

# UCLA

## UCLA Previously Published Works

### Title

Relating Nanoparticle Properties to Biological Outcomes in Exposure Escalation Experiments.

### Permalink

<https://escholarship.org/uc/item/1j9413cg>

### Journal

Environmetrics, 25(1)

### ISSN

1180-4009

### Authors

Patel, T

Low-Kam, C

Ji, Zx

et al.

### Publication Date

2014-02-01

### DOI

10.1002/env.2246

Peer reviewed

Published in final edited form as:

*Environmetrics*. 2014 February 1; 25(1): 57–68. doi:10.1002/env.2246.

## Relating Nanoparticle Properties to Biological Outcomes in Exposure Escalation Experiments

T. Patel<sup>a</sup>, D. Telesca<sup>a,b,\*</sup>, C. Low-Kam<sup>a,b</sup>, ZX. Ji<sup>b</sup>, HY. Zhang<sup>b</sup>, T. Xia<sup>b,c</sup>, J.I. Zinc<sup>b,d</sup>, and A. E. Nel<sup>b,c</sup>

<sup>a</sup>Department of Biostatistics, UCLA

<sup>b</sup>California Nanosystems Institute, UCLA

<sup>c</sup>Division of Nanomedicine, UCLA

<sup>d</sup>Department of Chemistry and Biochemistry, UCLA

### Abstract

A fundamental goal in nano-toxicology is that of identifying particle physical and chemical properties, which are likely to explain biological hazard. The first line of screening for potentially adverse outcomes often consists of exposure escalation experiments, involving the exposure of micro-organisms or cell lines to a library of nanomaterials. We discuss a modeling strategy, that relates the outcome of an exposure escalation experiment to nanoparticle properties. Our approach makes use of a hierarchical decision process, where we jointly identify particles that initiate adverse biological outcomes and explain the probability of this event in terms of the particle physicochemical descriptors. The proposed inferential framework results in summaries that are easily interpretable as simple probability statements. We present the application of the proposed method to a data set on 24 metal oxides nanoparticles, characterized in relation to their electrical, crystal and dissolution properties.

### Keywords

Dose-Response Models; Model Selection; Nanoinformatics; Smoothing Splines

## 1. INTRODUCTION

Nanomaterials are a large class of substances engineered at the molecular level to achieve unique mechanical, optical, electrical and magnetic properties. Unusual properties of these particles can be attributed to their small size (with one dimension less than 100nm), chemical composition, surface structure, solubility, shape and aggregation behavior. One often refers to these particle characteristics as physicochemical properties. These very properties allow for increasingly diverse biological interactions, therefore defining potential

\*Correspondence to: D. Telesca, UCLA Fielding School of Public Health, Department of Biostatistics. dtelesca@ucla.edu.

### SUPPLEMENTARY MATERIALS

Web Appendices and supplementary materials referenced in the manuscript are available with this paper at the *Environmetrics* website on Wiley Online Library.

hazard concerns. Furthermore, the potential bio-hazard of these compounds is coupled with increased likelihood of environmental and occupational release as nano-tech applications are ever growing in science, medicine and industry (Nel *et al.* 2006; Stern and McNeil 2008).

A first line of screening for potentially hazardous particles is often represented by *in vitro* studies, where micro-organisms or cell cultures are exposed to nanomaterials in a controlled experimental setting. A typical *in vitro* study would measure a biological outcome (LDH, MTS, etc.) in relation to a dose escalation protocol, though, more generally, *in vitro* studies aim to measure how a biological response  $y(w)$  changes in relation to a prescription of increased particle exposure  $w$ , varying according to an exposure escalation protocol  $\mathcal{W}$ , (Meng *et al.* 2010; George *et al.* 2011; Zhang *et al.* 2011). In Fig. 1, for example, we illustrate a more general design, where 24 metal-oxide nanoparticles are monitored in relation to a cellular membrane damage assay (Propidium Iodide (PI) absorption), measured contemporaneously over a grid of eleven doses and seven hours of exposure.

A first statistical challenge is encountered in relation to modeling the outcome  $y(w)$ . Standard dose-response models focus almost exclusively on dose-escalation protocols and often rely on parametric assumptions, defining dose-response relationships as sigmoidal curves (Ritz 2010, Piegorsch *et al.* (2013)). This practice is likely not warranted, when considering more general exposure escalation designs. Sigmoidal dose-response curves may be inadequate, for example, if one considers non-lethal assays (intracellular calcium flux, mitochondrial depolarization, etc.) as they carry assumptions of monotonicity. Furthermore, particle solutions at varying concentrations may result in different particle aggregation behavior, potentially reducing total cellular exposure at higher concentrations (Hinderliter *et al.* 2010). These considerations challenge standard dose-response summaries, e.g. effective concentrations ( $EC_{\alpha}$ ), slope coefficients, no adverse effect dose levels, benchmark dose levels, etc., and motivate our work. In §2 we replace parametric assumptions about dose-response relationships with a general smooth surface response model over  $\mathcal{W}$  and propose a statistical notion of toxicity, based on mixture modeling (McLachlan and Peel 2000).

Beyond the identification of potentially hazardous particles in the observed sample, from a predictive and inferential perspective it is important to understand how particles physical and chemical properties affect the way they interact at the nano-bio interface. The use of a combination of molecular properties to predict a compound's behavior with respect to biological end-points is a well accepted concept in the predictive toxicology of chemicals (Schultz *et al.* 2003). In nano-informatics, the literature refers to this class of models as nano quantitative structure-activity relationships (nano-QSAR) models (Shaw *et al.* 2008; Puzyn *et al.* 2011; Liu *et al.* (2011)). Shaw *et al.* (2008) present a nano-QSAR model based on hierarchical and consensus clustering. Nanoparticles are classified into groups and clustering patterns are informally compared in relation to particles properties, that tend to be assigned to the same class. Puzyn *et al.* (2011) characterize particle toxicity in terms of the dose concentration that brought about 50% reduction in bacterial cell viability ( $EC_{50}$ ). Given this summary, the relationship between  $EC_{50}$  values and particle descriptors is modeled via linear regression, with selection of covariates based on minimizing a cross validation error. Similarly, Liu *et al.* (2011) use logistic regression to model the probability that a particle is toxic, given covariates. A particle is defined as toxic at a given concentration, based on a

statistically significant difference in mean response, as compared to the background response in unexposed cells.

These techniques often rely on previously defined data summaries (effective concentrations, dose-response slopes, benchmark doses, etc.) and fail to account for the uncertainty inherent in the estimation of these summaries, when relating them to covariate information (particles physicochemical properties). We propose a modeling strategy that provides a definition of toxicity in relation to very general exposure escalation protocols and, at the same time, represents the probability of toxicity as a function of particle physical and chemical properties, without the loss of information associated with step-wise procedures. The proposed methodology is appropriate for limited data sets and easily includes data integration and a framework for advanced dimension reduction through variable selection.

The remainder of this manuscript is organized as follows. In §2 we introduce the proposed model. In §3 we discuss parameter estimation and associated inferential details. We illustrate the application of the proposed methodology in §4 through the analysis of 24 metal-oxide nanomaterials. We discuss our proposal, potential limitations and extensions in §5.

## 2. MODEL FORMULATION

### 2.1. A Statistical Definition of Toxicity

Standard protocols for in-vitro assays aim to measure the biological response  $y_{ij}(w)$  associated with exposure to particle  $i$ , ( $i = 1, \dots, n$ ). The experiment is replicated  $m$  times, with replicates indexed by  $j$ , ( $j = 1, \dots, m$ ) and the response  $y$  is evaluated with respect to a protocol of increased exposure to nanoparticles indexed by  $w \in \mathcal{W} \subset \mathbb{R}^{\nu+}$ . Increased exposure is most often implemented as a dose escalation experiment, in which case  $\nu = 1$  and  $\mathcal{W} = [0, D]$ , where  $D$  is the largest dose. However, experimental protocols considering both dose escalation and repeated measurements, obtained over a total exposure duration interval ( $t \in (0, T]$ ), are becoming more prevalent (George *et al.* 2011; Patel *et al.* 2011). In this last case  $\nu = 2$  and  $\mathcal{W} = [0, D] \times [0, T]$ . We assume that raw data have been normalized and purified from experimental artifacts.

