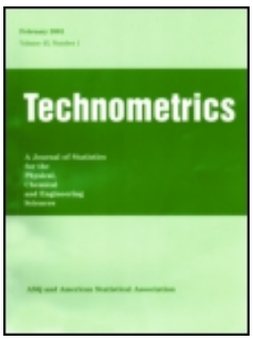

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: https://amstat.tandfonline.com/loi/utch20

Bayesian Modeling for Physical Processes in Industrial Hygiene Using Misaligned Workplace Data

João V. D. Monteiro, Sudipto Banerjee & Gurumurthy Ramachandran

To cite this article: João V. D. Monteiro, Sudipto Banerjee & Gurumurthy Ramachandran (2014) Bayesian Modeling for Physical Processes in Industrial Hygiene Using Misaligned Workplace Data, Technometrics, 56:2, 238-247, DOI: 10.1080/00401706.2013.836988

To link to this article: https://doi.org/10.1080/00401706.2013.836988

View supplementary material

Accepted author version posted online: 06 Sep 2013.
Published online: 16 May 2014.

Submit your article to this journal

Article views: 193

View Crossmark data

Citing articles: 2 View citing articles

Full Terms & Conditions of access and use can be found at
https://amstat.tandfonline.com/action/journalInformation?journalCode=utch20

Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

Bayesian Modeling for Physical Processes in Industrial Hygiene Using Misaligned Workplace Data

João V. D. Monteiro

Department of Statistical Sciences
Duke University
Durham, NC 27708
(jm397@stat.duke.edu)

Sudipto Banerjee

School of Public Health
Division of Biostatistics
University of Minnesota
Minneapolis, MN 55455
(baner009@umn.edu)

Gurumurthy Ramachandran

School of Public Health
Division of Environmental and Occupational Health
University of Minnesota
Minneapolis, MN 55455
(ramac002@umn.edu)

In industrial hygiene, a worker's exposure to chemical, physical, and biological agents is increasingly being modeled using deterministic physical models that study exposures near and farther away from a contaminant source. However, predicting exposure in the workplace is challenging and simply regressing on a physical model may prove ineffective due to biases and extraneous variability. A further complication is that data from the workplace are usually *misaligned*. This means that not all timepoints measure concentrations near and far from the source. We recognize these challenges and outline a flexible Bayesian hierarchical framework to synthesize the physical model with the field data. We reckon that the physical model, by itself, is inadequate for enhanced inferential and predictive performance and deploy (multivariate) Gaussian processes to capture uncertainties and associations. We propose rich covariance structures for multiple outcomes using latent stochastic processes. This article has supplementary material available online.

KEY WORDS: Bayesian melding; Cross-covariances; Gaussian processes; Linear ordinary differential equations; Markov chain Monte Carlo; Occupational exposure models

1. INTRODUCTION

A key concern of industrial hygiene is the estimation of a worker's exposure to chemical, physical, and biological agents. One goal of exposure modeling is to represent the physical processes generating chemical concentrations in the workplace. Physical models in industrial hygiene include a *source* (for contaminant generation) and allow for the transport and fate of the contaminant over time to predict concentrations (Nicas 1996; Nicas and Jayjock 2002). A common setting assumes different physical behavior for the concentrations—one close to the contaminant source ("near field") and another farther away ("far field"). The resulting physical model is known as a two-zone model and is described by ordinary differential equations (ODEs) that model the rate of change in concentrations using some physical parameters (Ramachandran 2005). These parameters ("inputs"), in theory, determine when the system attains steady state. A purely conceptual approach would assign "plausible" values to these inputs, usually assuming steady-state concentrations (e.g., Keil et al. 2009). An accurate representation will deliver better concentration estimates and facilitate subsequent exposure management. This, however, is challenging because the workplace is notoriously complex and deterministic physical models are unlikely to provide an adequate representation. It is, therefore, becoming increasingly clear that a synergy of physical and statistical models is needed to better estimate the processes in the workplace.

Inference is improved by using observations from the workplace. Concentration is typically measured over a finite set of timepoints. The two-zone setting produces bivariate concentration measurements—one from the "near field" and another from the "far field." Typically, however, there are discrepancies between the observations and the deterministic physical model since the physical model assumptions are violated in real workplace environments. A plausible choice for inputs to the two-zone model could, perhaps, be obtained by training them using trial-and-error until satisfactory agreement between the output and concentration measurements is achieved. That approach, however, is unattractive. Not only can finding satisfactory agreement between the observations and the physical model's output be difficult, even if they agree the approach fails to account for the uncertainty in estimation and prediction. Model assessment

© 2014 American Statistical Association and the American Society for Quality
TECHNOMETRICS, MAY 2014, VOL. 56, NO. 2
DOI: 10.1080/00401706.2013.836988

would be completely ad hoc as well. A more principled approach estimates the physical model's unknown inputs from the concentration measurements by making use of prior information on the input parameters. Usually, some prior information regarding the inputs to the physical model is available based upon physical considerations implied by the model or from experts with experience in workplace environments. A Bayesian modeling framework that allows synthesis of information from different sources is, therefore, attractive.

