

Supporting Information

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, WV 26506, USA, Department of Statistical Sciences and Operations
Research, Richmond, VA 23284 USA, National Institute for Occupational Safety and Health
(NIOSH), Morgantown, WV 26505 USA, and Department of Mechanical and Aerospace
Engineering, West Virginia University, Morgantown, WV 26506, USA*

E-mail: feng.yang@mail.wvu.edu

Number of pages:13

Number of figures:0

Number of tables: 1

*To whom correspondence should be addressed

[†]Department of Industrial and Management System Engineering, West Virginia University

[‡]Department of Statistical Sciences and Operations Research

[¶]National Institute for Occupational Safety and Health (NIOSH)

[§]Department of Mechanical and Aerospace Engineering, West Virginia University

Review of Standard Stochastic Kriging (SK)

Standard stochastic kriging (SK), which was first introduced by Ankenman et al.,¹ 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

$$\mathcal{Y}_j(\mathbf{x}) = Y(\mathbf{x}) + \varepsilon_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + \mathbf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x}), \quad (\text{S1})$$

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: $Y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + \mathbf{M}(\mathbf{x})$. $\mathbf{f}(\mathbf{x})$ 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,¹ this work adopts $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta_0$ unless stated otherwise. The term \mathbf{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 $\mathbf{M}(\mathbf{x})$ and $\mathbf{M}(\mathbf{x}')$ 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),¹ the stochastic nature of $\mathbf{M}(\mathbf{x})$ 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

$$\bar{\mathcal{Y}} = (\bar{\mathcal{Y}}(\mathbf{x}_1), \bar{\mathcal{Y}}(\mathbf{x}_2), \dots, \bar{\mathcal{Y}}(\mathbf{x}_I))^{\top} \quad (\text{S2})$$

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

$$\bar{\varepsilon} = (\bar{\varepsilon}(\mathbf{x}_1), \bar{\varepsilon}(\mathbf{x}_2), \dots, \bar{\varepsilon}(\mathbf{x}_I))^{\top}, \quad (\text{S3})$$

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

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

$$\text{Cov}[\mathbf{M}(\mathbf{x}), \mathbf{M}(\mathbf{x}')] = \delta^2 \cdot K(\mathbf{x}, \mathbf{x}') \quad (\text{S4})$$

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;² Qian et al. 2008³), and one of the most popular structures in practice is the

exponential correlation function

$$K(\mathbf{x}, \mathbf{x}') = \exp \left\{ \sum_{h=1}^d -\theta_h |x_h - x'_h|^p \right\}. \quad (\text{S5})$$

In (S5), $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ 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.⁴

With a selected correlation function such as (S5), the $I \times I$ variance-covariance matrix $\Sigma_{\mathbf{M}}$ is constructed as follows for a data set including I distinct design points

$$\Sigma_{\mathbf{M}} = \delta^2 \mathbf{R}(\boldsymbol{\theta}) = \delta^2 \begin{pmatrix} 1 & K(\mathbf{x}_1, \mathbf{x}_2) & \dots & K(\mathbf{x}_1, \mathbf{x}_I) \\ K(\mathbf{x}_2, \mathbf{x}_1) & 1 & \dots & K(\mathbf{x}_2, \mathbf{x}_I) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{x}_I, \mathbf{x}_1) & K(\mathbf{x}_I, \mathbf{x}_2) & \dots & 1 \end{pmatrix}. \quad (\text{S6})$$

In (S6), $\mathbf{R}(\boldsymbol{\theta})$ 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

$$\Sigma_{\mathbf{M}}(\mathbf{x}_0, \cdot) = \delta^2 \mathbf{v}(\mathbf{x}_0, \boldsymbol{\theta}) = \delta^2 \begin{pmatrix} K(\mathbf{x}_0, \mathbf{x}_1) \\ K(\mathbf{x}_0, \mathbf{x}_2) \\ \vdots \\ K(\mathbf{x}_0, \mathbf{x}_I) \end{pmatrix}, \quad (\text{S7})$$