In this general setting, it is indeed difficult to identify an obvious notion of toxicity. In fact, some biological assays may not address toxicity all-together and simply measure sublethal biological reactions. To be precise, in the absence of reasonable expectations for a standard sigmoidal dose-response curve, common measures of toxicity, including effective concentrations or hill coefficients are clearly ill-defined. Some progress can be made when considering a test library of  $n$  nanomaterials, by defining toxicity using empirical evidence about the existence of subgroups of particles. Our approach is based on defining two possible exposure-response categories, as follows:

$$\gamma_i = 0 \text{ if particle } i \text{ is non-toxic} \quad \gamma_i = 1 \text{ if particle } i \text{ is toxic.}$$

The foregoing model supports both the statistical definition of toxicity, via a mixture approach, and the formulation of probability statements describing the relationships between

toxicity and nanoparticle properties. Let  $\mathbf{y}_i$  be a vector of measurements collecting all evidence associated with particle  $i$ .<sup>†</sup> For each particle  $i$ , we define a sampling model

$$\mathbf{y}_i | \gamma_i \sim f_{i\gamma_i}(\cdot), \gamma_i \in \{0, 1\}, i=1, \dots, n.$$

Densities  $f_{i0}(\cdot)$  describe sampling variability over the entire exposure escalation domain  $\mathcal{W}$ , which is not distinguishable from the the case of no exposure. This invariance of  $f_{i0}(\cdot)$  with respect to  $\mathcal{W}$  allows for the interpretation of  $f_{i0}(\cdot)$  as a model for normal biological variation associated with exposure to non-toxic particles. Densities  $f_{i1}(\cdot)$  describe sampling variability as a function of the exposure escalation protocol  $\mathcal{W}$ , therefore allowing for the interpretation of  $f_{i1}(\cdot)$  as model of biological variation associated with exposure to toxic or bio-active particles. If we identify with  $\pi_i = P(\gamma_i = 1)$  the pre-experimental probability of particle  $i$  being toxic, then for each particle we can readily convert exposure escalation measurements  $\mathbf{y}_i$  into a probability of toxicity  $p_i^+$ . Specifically, we have

$$p_i^+ = P(\gamma_i = 1 | \mathbf{y}_i, f_{i0}, f_{i1}, \pi_i) = \frac{\pi_i f_{i1}(\mathbf{y}_i)}{(1 - \pi_i) f_{i0}(\mathbf{y}_i) + \pi_i f_{i1}(\mathbf{y}_i)}. \quad (1)$$

These quantities offer an interpretable and unequivocal scale of evidence in the classification of potentially hazardous nanomaterials. From a modeling perspective, the limitations of standard dose-response analysis and the importance of mixture modeling ideas have a been recognized by several authors (Boos and Brownie 1991, Piegorsch *et al.* 2013), even though, to our knowledge, we are the first to apply this strategy in the joint analysis of multiple biological stressors, specifically nanomaterials.

## 2.2. Distributional Assumptions and Semi-parametric Representation

Let  $y_{ij}(w)$  denote a response corresponding to nanoparticle  $i$  ( $i = 1, \dots, n$ ) and replicate  $j$  ( $j = 1, \dots, m$ ), at exposure  $w \in \mathcal{W}$ . In practice, observations are obtained over a discrete exposure sampling grid  $(w_1, \dots, w_k) \in \mathcal{W} \subset \mathbb{R}^{\nu^+}$ . However, for ease of presentation and without loss of generality, we assume that exposure points are defined in a continuous domain. We make precise our definition of toxicity by introducing the following smooth response model

$$y_{ij}(w) = \alpha_i + \pi_i m_i(w) + \varepsilon_{ij}(w), \quad (2)$$

where  $\alpha_i$  represents the background signal at exposure zero,  $m_i(w)$  is an exposure-response curve or surface,  $\varepsilon_{ij}(w) \sim N(0, \sigma_\varepsilon^2 / \tau_i)$  and  $\tau_i \mid \nu \sim \text{Gamma}(\nu/2, \nu/2)$ .

More precisely, assuming  $m_i(w) = 0$  for at least one  $w \in \mathcal{W}$ , the forgoing representation describes a particle  $i$  to be bio-active with probability  $\pi_i$ . The coefficient  $\alpha_i$  is interpreted as the background signal and  $m_i(w): \mathcal{W} \rightarrow \mathbb{R}$  is interpreted as the above-background response hyper-surface, defining signal variation over the exposure-escalation domain  $\mathcal{W}$ . More stringent constraints can be required if there is no possibility of cellular proliferation. In

<sup>†</sup>We will make our notation more precise later, as specific definitions of dimension and structure will depend on the specific design of exposure escalation. For now, we find it useful to maintain our notation at a high degree of generality.

which case:  $m_i(w) = 0$  for all  $w \in \mathcal{W}$  and  $m_i(w) > 0$  for at least some  $w \in \mathcal{W}$  defines a model of strict toxicity. For the moment we think of  $m_i(w)$  as a general smooth function defined over a  $p$ -dimensional space and describe practical representation issues later in the manuscript.

The sampling distribution of  $y_{ij}(w)$  is modeled in terms of the error term  $\varepsilon_{ij}(w)$  as a scale mixture of normal random variables. The error variance is defined in terms of the measurement error  $\sigma_\varepsilon^2$ , and a particle-specific random variance inflation parameter  $\tau_i$ , resulting marginally in a t-distributed error structure, with  $\nu$  degrees of freedom (West 1984).

It is convenient to introduce a data-augmented representation of the model in (2). More precisely, we define a  $n$ -dimensional vector of latent binary indicators  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_n)' \in \{0, 1\}^n$ , with  $\gamma_i | \boldsymbol{\pi}_i \sim \text{Bern}(\boldsymbol{\pi}_i)$ . Conditioning on  $\boldsymbol{\gamma}_i$  we have

$$E[y_{ij}(w) | \gamma_i] = \begin{cases} \alpha_i & \text{if } \gamma_i = 0 \\ \alpha_i + m_i(w) & \text{if } \gamma_i = 1. \end{cases} \quad (3)$$

This representation makes explicit the trans-dimensional nature of the proposed model and will be used as a key component in the construction of a covariate model. We note that likelihood identifiability is contingent on the condition  $m_i(w) = 0$  for at least one  $w \in \mathcal{W}$ , which includes possible positivity restrictions on  $w_i(w)$  as a special case.

### 2.3. Modeling Dependence on Covariates

Nanoparticles are often described in terms of their physical and chemical properties. We propose to use these descriptors to inform particles probabilities of toxicity. Let  $\mathbf{X}$  denote an  $n \times p$  covariate matrix,  $\mathbf{x}_i$  be the  $i$ -th row of  $\mathbf{X}$  and  $g(\cdot): [0, 1] \rightarrow \mathbb{R}$  denote a general link function. We relate toxicity indicators  $\gamma_i$  to  $\mathbf{x}_i$  as follows

$$g(P(\gamma_i=1 | \boldsymbol{\lambda}; \mathbf{x}_i)) = \mathbf{x}_i' \boldsymbol{\lambda}, \quad (4)$$

where  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)'$  is a vector of regression coefficients. Common implementations define  $g(\cdot)$  as a probit or logit link, often according to computational or analytical convenience. The specific interpretation of  $\boldsymbol{\lambda}$  will depend on  $g(\cdot)$ . However, the vector  $\boldsymbol{\lambda}$  can be generally interpreted as a summary of how variability in the particle properties explain variability in the probability of toxicity. This formulation is inherently related to the class of hierarchical mixture of experts introduced by Jordan and Jacobs (1994), as the probability of particle bio-active potential is now informed by the particle's physicochemical characteristics  $\mathbf{x}_i$ .

For regularization purposes, especially in instances where  $p$  is large in relation to  $n$ , it may be appropriate to consider covariate subset selection (George and McCulloch 1993). A possible representation of the problem relies again on data-augmentation, by defining a  $p$ -dimensional vector of variable inclusion indicators  $\boldsymbol{\rho} = (\rho_1, \dots, \rho_p)' \in \{0, 1\}^p$  and a reduced design matrix  $\mathbf{X}_\rho$ , which only includes the columns of  $\mathbf{X}$  corresponding to the non-zero

elements of  $\boldsymbol{\rho}$ . Conditioning on  $\boldsymbol{\rho}$ , we assume  $g(P(\gamma_i=1|\boldsymbol{\lambda}, \boldsymbol{\rho}; \mathbf{x}_{i\rho})) = \mathbf{x}'_{i\rho} \boldsymbol{\lambda}_\rho$ . In the foregoing model,  $\boldsymbol{\lambda}_\rho$  is interpreted as a  $|\boldsymbol{\rho}|$ -dimensional vector that includes all nonzero elements of  $\boldsymbol{\lambda}$ . Alternative regularization strategies are easily implemented and may rely, for example, on penalized likelihood formulations (Tibshirani 1996).

#### 2.4. Implementation for a Dose-Duration Escalation Protocol

Our discussion in §2.2 assumed a known exposure response hyper-surface  $m_i(w): \mathcal{W} \rightarrow \mathbb{R}^\nu$ . Here we discuss practical representation details in relation to a dose-duration escalation protocol. We illustrate this specific case as experimental protocols designed to monitor a biological outcome over several doses and durations of exposure to nanomaterials are becoming increasingly prevalent (George *et al.* 2011; Patel *et al.* 2011). Furthermore, standard dose escalation protocols are easily obtained as a special case of this more general design.

Specifically, we consider the case of  $\mathcal{W} = [0, D] \times (0, T]$ , so that the dose-response surface  $m_i(w) = m_i(d, t): [0, D] \times [0, T] \rightarrow \mathbb{R}$  spans two dimensions: dose  $d \in [0, D]$  and time  $t \in [0, T]$ . Requiring that  $m_i(d=0, t=0) = 0$ , allows for the interpretation of  $\alpha_i$  as background fluorescence. Furthermore, requiring  $m_i(d, t) > 0$  for at least one dose/time combination insures likelihood identifiability.

