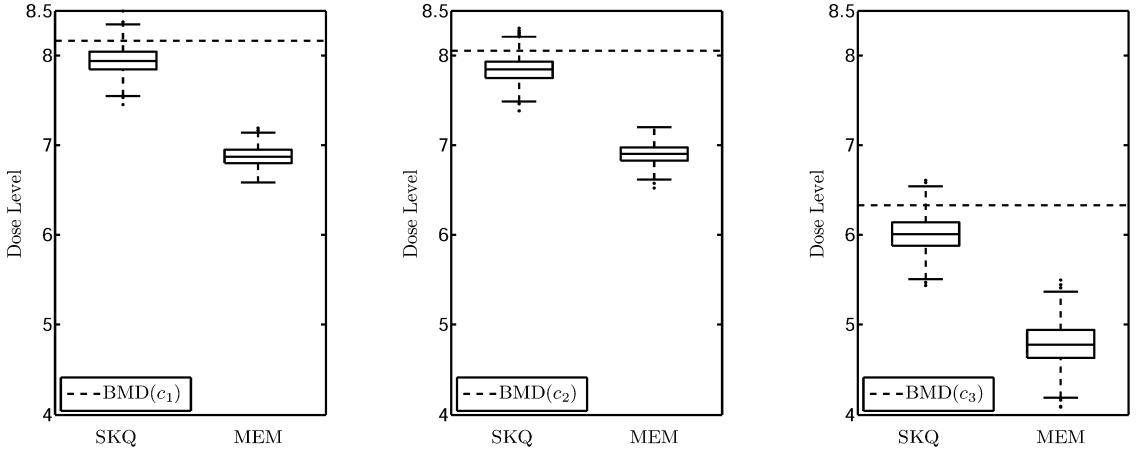


(a) MEM fitted dose-response curves (b) SKQ fitted dose-response curves

Figure 4. Comparison of the dose−response fitting results from MEM and SKQ.

Table 4. Comparison of MEM and SKQ in Terms of CI Coverage Probabilities for the Expected Response Y(·)

	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



(a) Subpopulation c_1 (b) Subpopulation c_2 (c) Subpopulation c_3

Figure 5. Box plots for the BMDLs resulting from the two modeling methods.

demonstration, the BMR is set as 42×10^3/mouse in this case. (Recall that the response of interest is BAL PMNs measured in the units of 10^3/mouse).

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

$$\{\widehat{\mathrm{BMDL}}^{(r)}_{\mathrm{MEM}}(c_q); q = 1,2,3; r = 1,2,\ldots,1000\}, \quad (34)$$

The BMDLs obtained from the 1000 SKQ macro-replications are represented as

$$\{\widehat{\mathrm{BMDL}}^{(r)}_{\mathrm{SKQ}}(c_q); q = 1,2,3; r = 1,2,\ldots,1000\}, \quad (35)$$

The true values for BMD(c_q), q = 1,2,3, can be easily obtained for BMR = 42×10^3 based on the simulation model (eq 27) and are represented by the horizontal lines in Figure 5a−c corresponding to the three subpopulations. Figure 5a is devoted to the subpopulation c_1, and the two box plots are generated based on $\{\widehat{\mathrm{BMDL}}^{(r)}_{\mathrm{MEM}}(c_1);\ r = 1,2,...,1000\}$ and $\{\widehat{\mathrm{BMDL}}^{(r)}_{\mathrm{SKQ}}(c_1);\ r = 1,2,...,1000\}$. Figure 5b and c are plotted for the other two subpopulations in the same manner. Clearly, the BMDLs given by SKQ are much closer to the true BMD than those provided by MEM.