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 (S3). Under the i.i.d assumption for random errors, Σ_ε is a diagonal matrix

$$\Sigma_\varepsilon = \text{diag}\{\text{Var}[\varepsilon(\mathbf{x}_1)]/n(\mathbf{x}_1), \text{Var}[\varepsilon(\mathbf{x}_2)]/n(\mathbf{x}_2), \dots, \text{Var}[\varepsilon(\mathbf{x}_I)]/n(\mathbf{x}_I)\}. \quad (\text{S8})$$

Estimation and Inference by Standard Stochastic Kriging

Recall that the random response can be written as

$$\mathcal{Y}_j(\mathbf{x}) = \beta_0 + \mathbf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x}), \quad (\text{S9})$$

As stated in Ankenman et al. (2010),¹ the SK-based modeling and inference requires the following assumption.

Assumption 1 *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. 2003,²), 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 (S6). 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 1, the log-likelihood function of the unknown parameters $(\beta_0, \delta^2, \boldsymbol{\theta})$ is

$$\ln \mathcal{L}(\beta_0, \delta^2, \boldsymbol{\theta}) = -\ln[(2\pi)^{I/2}] - \frac{1}{2} \ln[|\delta^2 \mathbf{R}(\boldsymbol{\theta}) + \Sigma_\varepsilon|] - \frac{1}{2} (\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_I)^\top [\delta^2 \mathbf{R}(\boldsymbol{\theta}) + \Sigma_\varepsilon]^{-1} (\bar{\mathcal{Y}} - \beta_0 \mathbf{1}_I), \quad (\text{S10})$$

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

Ankenman et al. (2010)¹ summarizes the SK modeling steps as follows.

1. Obtain the estimated Σ_ε :

$$\widehat{\Sigma}_\varepsilon = \text{diag}\{\widehat{\text{Var}}[\varepsilon(\mathbf{x}_1)]/n(\mathbf{x}_1), \widehat{\text{Var}}[\varepsilon(\mathbf{x}_2)]/n(\mathbf{x}_2), \dots, \widehat{\text{Var}}[\varepsilon(\mathbf{x}_I)]/n(\mathbf{x}_I)\} \quad (\text{S11})$$

where

$$\widehat{\text{Var}}[\varepsilon(\mathbf{x}_i)] = \frac{1}{n(\mathbf{x}_i) - 1} \sum_{j=1}^{n(\mathbf{x}_i)} (\mathcal{Y}_j(\mathbf{x}_i) - \bar{\mathcal{Y}}(\mathbf{x}_i))^2, \quad i = 1, 2, \dots, I. \quad (\text{S12})$$

2. Using $\widehat{\Sigma}_\varepsilon$ instead of Σ_ε , maximize the log-likelihood (S10) over $(\widehat{\beta}_0, \widehat{\delta}^2, \widehat{\boldsymbol{\theta}})$.
3. Estimate the expected response $\mathbf{Y}(\mathbf{x}_0)$ by

$$\widehat{\mathbf{Y}}(\mathbf{x}_0) = \widehat{\beta}_0 + \widehat{\delta}^2 \mathbf{v}(\mathbf{x}_0, \widehat{\boldsymbol{\theta}})^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}) + \widehat{\Sigma}_\varepsilon]^{-1} (\bar{\mathcal{Y}} - \widehat{\beta}_0 \mathbf{1}_I), \quad (\text{S13})$$

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

$$\widehat{\text{MSE}}[\widehat{\mathbf{Y}}(\mathbf{x}_0)] = \widehat{\delta}^2 - \widehat{\delta}^4 \mathbf{v}(\mathbf{x}_0, \widehat{\boldsymbol{\theta}})^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}) + \widehat{\Sigma}_\varepsilon]^{-1} \mathbf{v}(\mathbf{x}_0, \widehat{\boldsymbol{\theta}}) + \eta^\top \eta (\mathbf{1}_I^\top [\widehat{\delta}^2 \mathbf{R}(\widehat{\boldsymbol{\theta}}) + \widehat{\Sigma}_\varepsilon]^{-1} \mathbf{1}_I)^{-1} \quad (\text{S14})$$

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 $\mathbf{Y}(\mathbf{x}_0)$ is

$$\widehat{\mathbf{Y}}(\mathbf{x}_0) \pm z_{(1-\alpha)/2} \sqrt{\widehat{\text{MSE}}[\widehat{\mathbf{Y}}(\mathbf{x}_0)]} \quad (\text{S15})$$

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

Review of 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.⁵ for details. In this part, a brief review is provided for the readers' convenience.