It is often appropriate to think of  $m_i(\cdot)$  as a smooth function, therefore non-parametric representations may rely on two-dimensional P-splines (Lang and Brezger 2006). The literature about the application of splines to the analysis of dose-response data has often focused on the estimation of curves under shape restrictions (Ramsay 1988; Desquilbet and Mariotti 2010; Wang and Shel 2010). The topic is too vast to review here. However, we point out that strict topological constraints may not be warranted in nanotox experiments and therefore recommend simpler penalization strategies.

We consider the representation of  $m_i(d, t)$  using  $s_d$  and  $s_t$  equally spaced interior knots

$$0 < \eta_{1_d} < \eta_{2_d} < \dots < \eta_{s_d-1} < \eta_{s_d} < D; \text{ and } 0 < \eta_{1_t} < \eta_{2_t} < \dots < \eta_{s_t-1} < \eta_{s_t} < T.$$

Each set of interior knots generates a corresponding set of  $M_d = s_d + \ell$  and  $M_t = s_t + \ell$  B-spline basis functions of order  $\ell$ . Let  $\mathcal{B}_m(\cdot)$  denote the  $m^{\text{th}}$  basis and  $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{i, M_d M_t}, \dots, \beta_{i, M_d M_t})'$  be a  $(M_d M_t)$ -dimensional vector of spline coefficients. The function  $m_i(d, t; \boldsymbol{\beta}_i)$  can then be represented as the tensor product of one-dimensional B-splines:

$$m_i(d, t; \boldsymbol{\beta}_i) = \sum_{m_d=1}^{M_d} \sum_{m_t=1}^{M_t} \beta_{i, m_d m_t} \mathcal{B}_{m_t}(t) \mathcal{B}_{m_d}(d); i=1, \dots, n. \quad (5)$$

Using the foregoing representation, the identifiability restriction  $m_i(d=0, t=0; \boldsymbol{\beta}_i) = 0$ , is implemented by fixing  $\beta_{i1} = 0$ , for all particles. Similarly, if positivity of  $m_i(\cdot)$  is required, it is easily implemented by restricting  $\boldsymbol{\beta}_i \in \mathbb{R}^{M_d M_t+}$  for all  $i = 1, \dots, n$ . In our implementation, we will simply require that not all  $\beta_{i\ell}$  equal 0, ( $i = 1, \dots, n$ ;  $\ell = 1, \dots, M_d M_t$ ).

Model complexity and flexibility in  $m_i(d, t)$  is a function of the spline order  $\ell$  and the number of interior knots  $s_d$  in the dose and  $s_t$  in the time domain. Typical implementations rely on cubic splines  $\ell = 4$  and a large number of interior knots. Automatic smoothing is then obtained via regularization procedures that include penalized likelihood (Eilers and Marx 1996) or smoothing priors (Besag and Kooperberg 1995, Lang and Brezger 2004).

### 3. ESTIMATION AND INFERENCE

#### 3.1. Bayesian Hierarchical Representation

Estimation of the model parameters and inference can be carried out using several inferential strategies. In particular, penalized likelihood estimation could be implemented in order to account for regularization of the exposure response surface and regression coefficients in the covariate model.

A more natural representation of the proposed probability scheme is based on hierarchical models (Carlin and Louis 2000). This choice is guided by the fact that the proposed model is defined in terms of large vectors of related parameters and Bayesian hierarchical models lead to shrinkage estimates with good statistical properties (Berger 1985).

In this article we consider the following representation.

**Stage 1: Sampling Model.** We rewrite the model first introduced in (2), conditioning on all coefficients defining its semi-parametric structure. We have,

$$y_{ij}(w) | \gamma_i, \alpha_i, \beta_i, \sigma_\varepsilon^2, \tau_i \sim N \left( \alpha_i + I(\gamma_i=1) m_i(w, \beta_i), \sigma_\varepsilon^2 / \tau_i \right); \quad (6)$$

for particles  $i = 1, \dots, n; j = 1, \dots, m, w \in \mathcal{W}$ .

**Stage 2: Particle Level Model.** At the level of particle  $i$ , we specify a model for the toxicity indicators  $\gamma_i$  and the response surface coefficients  $\beta_i$ .

The probability of toxicity is assumed to depend on a subset of covariate  $\mathbf{X}_\rho$ , indexed by inclusion indicators  $\boldsymbol{\rho}$ . We follow Albert and Chib (1993) and introduce a  $n$ -dimensional vector of latent variables  $\mathbf{z} = (z_1, \dots, z_n)'$ , so that we can write

$$\gamma_i | z_i = I(z_i > 0), \text{ and } z_i | \boldsymbol{\lambda}, \boldsymbol{\rho}; \mathbf{x}_{i\rho} \sim N(\mathbf{x}'_{i\rho} \boldsymbol{\lambda}_\rho, 1). \quad (7)$$

The foregoing implementation defines a probit link for  $\boldsymbol{\gamma}$  and represents covariate selection as introduced in §2.3.

Distributional assumptions for  $\beta_i$  are based on considerations of smoothness and conjugacy (Besag and Kooperberg 1995, Lang and Brezger 2004). Specifically, we use a prior based on the Kronecker product of penalty matrices defined in of the dose and time domain, so that

$$\beta_i \sim N_{M_d M_t} \left( 0, \boldsymbol{\Sigma}_\beta = \sigma_\beta^2 (K_d \otimes K_t)^{-1} \right). \quad (8)$$

The penalty matrices  $K_d : M_d \times M_d$  and  $K_t : M_t \times M_t$ , are constructed as follows:



$$K_d = \begin{bmatrix} 1+\delta & -2 & 1 & 0 & \cdots & \cdots & 0 \\ -2 & 5 & -4 & 1 & \ddots & & \vdots \\ 1 & -4 & 6 & \ddots & \ddots & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & -4 & 1 & 0 \\ \vdots & \ddots & \ddots & -4 & 6 & -4 & 1 \\ \vdots & & \ddots & 1 & -4 & 5 & -2 \\ 0 & \cdots & \cdots & 0 & 1 & -2 & 1+\delta \end{bmatrix}; \delta > 0. \quad (9)$$

The matrix  $K_i$  is similarly constructed. A small positive constant  $\delta$  is introduced to insure propriety of the prior, without fundamentally changing the spatial smoothing properties of the proposed distribution. Detailed derivations and alternatives are discussed in Clayton (1996), as well as Lang and Brezger (2004).

For completeness the prior on the background level is chosen to be conjugate and non informative, so that  $\alpha_i | \sigma_\varepsilon^2, \tau_i \sim N(0, c\sigma_\varepsilon^2/\tau_i)$ , where  $c$  is set to a large value.

**Stage 3: Priors.** Our choice of priors is guided by conditional conjugacy as well as principles of multiplicity correction. In particular we consider a Zellner  $g$ -prior (Zellner 1986) for the covariate model coefficients  $\lambda_\rho | \rho \sim N(0, g_\rho(\mathbf{X}'_\rho \mathbf{X}_\rho)^{-1})$  and an exchangeable Bernoulli prior for the inclusion indicators  $\rho$ , so that  $p(\rho | \psi) = \psi^{|\rho|} (1 - \psi)^{p-|\rho|}$ . As described by Lee *et al.* (2003), choosing small values for  $\psi$ , leads to parsimonious models by restricting the number of covariates included. The values of  $\psi$  can also be tuned to include prior knowledge about the importance of certain physicochemical properties. Most commonly, a prior is placed on  $\psi$ , such that  $\psi \sim \text{Beta}(a_\pi, b_\pi)$ , and multiplicity correction is achieved automatically, as described by Scott and Berger (2006). A data dependent  $g_\rho$  is often recommended and we follow Fernandez *et al.* (2001) in setting  $g_\rho = \max(n, |\rho|^2)$ .

Finally, we complete the model by specifying conditionally conjugate prior distributions on variance components:<sup>‡</sup>

$$1/\sigma_\varepsilon^2 \sim \text{Gamma}(a_\varepsilon, b_\varepsilon); 1/\sigma_{\alpha_i}^2 \sim \text{Gamma}(a_{\alpha_i}, b_{\alpha_i}); 1/\sigma_{\beta_i}^2 \sim \text{Gamma}(a_{\beta_i}, b_{\beta_i}).$$

### 3.2. Posterior Simulation via MCMC

We base our inference on Markov Chain Monte Carlo (*MCMC*) simulations. Closed-form full conditional distributions are available for all parameters, therefore the proposed posterior simulation algorithm utilizes a Gibbs sampler to directly update parameters component-wise (Geman and Geman 1984, Gelfand and Smith 1990).

Specific steps in the simulation are inherently trans-dimensional and some care is needed in the identification of Gibbs components to be updated jointly. In particular, we implement a

<sup>‡</sup>We assume,  $x \sim \text{Gamma}(a, b)$  denotes a Gamma distributed random quantity with shape  $a$  and rate  $b$ , such that  $E(x) = a/b$ .

chain with transition kernels moving toxicity indicators  $\boldsymbol{\gamma}$  and covariate inclusion indicators  $\boldsymbol{\rho}$ , always with respect to their marginal posterior distributions, integrating over  $\boldsymbol{\beta}$  and  $\boldsymbol{\lambda}$  respectively. Given  $\boldsymbol{\gamma}$  and  $\boldsymbol{\rho}$ , the model structure and dimension is fully determined and other model parameters can be updated in a standard fashion. The proposed sampling scheme can be summarized as follows.