Each BMDL estimate in eqs 34 and 35 corresponds to a one-sided 95% CI: [BMDL,∞]. Table 5 compares the coverage probabilities of the CIs obtained from MEM and SKQ. It is shown that the coverage probabilities of SKQ are close to the nominal level 95%, whereas MEM's estimated probabilities are all 1.000. Therefore, as shown in Figure 5 and Table 5, SKQ is able to give more informative CIs of the BMD compared to MEM.

dx.doi.org/10.1021/sc500102h | ACS Sustainable Chem. Eng. 2014, 2, 1581−1591

Table 5. Comparison of MEM and SKQ in Terms of CI Coverage Probabilities for 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

SUMMARY

A new semi-parametric statistical model, SKQ (stochastic kriging with qualitative factors), has been developed in this work. SKQ is the first kriging-based model able to take into account the three types of variability steming from quantitative factors, qualitative factors, and uncountable 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 over the existing methods, SK (standard stochastic kriging) and MEM. SKQ is able to pool information across multiple data sources, accommodate general data features, and provide more informative estimation/inference from given data. For clarity and succinctness of the presentation, the two cases for this paper were designed to involve a relatively small number of data sources (subpopulations). It is worth noting that when applying SKQ to data of 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 (information pooling effects) is more pronounced with more sources of data. All the SKQ-related algorithms are currently implemented in MATLAB programs, which are available upon request from the authors. To facilitate SKQ modeling and inference by biologists, a VBA (Visual Basic for Application) Excel macro tool is under development. The VBA tool intends to allow users who possess the basic ability to use Excel to perform SKQ by interacting with Excel spreadsheets and menus.

As a final note, the focus of this work is on efficient modeling of given data. However, SKQ's ability to provide more informative statistical inference from limited data certainly opens opportunities for efficient design of experiments (DOE). Given some subsets of data already collected and likely from different sources, how should we 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, and in our ongoing research, a SKQ-based DOE method is under development to achieve experimental efficiency.

ASSOCIATED CONTENT

Supporting Information

Review of standard stochastic kriging (SK) and mixed-effects model (MEM), true expected exposure−response models for case 1 and an estimation data set (EDS) for case 2. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: feng.yang@mail.wvu.edu. Phone: 304-293-9477. Fax: 304-293-4970.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation Grant CBET-1065931 and the National Institute of Environmental Health Sciences Grant 1RC2ES018742-01. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

REFERENCES

(1) Zhi, M.; Manivannan, A.; Meng, F.; Wu, N. Highly conductive electrospun carbon nanofiber/MnO_2 coaxial nano-cables for high energy and power density supercapacitors. *J. Power Sources* **2012**, *208*, 345−353.

(2) Xiang, C.; Li, M.; Zhi, M.; Manivannan, A.; Wu, N. Reduced graphene oxide/titanium dioxide composites for supercapacitor electrodes: shape and coupling effects. *J. Mater. Chem.* **2012**, *22*, 19161−19167.

(3) Li, M.; Wang, Q.; Shi, X.; Hornak, L. A.; Wu, N. Detection of mercury(II) by quantum dot-DNA-gold nanoparticle ensemble based nanosensor via nanometal surface energy transfer. *Anal. Chem.* **2011**, *83*, 7061−7065.

(4) Wu, N.; Zhao, M.; Zheng, J.-G.; Jiang, C.; Myers, B.; Li, S.; Chyu, M.; Mao, S. X. Porous CuO-ZnO nanocomposite for sensing electrode of high-temperature CO solid-state electrochemical sensor. *Nanotechnology* **2005**, *16*, 2878−2881.

(5) Wu, N.; Chen, Z.; Xu, J.; Chyu, M.; Mao, S. X. Impedance-metric Pt/YSZ/Au-Ga_2O_3 sensors for CO detection at high-temperature. *Sens. Actuators, B* **2005**, *110*, 49−53.

(6) Li, M.; Zhang, J.; Suri, S.; Sooter, L. J.; Ma, D.; Wu, N. Detection of adenosine triphosphate with an aptamer biosensor based on surface-enhanced Raman scattering. *Anal. Chem.* **2012**, *84*, 2837−2842.