Formulation and Assumptions for Mixed-Effects Model (MEM)

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

$$\mathcal{Y}_j(\mathbf{x}, c_q) = g(\mathbf{x}, \boldsymbol{\alpha}_{c_q}) + \varepsilon_j(\mathbf{x}, c_q); \quad q = 1, 2, \dots, Q, \quad (\text{S16})$$

In Equation (S16), 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; and \mathbf{x} denotes the vector of quantitative factors.

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 :

$$\text{Var}[\varepsilon(\mathbf{x}, c_q)] = \sigma^2 h^2(g(x, \boldsymbol{\alpha}_{c_q}), \boldsymbol{\gamma}); \quad q = 1, 2, \dots, Q \quad (\text{S17})$$

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

$$\boldsymbol{\alpha}_{c_q} = \mathbf{d}(\boldsymbol{\nu}, \boldsymbol{\alpha}) + \mathbf{b}_{c_q}, \quad (\text{S18})$$

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 STATEMENT OF THE RESEARCH PROBLEM, all the qualitative factors in \mathbf{z} are candidate components for ν . 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

$$\mathbf{b}_{c_q} \sim \text{Norm}(\mathbf{0}, \mathbf{D}) \quad (\text{S19})$$

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

The simplest example form of (S18) is

$$\boldsymbol{\alpha}_{c_q} = \boldsymbol{\alpha} + \mathbf{b}_{c_q}, \quad (\text{S20})$$

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.⁶ 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 (S17) is used to describe the dependence of variance upon the response. (iii) As given in (S19), 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.

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.⁵⁻⁷ 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 esti-

mated variance model; and $\widehat{\mathbf{D}}$, the estimated covariance matrix of $\widehat{\boldsymbol{\alpha}}_{c_q}$.

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

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

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

$$\widehat{Y}(\mathbf{x}_0, c_q) \pm t_{1-\alpha/2, v} \times \sqrt{\widehat{\text{Var}}[\widehat{Y}(\mathbf{x}_0, c_q)]} \quad (\text{S21})$$

where t denotes Student's t distribution with degree of freedom v . Following the context of BMD estimation in Subsection Inverse Estimation and Inference, 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:

$$\widehat{\text{BMD}}(c_q) = g^{-1}(\text{BMR}, \widehat{\boldsymbol{\alpha}}_{c_q}). \quad (\text{S22})$$

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:

$$\widehat{\text{Var}}[\widehat{\text{BMD}}(c_q)] = (g\boldsymbol{\alpha}^{-1})^\top \widehat{\mathbf{D}} g\boldsymbol{\alpha}^{-1} \quad (\text{S23})$$

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:

$$[\widehat{\text{BMD}}(c_q) - t_{1-\alpha, v} \sqrt{\widehat{\text{Var}}[\widehat{\text{BMD}}(c_q)]}, \infty) \quad (\text{S24})$$

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

True Expected Exposure-Response Model for Case 1

As part of the simulation model for Case 1 (Subsection Case 1: Modeling Dose-Time-Response Data), the following two models (S25-S26) provide the true expected response surfaces for the two subpopulations (short and long TiO_2 nanobelts) respectively. Both of the models take the form of a single-hidden layer feedforward network (SLFN),^{8,9} and are estimated from the real dose-time-response data in NIOSH's in-vivo study of TiO_2 nanopar-

titles.¹⁰

$$Y(\mathbf{x}, c_1) = 67.55 + \frac{0.14}{\exp(-12.85 + 0.41x_1 + 0.06x_2)} + \frac{-33.23}{\exp(-1.39 + 0.12x_1 - 0.02x_2)} \quad (\text{S25})$$