**1. Trans-dimensional updates**—We begin by drawing  $\boldsymbol{\gamma}$  from its marginalized conditional distribution, obtained by integrating the conditional posterior over  $\boldsymbol{\beta}$ . More precisely, for each particle  $i$  ( $i = 1, \dots, n$ )

$$p(\gamma_i | \mathbf{y}_i, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \sigma_\varepsilon^2, \boldsymbol{\tau}, \boldsymbol{\rho}) \propto \int p(\mathbf{y}_i | \boldsymbol{\beta}_i, \boldsymbol{\alpha}, \sigma_\varepsilon^2, \boldsymbol{\tau}) p(\boldsymbol{\beta}_i | \gamma_i, \boldsymbol{\Sigma}_\beta) d\boldsymbol{\beta}_i p(\gamma_i | \boldsymbol{\lambda}, \boldsymbol{\rho});$$

where defining  $\boldsymbol{\Omega}_{\beta_i} = \frac{\tau_i}{\sigma_\varepsilon^2} \mathbf{H}_i' \mathbf{H}_i + \boldsymbol{\Sigma}_\beta^{-1}$ , gives us

$$P(\gamma_i = 1 | \mathbf{y}_i, \boldsymbol{\alpha}, \boldsymbol{\lambda}, \sigma_\varepsilon^2, \boldsymbol{\tau}, \boldsymbol{\rho}) = \left[ \frac{\Phi(-\mathbf{x}'_{i\rho} \boldsymbol{\lambda}_\rho)}{\Phi(\mathbf{x}'_{i\rho} \boldsymbol{\lambda}_\rho) |\boldsymbol{\Sigma}_\beta|^{-\frac{1}{2}} |\boldsymbol{\Omega}_{\beta_i}|^{-\frac{1}{2}} \exp \left\{ \frac{\tau_i^2}{2\sigma_\varepsilon^4} \tilde{\mathbf{y}}_i' \mathbf{H}_i \boldsymbol{\Omega}_{\beta_i}^{-1} \mathbf{H}_i' \tilde{\mathbf{y}}_i \right\}} + 1 \right]^{-1}.$$

In the foregoing formula,  $\Phi(\cdot)$  is the cdf of a standard normal distribution. The derivation of this result, and precise construction of the vectors  $\tilde{\mathbf{y}}_i$  and matrices  $\mathbf{H}_i$  are provided in Appendix A. If  $\boldsymbol{\beta}$  is subjected to positivity restrictions, the foregoing integral must be evaluated over a positive hyper-quadrant, therefore requiring approximate Monte Carlo evaluation. Alternatively a reversible jumps transition scheme may be devised to move jointly over the space of  $(\gamma_i, \boldsymbol{\beta}_i)$  (Green 1995).

A similar scheme was adapted for sampling from  $\boldsymbol{\rho} | \boldsymbol{\lambda}, \mathbf{z}, \boldsymbol{\gamma}$ . The marginalized conditional distribution of  $\boldsymbol{\rho}$ , obtained by integrating over  $\boldsymbol{\lambda}$  is given by

$$p(\boldsymbol{\rho} | \mathbf{z}, \psi) \propto \int p(\mathbf{z} | \boldsymbol{\rho}, \boldsymbol{\lambda}) p(\boldsymbol{\lambda} | \boldsymbol{\rho}) d\boldsymbol{\lambda} p(\boldsymbol{\rho} | \psi) \propto \frac{1}{\sqrt{(1+g_\rho)}} \exp \left[ \frac{g_\rho}{2(1+g_\rho)} \mathbf{z}' \mathbf{X}_\rho (\mathbf{X}'_\rho \mathbf{X}_\rho)^{-1} \mathbf{X}'_\rho \mathbf{z} \right] \psi^{|\boldsymbol{\rho}|} (1-\psi)^{I-|\boldsymbol{\rho}|}.$$

**2. Fixed dimensional updates**—Given the current state of the latent indicators  $\boldsymbol{\gamma}$ , response surfaces are uniquely defined as in (3). Posterior sampling here is standard, and proceeds by updating spline coefficients  $\boldsymbol{\beta}_i$ , background response parameters  $\boldsymbol{\alpha}_i$ , and variance parameters  $\tau_i$  and  $\sigma_\varepsilon^2$  and  $\sigma_{\beta_i}^2$ , from their full conditional distributions via direct simulation. Similarly, given latent indicators  $\boldsymbol{\rho}$ , the covariate model is uniquely defined as in (7). Again, posterior sampling is standard and proceeds by updating regression coefficients  $\boldsymbol{\lambda}_\rho$  from  $p(\boldsymbol{\lambda}_\rho | \boldsymbol{\rho}, \mathbf{z})$ , and,  $\mathbf{z}$  from  $p(\mathbf{z} | \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\rho})$ , via Gibbs sampling. Full conditional distributions are given in Appendix A.

## 4. STRUCTURE ACTIVITY RELATIONSHIPS OF NANO METAL OXIDES

### 4.1. Case Study Background

We illustrate the proposed methodology by analyzing data on human bronchial epithelial cells (BEAS-2B), exposed to a library of 24 nano metal-oxides including: ZnO, CuO, CoO, Fe<sub>2</sub>O<sub>3</sub>, Fe<sub>3</sub>O<sub>4</sub>, WO<sub>3</sub>, Cr<sub>2</sub>O<sub>3</sub>, Mn<sub>2</sub>O<sub>3</sub>, Ni<sub>2</sub>O<sub>3</sub>, SnO<sub>2</sub>, CeO<sub>2</sub>, Al<sub>2</sub>O<sub>3</sub>, among others. We consider a measure of cellular membrane damage as measured by Propidium Iodine (PI) uptake. Specifically, cell cultures are treated with Propidium Iodide (PI) and in cells with damaged membranes PI is able to permeate the cell and bind to DNA, where it causes the nucleus to emit a red fluorescence. Each sample was also stained with a Hoechst dye, which causes all cell nuclei to emit a blue fluorescence, and allows for a count of the total number of cells. An analysis of the fluorescence readout, measured at varying wavelengths, results in a measure of the percentage of cells positive for each response. The outcome is measured over a grid of eleven doses and seven times (7-hours) of exposure over four replicates. The final response was purified from experimental artifacts and normalized using a logit transformation (Fig. 1).

All particles were characterized in terms of their size, dissolution, crystal structure, conduction energies, as well as many other particle descriptors (Zhang *et al.* 2012). Burello and Worth (2011) hypothesized that comparing the conduction and valence band energies to the redox potential of the reactions occurring within a biological system could predict nanoparticle toxicity of oxides. The normal cellular redox potential is in the range (−4.12 to −4.84 Ec). The conduction band energy of a particle is a measure of the energy sufficient to free an electron from an atom, and allow it to move freely within the material. Therefore, we are interested in the relationship between particles with conduction band energies within and outside the range of the cellular redox potential and their cytotoxicity profiles.

Another measure of interest is the metal dissolution rate of a particle. Particles that are highly soluble have the ability to shed metal ions, which can lead to nanoparticle toxicity (Xia *et al.* 2008; Xia *et al.* 2011; Zhang *et al.* 2012). Other potential risk factors include the particle primary size, a measure of length that defines the crystal structure ( $b(\text{Å})$ ), lattice energy ( $H_{\text{lattice}}$ ), which measures the strength of the bonds in the particles, and the enthalpy of formation of gaseous cation  $Me^{n+}$  ( $H_{Me^{n+}}$ ), which is a combined measure of the energy required to convert a solid to a gas and the energy required to remove  $n$  electrons from that gas.

### 4.2. Case Study Analysis and Results

The model described in §2 was fit to the metal-oxide data-set described in the previous section. We placed relatively diffuse  $\text{Gamma}(.01, .01)$  priors on the  $1/\sigma_e$  parameter,  $\text{Gamma}(1, .1)$  priors on all remaining precision parameters, and  $c = 100$  as the prior variance inflation constant of  $\alpha_i$ . We also placed a prior distribution on the degrees of freedom parameter  $\nu$ , for the t-distributed error described in Section 2.2. Specifically, the prior was modeled using a discrete uniform distribution on 1,2,4,8,16, and 32 degrees of freedom (Besag and Higdon 1999). Finally, for the covariate model we consider a default  $U(0, 1)$  prior for  $\psi$ .

In structuring the covariate matrix  $\mathbf{X}$ , we model  $\log(\text{dissolution})$  in a non-linear fashion as a spline function with a change point at  $\log(\text{dissolution})$  of 2.3. This choice is motivated by the fact that the nanoparticle library is really comprised of two separate classes, a set of particles with low dissolution and two particles (ZnO and CuO) with high dissolution, therefore it is wise to allow for a discontinuity at the boundary. Conduction band energy was coded as a binary covariate ( $I\{Ec \in (-4.12, -4.84)\}$ ), and all other ENM characteristics described above as continuous linear predictors.