(7) Ren, N.; Li, R.; Chen, L.; Wang, G.; Liu, D.; Wang, Y.; Zheng, L.; Tang, W.; Yu, X.; Jiang, H.; Liu, H.; Wu, N. In-situ construction of titanate/silver nanoparticle/titanate sandwich nanostructure on metallic titanium surface for bacteriostatic and biocompatible implants. *J. Mater. Chem.* **2012**, *22*, 19151−19160.

(8) Edler, L.; Poirier, K.; Dourson, M.; Kleiner, J.; Mileson, B.; Nordmann, H.; Renwickf, A.; Slobg, W.; Waltonf, K.; Wurtzenh, G. Mathematical modelling and quantitative methods. *Food Chem. Toxicol.* **2002**, *40*, 283−326.

(9) Bretz, F.; Hsu, J.; Pinheiro, J.; Liu, Y. Dose finding − A challenge in statistics. *Biometrical Journal* **2008**, *50*, 480−504.

(10) *Benchmark Dose Technical Guidance*; EPA/100/R-12/001; U.S. Environmental Protection Agency, Washington, DC, 2012.

(11) Crump, K. S. A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* **1984**, *4*, 854−871.

(12) Crump, K. Critical issues in benchmark calculations from continuous data. *Crit. Rev. Toxicol.* **2002**, *32*, 133−153.

(13) *Benchmark Dose Software*; U.S. Environmental Protection Agency: Washington, DC, 2012; Version 2.3.1.

(14) Davidian, M.; Giltinan, D. M. *Nonlinear Models for Repeated Measurement Data*; Chapman & Hall: Boca Raton, FL, 1995.

(15) Ankenman, B.; Nelson, B. L.; Staum, J. Stochastic kriging for simulation metamodeling. *Oper. Res.* **2010**, *58*, 371−382.

(16) Bailer, A. J.; Hughes, M. R.; Denton, D. L.; Oris, J. T. An empirical comparison of effective concentration estimators for evaluating aquatic toxicity test responses. *Environ. Toxicol. Chem.* **2000**, *19*, 141−150.

(17) Kass, R. E.; Steffey, D. Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *J. Am. Stat. Assoc.* **1989**, *84*, 717−726.

(18) *Current Intelligence Bulletin 65: Occupational Exposure to Carbon Nanotubes and Nanofibers*; Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health: Atlanta, GA, 2013.

(19) Slob, W. Dose-response modeling of continuous endpoints. *Toxicol. Sci.* **2002**, *66*, 298−312.

(20) Sand, S.; von Rosen, D.; Eriksson, P.; Fredriksson, A.; Viberg, H.; Victorin, K.; Filipsson, A. Dose-response modeling and benchmark calculations from spontaneous behavior data on mice neonatally