$$\begin{aligned} & + \frac{-0.08}{\exp(-8.57 + 1.22x_1 - 0.2x_2)} + \frac{-76.00}{\exp(-0.21 + 0.06x_1 - 0.01x_2)} \\ & + \frac{-3.49}{\exp(-0.40 + 0.20x_1 - 0.04x_2)} + \frac{-76.99}{\exp(-0.51 - 0.05x_1 + 0.03x_2)} \\ & + \frac{5.15}{\exp(-5.52 + 0.16x_1 + 0.04x_2)} + \frac{-0.03}{\exp(-25.89 + 2.99x_1 + 0.28x_2)} \\ & + \frac{32.77}{\exp(-0.62 + 0.05x_1 + 0.03x_2)} + \frac{1.09}{\exp(5.76 + 0.26x_1 - 0.07x_2)} \\ Y(\mathbf{x}, c_2) = & 71.74 + \frac{-104.84}{\exp(-0.66 + 0.09x_1 - 0.02x_2)} + \frac{-1.18}{\exp(7.66 - 0.45x_1 - 0.01x_2)} \quad (\text{S26}) \\ & + \frac{0.04}{\exp(28.01 - 3.62 + 0.24x_2)} + \frac{-4.50}{\exp(5.02 - 0.16x_1 - 0.04x_2)} \\ & + \frac{82.50}{\exp(0.74 + 0.04x_1 - 0.03x_2)} + \frac{-27.73}{\exp(0.83 + 0.13x_1 - 0.02x_2)} \\ & + \frac{35.84}{\exp(-1.18 + 0.04x_1 + 0.03x_2)} + \frac{-13.58}{\exp(-0.85 - 0.18x_1 + 0.03x_2)} \\ & + \frac{-2.61}{\exp(0.32 - 0.27x_1 + 0.01x_2)} + \frac{2.59}{\exp(-8.86 - 0.05x_1 + 0.08x_2)}. \end{aligned}$$

An Estimation Data Set (EDS) for Case 2

Table S1 shows an estimation data set (EDS) randomly generated from the simulation models of Case 2 (Subsection Case 2: Modeling Multi-Source Dose-Response Data). 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 S1: 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) Ankenman, B.; Nelson, B. L.; Staum, J. Stochastic kriging for simulation metamodeling. *Operations Research* **2010**, *58*, 371–382.
- (2) Santner, T. J.; Williams, B. J.; Notz, W. *The design and analysis of computer experiments*; Springer, 2003.
- (3) Qian, P. Z.; Wu, H.; Wu, C. J. Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* **2008**, *50*.
- (4) Rasmussen, C. E.; Christopher, K. I. W. *Gaussian processes for machine learning*; The MIT Press, 2006.
- (5) Davidian, M.; Giltinan, D. M. *Nonlinear models for repeated measurement data*; Chapman & Hall, 1995.

- (6) Giltinan, D. M.; Davidian, M. Assays for recombinant proteins: A problem in non-linear calibration. *Statistics in Medicine* **1994**, *13*, 1165–1179.
- (7) Steimer, J.-L.; Mallet, A.; Golmard, J.-L.; Boisvieux, J.-F. Alternative approaches to estimation of population pharmacokinetic parameters: comparison with the nonlinear mixed-effect model. *Drug metabolism reviews* **1984**, *15*, 265–292.
- (8) Huang, G.-B.; Chen, Y.-Q.; Babri, H. A. Classification ability of single hidden layer feedforward neural networks. *Neural Networks, IEEE Transactions on* **2000**, *11*, 799–801.
- (9) Yang, F. Neural network metamodeling for cycle time-throughput profiles in manufacturing. *European Journal of Operational Research* **2010**, *205*, 172–185.
- (10) Porter, D. W. et al. Differential mouse pulmonary dose and time course responses to titanium dioxide nanospheres and nanobelts. *Toxicological Sciences* **2013**, *131*, 179–193.