Our inferences are based on 20,000 MCMC samples from the posterior distribution, after discarding a conservative 60,000 iterations for burn-in. MCMC sampling was performed in R version 2.10.0, and convergence diagnostics were performed using the package CODA (Convergence Diagnostics and Output Analysis), (Plummer *et al.* 2006).

Figure 2 illustrates data and the posterior expected dose-time response for a sample of particles examined in this study. Figure 3 provides a plot of the estimated probability of toxicity for each of the 24 nanomaterials. We find that CuO,  $\text{Mn}_2\text{O}_3$ , ZnO,  $\text{Cr}_2\text{O}_3$ , CoO,  $\text{CO}_3\text{O}_4$ , and  $\text{Ni}_2\text{O}_3$  nanomaterials have a pronounced dose/time effect, as compared to the other 17 materials. These seven particles all have posterior mean probabilities of toxicity above 0.5 (Fig. 3), suggesting that they are highly likely to induce cytotoxicity. It is important to stress the fact that our chosen cut-off value of 0.5 is, in a sense, arbitrary and that reporting results as we did in Figure 3 is essential to illustrate the full spectrum of bio-active potential in the particle sample.

An estimate of the inclusion probability,  $\hat{\rho}$  was used to select a single covariate model. Specifically we follow Scott and Berger (2006) by selecting the median model, that is including all covariates for which  $\hat{\rho}_k > 0.5$ , ( $k = 1, \dots, p$ ). The median model included  $\log(\text{metal dissolution})$  and conduction band energy. Table 1 provides estimates of posterior inclusion probabilities and model averaging posterior summaries. Results for the median model are reported in Table 2, where we provide posterior summaries for the model including conduction band energy and  $\log(\text{metal dissolution rate})$  alone.

These regression coefficients are easily interpreted as simple probability statements related to changes in the probability of toxicity, as one varies a particle's physical and chemical properties. Figure 4 illustrates summary plots of the posterior probability of toxicity (with 95% credible bands) as a function of conduction band energy and metal dissolution. The *bottom* panel of Figure 4 provides the posterior mean probability of toxicity, as a function of both conduction band energy and  $\log(\text{metal dissolution})$ . We find that conduction band energies inside the redox potential of the cell, predict high probability of cytotoxicity. Similar inference is drawn for particles with a high metal dissolution rate. The mechanistic interpretation of our findings is based on the idea that oxidizing and reducing substances can create an imbalance in the normal intracellular state of the cell, either by the production of oxygen radicals or by reducing antioxidant levels. In particular, the potential for oxidative stress might be predicted by comparing the energy structure of oxides, as measured by their conduction and valance band energy levels, to the redox potential of the biomolecules maintaining the cellular homeostasis. When these two energy levels are comparable, it can possibly allow for the transfer of electrons, and subsequent imbalance in the normal

intracellular state. Our model confirms the substantive findings in Burello and Worth (2011) and Zhang *et al.* (2012)).

Finally, we assess goodness of fit using posterior predictive checks. The conditional predictive ordinate (*CPO*), as defined by Geisser (1980), is the predictive density of observation  $\ell$  given all other observations, and can be used as a diagnostic tool for detecting observations with poor model fit. Large values of  $-\log(CPO)$  indicate observations that are not being fitted well. The *top* panel of Figure 5, provides a plot of  $-\log(C\hat{P}O_i(d, t))$  for our final model. In general, low values of  $-\log(CPO_i(d, t))$  indicate good model fit. The *middle* panel indicates that the largest values of  $-\log(C\hat{P}O_i(d, t))$  tend to be observations with large exposure times. This is expected, as cell death is sometimes followed by the dissolution of cell nuclei, hindering the measurement of cellular response. Next, we provide plots of the probability integral transform histogram for the entire model, as described by Gneiting *et al.* (2007). The plot is provided in the *bottom* panel of Figure 5, and visual assessment indicates that it is close to uniformity, suggesting relatively good posterior predictive calibration. In this case study, a subsequent analysis performed excluding potentially influential data-points revealed no changes in our inferences. We would however like to stress the importance of final model diagnostics any times one reaches for highly structured models for data analysis. Additional summaries and formal definition of these diagnostic tools are detailed in Appendix B.

## 5. DISCUSSION

In this manuscript we propose a modeling framework for general exposure escalation experiments, based on a newly defined measure of toxicity. Precisely, we consider the representation and estimation of the probability that each particle or compound would initiate adverse biological reactions. This measure is seamlessly integrated into very general exposure escalation protocols, including multi-dimensional models, that account for dose and duration kinetics jointly. Furthermore, the proposed summary is warranted with respect to non-lethal biological outcomes, as well as for experiments in nano-toxicology, where particle aggregation behavior may hinder reasonable expectations of sigmoidal dose response trajectories.

One dimension that is potentially lost in the model, if implemented in terms of a general exposure-response surface, is the possible interpretation of standard dose-response parameters for canonical risk assessment. However, if parametric assumptions are warranted by the experimental setting, our model is easily represented in terms of these standard parametric formulations. Furthermore, most quantities related to standard risk assessment parameters can still be easily obtained numerically, by post-processing posterior samples.

We show how the probability of toxicity can be used to link nanomaterials physicochemical properties to non-linear and multidimensional cytotoxicity profiles, while accounting for the uncertainty in the estimation of this summary. This is in stark contrast with the current practice of step-wise analyses. In our case study we chose to carry out the analysis through the use of standard diffuse prior setting on the latent regression model, which in turn translates into over dispersed Beta priors on the pre-experimental probabilities of toxicity. It

is however important to stress how the proposed formulation allows for the explicit inclusion of pre-experimental evidence through prior calibration related to particle properties that are known to be potentially hazardous. This last exercise may be crucial in risk assessment settings, where large amounts of evidence already exist in relation to specific classes of nanomaterials. The proposed methodology is also appropriate for limited data sets, as it includes data integration and a framework for advanced dimension reduction through variable selection. We account for the non-robust nature of the data by allowing for particle specific variance inflation, resulting in a t-distributed model for the error structure. Finally, the hierarchical representation of the model allows for easy and robust estimation, as well as rich inference and informative model diagnostics.

Our proposal is easily extended to account for multiple cytotoxicity parameters, in the form of multivariate dependent observations. A reasonable dependence scheme can, for example, assume data to be dependent within outcome and particle, as well as between outcomes for the same particle. Further generalizations may relax the assumption of linearity relating particle properties to probabilities of toxicity. This exercise would however, find justification only if larger sets of nanomaterial libraries become available for biological testing.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Primary support was provided by the U.S. Public Health Service Grant U19 ES019528 (UCLA Center for Nanobiology and Predictive Toxicology). This work was also supported by the National Science Foundation and the Environmental Protection Agency under Cooperative Agreement Number DBI-0830117. Any opinions, findings, conclusions or recommendations expressed herein are those of the author(s) and do not necessarily reflect the views of the National Science Foundation or the Environmental Protection Agency. This work has not been subjected to an EPA peer and policy review.