exposed to 2,2′,4,4′,5-pentabromodiphenyl ether. *Toxicol. Sci.* **2004**, *81*, 491−501.
(21) Oberdorster, G.; Oberdorster, E.; Oberdorster, J. Nanotoxicology: An emerging discipline evolving from studies of ultrafine particles. *Environ. Health Perspect.* **2005**, *113*, 823−839.
(22) Woodall, G. M.; Gift, J. S.; Foureman, G. L. Empirical Methods and Default Approaches in Consideration of Exposure Duration in Dose-Response Relationships. In *General, Applied, and Systems Toxicology*; John Wiley and Sons, Inc.: New York, **1999**.
(23) Sun, K.; Krause, G. F.; Mayer, F. L.; Ellersieck, M. R.; Basu, A. P. Estimation of acute toxicity by fitting a dose-time-response surface. *Risk Anal.* **1995**, *15*, 247−252.
(24) Dunn, G. Optimal designs for drug, neurotransmitter and hormone receptor assays. *Stat. Med.* **1988**, *7*, 805−815.
(25) Chen, X.; Ankenman, B. E.; Nelson, B. L. Enhancing stochastic kriging metamodels with gradient estimators. *Oper. Res.* **2013**, *61*, 512−528.
(26) Chen, X.; Ankenman, B. E.; Nelson, B. L. The effects of common random numbers on stochastic kriging metamodels. *ACM Trans. Model. Comput. Simul.* **2012**, *22*, 7.
(27) Seber, G.; Wild, C. *Nonlinear Regression*; Wiley, New York, 2005.
(28) Wang, K.; Yang, F.; Porter, D. W.; Wu, N. Two-stage experimental design for dose−response modeling in toxicology studies. *ACS Sustainable Chem. Eng.* **2013**, *1*, 1119−1128.
(29) Giltinan, D. M.; Davidian, M. Assays for recombinant proteins: A problem in non-linear calibration. *Stat. Med.* **1994**, *13*, 1165−1179.
(30) Zeng, Q.; Davidian, M. Bootstrap-adjusted calibration confidence intervals for immunoassay. *J. Am. Stat. Assoc.* **1997**, *92*, 278−290.
(31) Zhang, J.; Bailer, A. J.; Oris, J. T. Bayesian approach to estimating reproductive inhibition potency in aquatic toxicity testing. *Environ. Toxicol. Chem.* **2012**, *31*, 916−927.
(32) Qian, P. Z.; Wu, H.; Wu, C. J. Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* **2008**, *50*.
(33) Han, G.; Santner, T. J.; Notz, W. I.; Bartel, D. L. Prediction for computer experiments having quantitative and qualitative input variables. *Technometrics* **2009**, *51*.
(34) Santner, T. J.; Williams, B. J.; Notz, W. *The Design and Analysis of Computer Experiments*; Springer-Verlag: New York, 2003.
(35) Rasmussen, C. E.; Christopher, K. I. W. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, 2006.
(36) Zhou, Q.; Qian, P. Z.; Zhou, S. A simple approach to emulation for computer models with qualitative and quantitative factors. *Technometrics* **2011**, *53*.
(37) Filipsson, A.; Sand, S.; Nilsson, J.; Victorin, K. The benchmark dose method − Review of available models and recommendations for application in health risk assessment. *Crit.. Rev. Toxicol.* **2003**, *33*, 505−542.
(38) Hastie, T.; Tibshirani, R.; Friedman, J. J. H. *The Elements of Statistical Learning*; Springer: New York, 2001; Vol. *1*.
(39) Kirk, P. D.; Stumpf, M. P. Gaussian process regression bootstrapping: Exploring the effects of uncertainty in time course data. *Bioinformatics* **2009**, *25*, 1300−1306.
(40) Shao, J.; Tu, D. *The Jackknife and Bootstrap*; Springer-Verlag: New York, 1995.
(41) Schelin, L.; Sjöstedt-de, S. L. Kriging prediction intervals based on semiparametric bootstrap. *Math. Geosci.* **2010**, *42*, 985−1000.
(42) Sjöstedt-de, S. L. The bootstrap and kriging prediction intervals. *Scand. J. Stat.* **2003**, *30*, 175−192.
(43) Davison, A. C.; Hinkley, D. V. *Bootstrap Methods and Their Application*; Cambridge University Press: Cambridge, U.K., 1997; Vol. *1*.
(44) Efron, B.; Tibshirani, R. J. *An Introduction to the Bootstrap*; CRC Press: Boxca Raton, FL, 1993; Vol. *57*.
(45) Porter, D. W.; et al. Differential mouse pulmonary dose and time course responses to titanium dioxide nanospheres and nanobelts. *Toxicol. Sci.* **2013**, *131*, 179−193.
(46) Yang, F. Neural network metamodeling for cycle time-throughput profiles in manufacturing. *Eur. J. Oper. Res.* **2010**, *205*, 172−185.
(47) Kelton, W. D.; Law, A. M. *Simulation Modeling and Analysis*; McGraw Hill: Boston, MA, 2000.