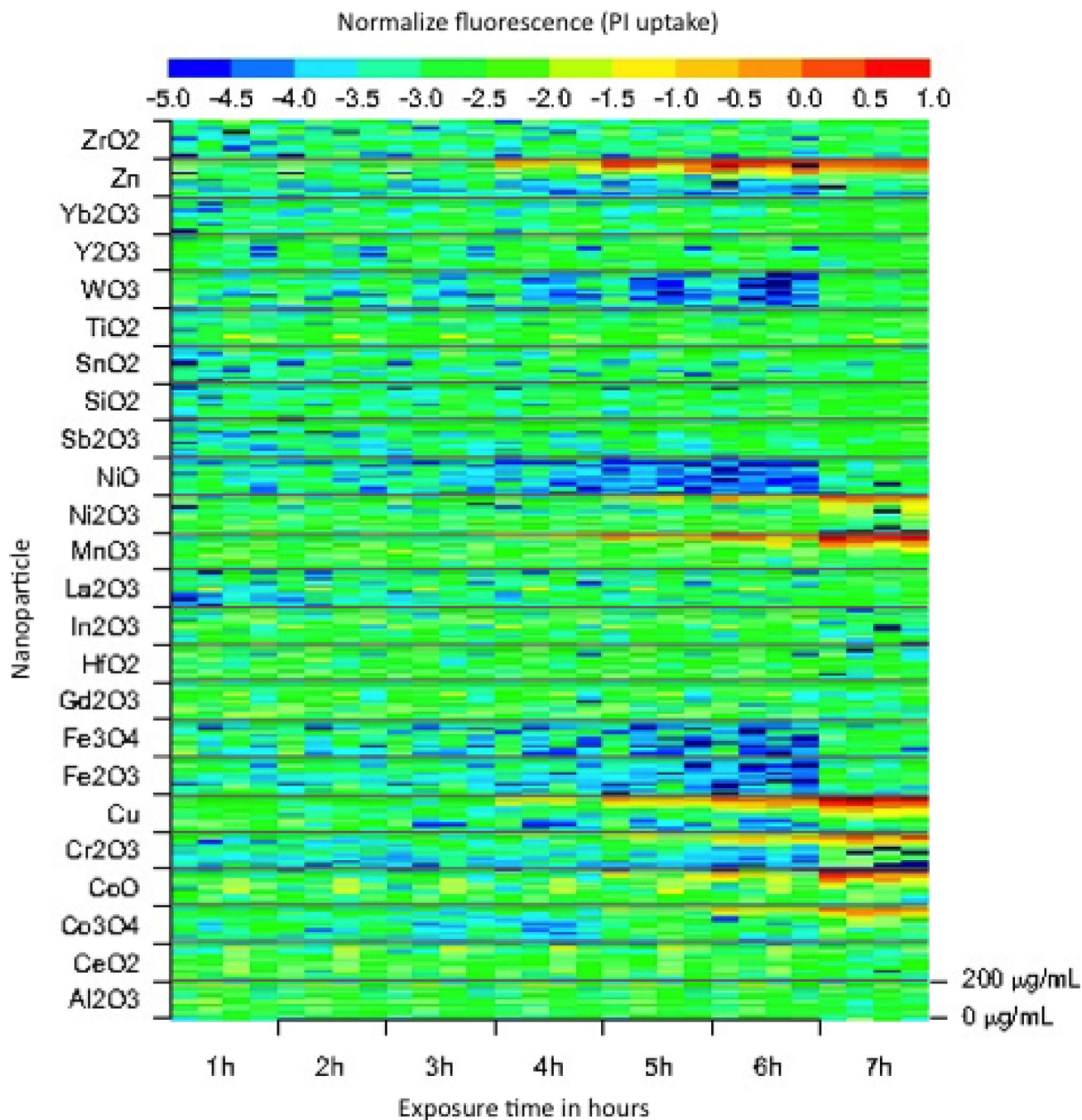
## REFERENCES

- Albert J, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of American Statistical Association*. 1993; 88(422):669–679.
- Berger, JO. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag; 1985.
- Besag J, Higdon D. Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*. 1999; 61(4):691–746.
- Besag J, Kooperberg C. On conditional and intrinsic autoregression. *Biometrika*. 1995; 82:733–746.
- Boos DD, Brownie C. Mixture models for continuous data in dose-response studies when some animals are unaffected by treatment. *Biometrics*. 1991; 47(4):1489–1504. [PubMed: 1786327]
- Burello E, Worth P. Qsar modeling of nanomaterials. *WIREs Nanomedicine and Nanobiotechnology*. 2011; 3:298–306. [PubMed: 21384562]
- Carlin, BP.; Louis, TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Boca Raton, FL: Chapman and Hall; 2000.
- Clayton, D. Markov Chain Monte Carlo in Practice, eds. London: Chapman and Hall; 1996. p. 275-301.
- Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. *Statistics in Medicine*. 2010; 29(9):1037–1057. [PubMed: 20087875]

- Eilers EHC, Marx BD. Flexible smoothing with  $b$ -splines and penalties. *Statistical Science*. 1996; 11(2):89–102.
- Fernandez C, Ley E, Steel M. Benchmark priors for bayesian model averaging. *Journal of Econometrics*. 2001; 100:381–427.
- Geisser S. Discussion on sampling and bayes inference in scientific modeling and robustness. *Journal of the Royal Statistical Society: Series A*. 1980; 143:416–417.
- Gelfand A, Smith A. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 1990; 85:398–409.
- Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1984; 6:721–741. [PubMed: 22499653]
- George E, McCulloch R. Variable selection via gibbbs sampling. *Journal of the American Statistical Association*. 1993; 88:881–889.
- George S, Xia T, Rallo R, Zhao Y, Ji Z, Lin S, Wang X, Zhang H, France B, Schoenfeld D, Damoiseaux R, Liu R, Lin S, Bradley K, Cohen Y, Nel A. Use of a high-throughput screening approach coupled with in vivo zebrafish embryo screening to develop hazard ranking for engineered nanomaterials. *ACS Nano*. 2011; 5(3):1805–1817. [PubMed: 21323332]
- Gneiting T, Balabdaoui F, Raftery A. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2007; 69(2):243–268.
- Green P. Reversible jump markov chain monte carlo computation and Bayesian model determination. *Biometrika*. 1995; 82(4):711–732.
- Hinderliter PM, Minard KR, Orr G, Chrisler WB, Thrall BD, Pounds JG, Teeguarden JG. A computational model of particle sedimentation, diffusion and target cell dosimetry for in vitro toxicity studies. *Particle and Fibre Toxicology*. 2010; 7
- Jordan MI, Jacobs RA. Hierarchical mixtures of experts and the em algorithm. *Neural Computation*. 1994; 6:181–214.
- Lang S, Brezger A. Bayesian p-splines. *Journal of Computational and Graphical Statistics*. 2004; 13(1):183–212.
- Lang S, Brezger A. Generalized structured additive regression based on Bayesian p-splines. *Computational Statistics and Data Analysis*. 2006; 50:967–991.
- Lee K, Sha N, Dougherty E, M V, Mallick B. Gene selection: a Bayesian variable selection approach. *Bioinformatics*. 2003; 19(1):90–97. [PubMed: 12499298]
- Liu R, Rallo R, George S, Ji Z, Nair S, Nel A, Cohen Y. Classification nanosar development for cytotoxicity of metal oxide nanoparticles. *Small*. 2011; 7(8):1118–1126. [PubMed: 21456088]
- McLachlan, G.; Peel, D. *Finite Mixture Models*. New York: Wiley Series in Probability and Statistics; 2000.
- Meng H, Liong M, Xia T, Li Z, Ji Z, Zink J, E NA. Engineered design of mesoporous silica nanoparticles to deliver doxorubicin and p-glycoprotein sirna to overcome drug resistance in a cancer cell line. *ACS nano*. 2010; 4(8)
- Nel A, Xia T, Mädler L, Li N. Toxic potential of materials at the nanolevel. *Science*. 2006; 311(5761): 622–627. [PubMed: 16456071]
- Patel T, Telesca D, George S, Nel A. Toxicity profiling of engineered nanomaterials via multivariate dose response surface modeling. *COBRA Preprint Series*. 2011 Article 88.
- Piegorsch WW, An L, Wickens AA, Webster West R, Peña EA, Wu W. Information-theoretic model-averaged benchmark dose analysis in environmental risk assessment. *Environmetrics*. 2013 n/a–n/a.
- Plummer M, Best N, Cowles K, Vines K. CODA: Convergence diagnosis and output analysis for MCMC. *R News*. 2006; 6(1):7–11.
- Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari T, Michalkova A, Hwang H, Toropov A, Leszczynska D, Leszczynski J. Using nano-qsar to predict the cytotoxicity of metal oxide nanoparticles. *Nature Nanotechnology*. 2011; 6:175–178.
- Ramsay JO. Monotone Regression Splines in Action. *Statistical Science*. 1988; 3:425–441.

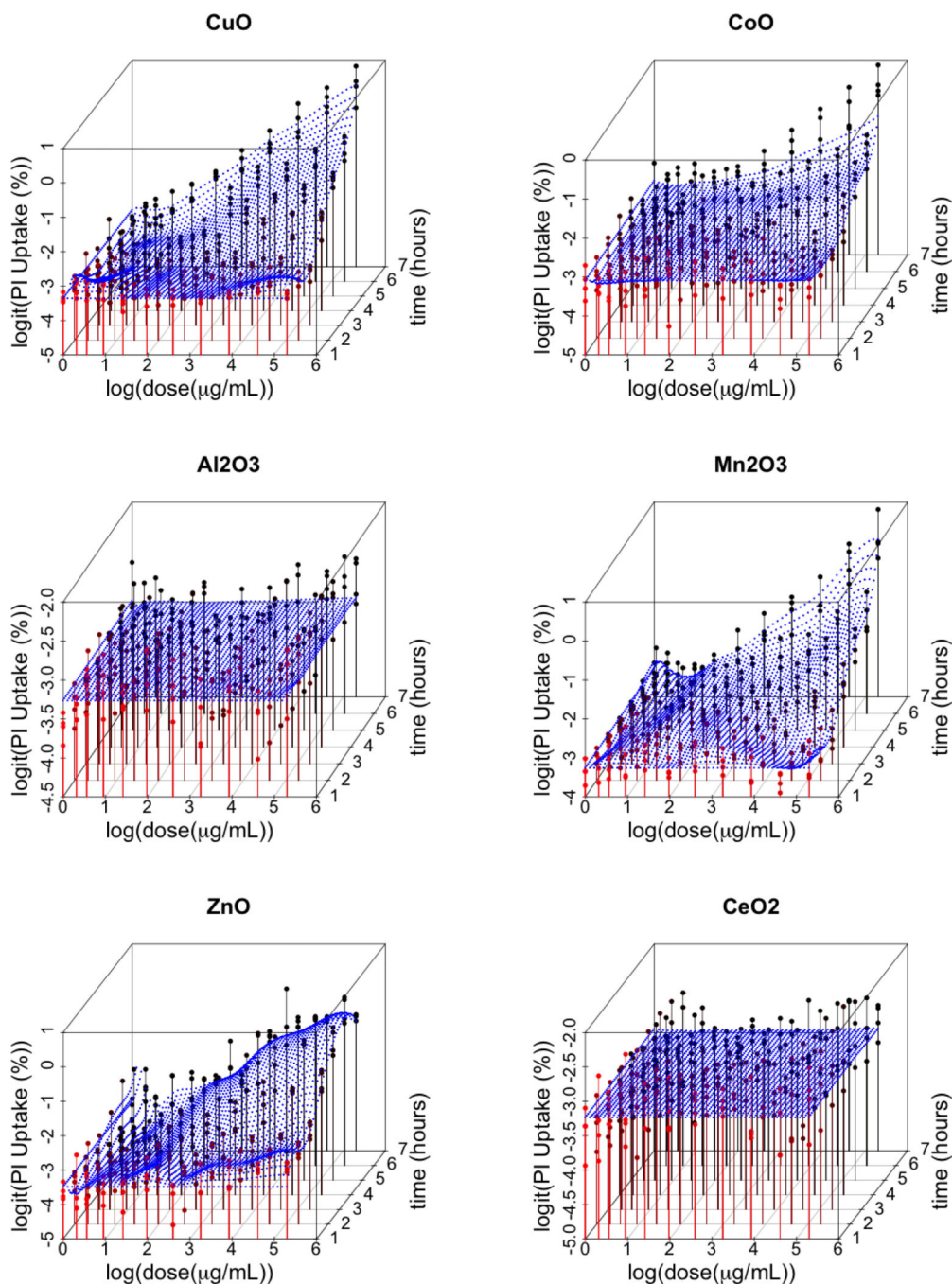
- Ritz C. Toward a unified approach to dose-response modeling in ecotoxicology. *Environmental Toxicology and Chemistry*. 2010; 29(1):220–229. [PubMed: 20821438]
- Schultz W, Cronin M, Netzevab T. The present status of qsar in toxicology. *Journal of Molecular Structure: THEOCHEM*. 2003; 622(1–2):23–38.
- Scott J, Berger J. An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*. 2006; 136(7):2144–2162.
- Shaw S, Westly E, Pittet M, Subramanian A, Schreiber S, Weissleder R. Perturbational profiling of nanomaterial biologic activity. *Proceedings of the National Academy of Sciences*. 2008; 105(21):7387–7392.
- Stern S, McNeil S. Nanotechnology safely concerns revisited. *Toxicological Sciences*. 2008; 101(1):4–21. [PubMed: 17602205]
- Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*. 1996; 58(1):267–288.
- Wang X, Shel J. A class of grouped Brunk estimators and penalized spline estimators for monotone regression. *Biometrika*. 2010; 97(3):585–601.
- West M. Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Association. Series B (Methodological)*. 1984; 46(3):431–439.
- Xia T, Kovochich M, Liong M, Mädler L, Gilbert B, Shi HB, Yeh JI, Zink JI, Nel AE. Comparison of the mechanism of toxicity of zinc oxide and cerium oxide nanoparticles based on dissolution and oxidative stress properties. *ACS Nano*. 2008; 2(10):2121–2134. [PubMed: 19206459]
- Xia T, Zhao Y, Sager T, George S, Pokhrel S, Li N, Schoenfeld D, Meng H, Lin S, Wang X, Wang M, Ji Z, Zink JI, Mädler L, Castranova V, Lin S, Nel AE. Decreased dissolution of ZnO by iron doping yields nanoparticles with reduced toxicity in the rodent lung and zebrafish embryos. *ACS Nano*. 2011; 5(2):1223–1235. [PubMed: 21250651]
- Zellner, A. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*. North-Holland: Elsevier; 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distribution regression using Bayesian variable selection; p. 233-243.
- Zhang H, Ji Z, Xia T, Meng H, Low-Kam C, Rong L, Suman P, Lin S, Wang X, Yu-Pei L, Wang M, Linjiang L, Rallo R, Damoiseaux R, Telesca D, Madler L, Cohen Y, Zink J, Nel A. Use of metal oxide nanoparticle band gap to develop a predictive paradigm for oxidative stress and acute pulmonary inflammation. *ACS Nano*. 2012; 6(5):4349–4368. [PubMed: 22502734]
- Zhang H, Xia T, Meng H, Xue M, George S, Ji Z, Wang X, Liu R, Wang M, France B, Rallo R, Damoiseaux R, Cohen Y, Bradley K, Zink JI, Nel AE. Differential expression of syndecan-1 mediates cationic nanoparticle toxicity in undifferentiated versus differentiated normal human bronchial epithelial cells. *ACS Nano*. 2011; 5(4):2756–2769. [PubMed: 21366263]





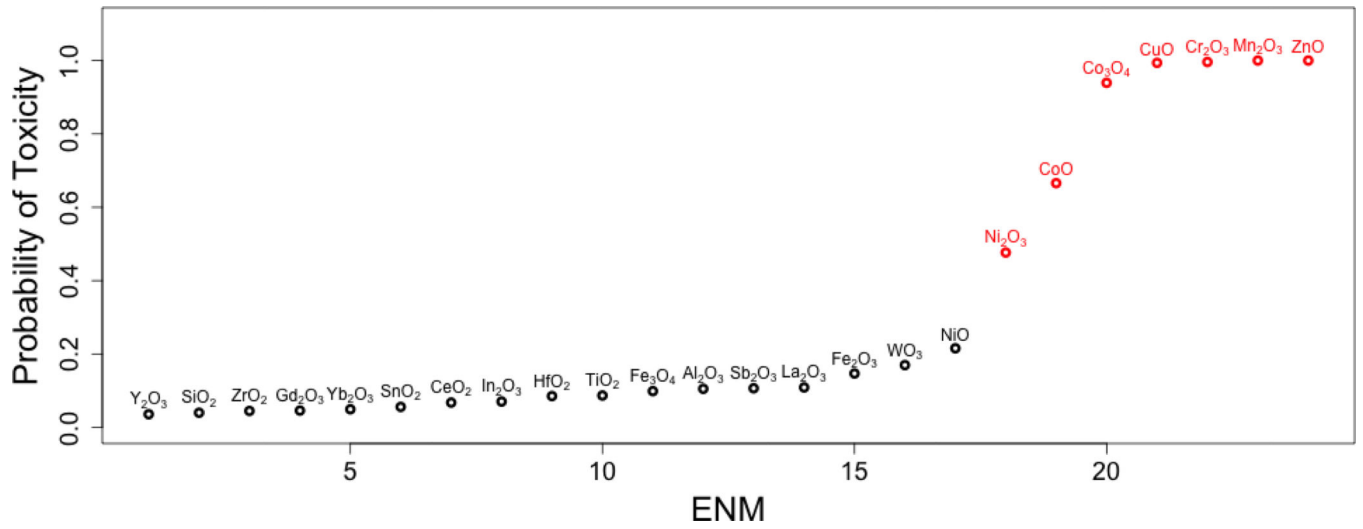
**Figure 1. PI uptake of BEAS-2B**

Normalized PI uptake associated with exposing BEAS-2B cells to increased doses of nanomaterials (from 0µg/ml to 200µg/ml). This image summarizes data for an exposure escalation experiment involving a library of 24 nano metal oxides. The exposure escalation protocol includes 11 dose levels and 7 hours of exposure. Each dose by duration combination is replicated 4 times.

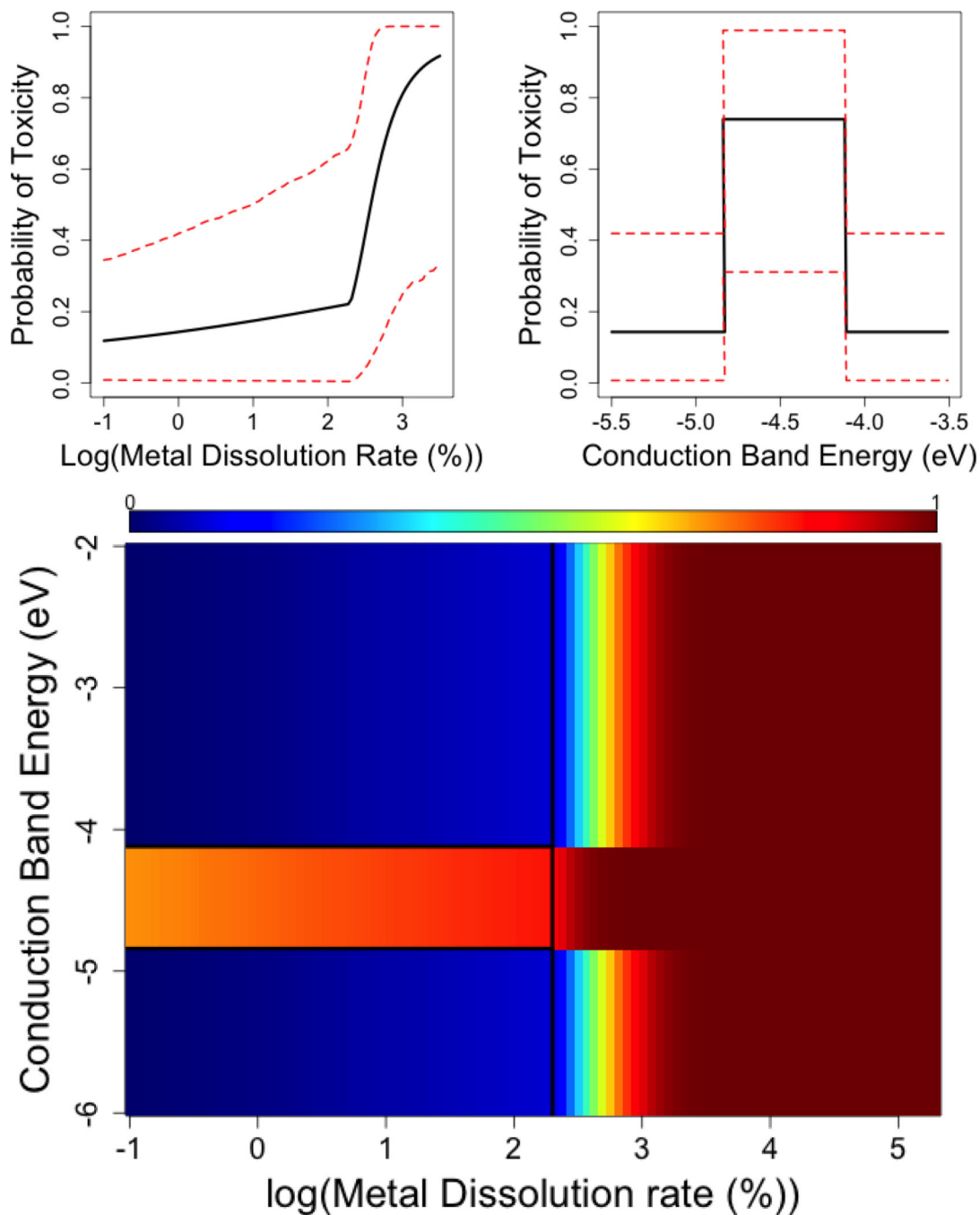


**Figure 2. Fitted response surfaces for CuO, CoO, Al<sub>2</sub>O<sub>3</sub>, Mn<sub>2</sub>O<sub>3</sub>, ZnO, CeO<sub>2</sub> ENMs**

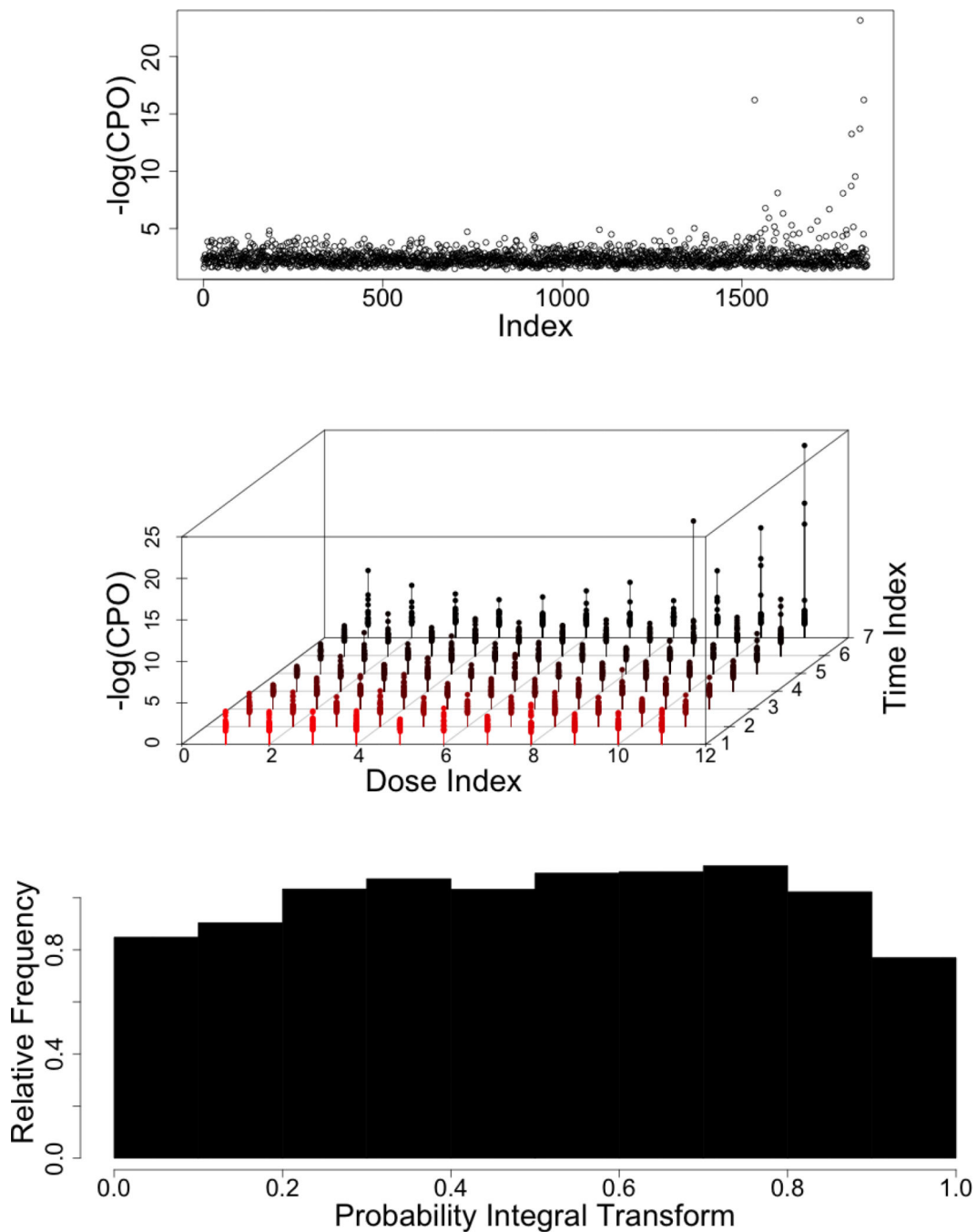
Estimated smooth response surfaces of cellular membrane damage, as measured by Propidium Iodine (PI) uptake, as a function of log(dose) and time. The color red represents response values corresponding to lower time points and the color black represents response values corresponding to higher time points.



**Figure 3. Posterior mean probability of toxicity for each ENM**  
 Particles with a probability of toxicity greater than .5 are indicated in red.



**Figure 4. Posterior summaries of the probability of toxicity as a function of conduction band energy and log metal dissolution rate**  
*(Top)* Posterior mean (black) and 95% posterior intervals (red) for the probability of toxicity as a function of log(metal dissolution rate), given conduction band energy outside the range of the cell redox potential. *(Middle)* Posterior mean (black) and 95% posterior intervals (red) for the probability of toxicity as a function of conduction band energy, given no metal dissolution. *(Bottom)* Posterior mean probability as a function of conduction band energy and log(metal dissolution rate). Red colored regions indicate greater probability of toxicity, whereas blue colored regions indicate low probability of toxicity.



**Figure 5. Graphical model diagnostics**

(Top) Estimate of  $-\log(\text{cpo}_i(d, t))$  for detecting observations with poor model fit. (Middle) Plot of  $-\log(\text{cpo}_i(d, t))$  as a function of dose and time, indicating any relationship between outlying observations and the administered dose or duration of exposure. (Bottom) Probability Integral Transform assessing empirical calibration of the posterior predictive distribution.

**Table 1**  
**Posterior summaries for regression coefficients corresponding to the model which includes predictors for seven different ENM physicochemical properties**

Posterior inclusion probabilities as well as posterior mean and associated 95% posterior intervals are provided for the model regression coefficients.

Parameter	Inclusion Probability	Posterior Summaries
Intercept	1.00	-0.50(-16.17, 18.31)
Conduction Band Energy	0.77	1.62(0.00, 3.81)
log(Metal Dissolution Rate)<2.3	0.69	0.07(-0.19, 0.39)
log(Metal Dissolution Rate)>2.3	0.69	2.81(-0.19, 12.90)
log(Primary Size)	0.31	0.00(-0.72, 0.71)
log(Crystal Structure (b(Å)))	0.35	-0.03(-2.11, 1.97)
log(Enthalpy of Formation ( $H_{Me^{n+}}$ ))	0.38	-0.20(-2.38, 1.37)
log(Lattice Energy ( $H_{lattice}$ ))	0.36	-0.18(-1.89, 2.10)

**Table 2**  
**Posterior summaries for regression coefficients corresponding to the final model**

The final model includes model includes predictors for to conduction band energy and log(metal dissolution rate). Posterior mean estimates and associated 95% posterior intervals are provided for the model regression coefficients.

<b>Parameter</b>	<b>Posterior Summaries</b>
Intercept	-1.22(-2.45, -0.20)
Conduction Band Energy	2.02(0.49, 3.75)
log(Metal Dissolution Rate)<2.3	0.11(-0.19, 0.41)
log(Metal Dissolution Rate)>2.3	3.52(-0.18, 8.49)