Synthesizing deterministic physical models with statistical models to achieve improved inference continues to garner attention. One approach, Bayesian melding (e.g., Raftery, Givens, and Zeh 1995; Poole and Raftery 2000; Fuentes and Raftery 2005; Ševčíková, Raftery, and Waddell 2007, 2011), achieves such synthesis by incorporating prior information on the inputs to the physical model, estimates them using their posterior distributions, and carries out subsequent predictive inference. In its simplest form, Bayesian melding proceeds from a hierarchical model that regresses on the physical model. See, for example, Zhang et al. (2009) and Raftery and Bao (2010) for two very different applications of this approach. We demonstrate, however, that straightforward Bayesian nonlinear regression can be highly ineffective in predicting exposure concentrations in industrial workplaces.

Stochastic processes are deployed to reckon with variability not accounted for by the physical model. Fuentes and Raftery (2005) and Berrocal, Gelfand, and Holland (2011) recently applied spatial processes to meld information from monitoring sites with output from numerical models. They focused largely upon spatial interpolation and predictions using independent runs of the numerical model. Assessing uncertainty in model inputs was precluded by the complexity of the physical models therein. In different applications, Ševčíková, Raftery, and Waddell (2007, 2011) proposed Bayesian melding with single-outcome land use and transportation models. There is also related literature in the domain of "computer models" (e.g., Kennedy and O'Hagan 2001; Santner, Williams, and Notz 2003; Bayarri et al. 2007). Here, the computer model is usually highly complex and computationally onerous to evaluate. Hence, a Gaussian process is used as a stochastic emulator or interpolator to approximate model outputs. We, on the other hand, are able to work with the exact mathematical model, which is cheap to evaluate, and focus upon flexibly modeling the discrepancies between the mathematical model and the physical output.

Apart from addressing a new domain of application and the theoretical implications therein, we specifically focus upon two pertinent statistical issues. First, we deal with associated bivariate outcomes that are not only related by the physical model, but are also likely to produce correlated residuals. We recognize that even when the physical model is easily tractable, either analytically or computationally, it is unable to account for extraneous variability in the workplace. This occurs almost invariably in industrial hygiene experiments—the physical model provides information on the overall trend but is too inflexible to capture variation at smaller scales, thereby impairing predictive performance. Second, the data from industrial workplaces are, more often than not, *misaligned*. This means that concentration measurements from two different outcomes may not always have been observed at the same timepoint. More precisely, we can imagine three sets of timepoints—one has observations from both the near and far fields, another has measurements from the near field only, and the third includes measurements from the far field only. Alternative terminology might refer to this setting as one of "missing data." In our context, such missingness is assumed *completely at random*. Interest focuses upon estimating and predicting the joint distribution of the concentrations in the two fields at any arbitrary timepoint.

Statistical modeling for temporal processes can proceed either by treating time as "discrete" or as "continuous" depending upon whether inference (e.g., prediction or interpolation) is sought at the same temporal resolution (e.g., "minutes" and "hours") at which the concentrations have been observed or whether it is sought at arbitrary resolutions. Here, we treat the concentrations as smooth functions of time and offer inference at arbitrary temporal resolutions. In this regard, our approach is arguably richer and especially attractive for handling temporal misalignment. Our key modeling ingredient is a multivariate Gaussian process. Apart from modeling the usual residual variability, our framework achieves the following analytical objectives: (i) approximate the trend (or bias) missed by the physical model for concentrations in both fields, (ii) capture correlations across time (with process realizations acting as time-varying random effects), and (iii) model the correlations among the outcomes when we have multiple outcomes. These objectives resemble those in the "calibration" of multi-output computer models, where Gaussian processes emulators for the physical model are used to estimate the inputs (see, e.g., Conti and O'Hagan 2010). Our contribution here is to offer a richer class of multivariate Gaussian processes to model the discrepancies between the physical model and the observed data.

The remainder of the article evolves as follows. In Section 2, we briefly describe the two-zone occupational exposure model. Section 3 introduces the misaligned experimental data we subsequently analyze. Section 4 presents our process-based Bayesian melding (PBBM) approach. In Section 5, we describe a fairly elaborate simulation study followed by an analysis of our workplace data. Finally, Section 6 concludes the article by presenting some discussion with an eye toward future work.

2. TWO-ZONE MODEL

The two-zone (or two-component) model assumes the presence of a contamination source in the workplace and that the region is composed of two well-mixed zones or fields. The zone very near and around the source is called the *near field*, while the rest of the room is called the *far field*. The far field completely encloses the near field and there is some amount of air exchange between them. Following convention, we assume that two distinct places that are in the same field have equal exposure concentration levels. Also, we assume that the contaminant's total mass is emitted at a constant rate G and that there is an airflow rate between the near field and far field equal to β. The final assumption considers that there are supply and exhaust flow rates that are taken to be the same and equal to Q. Figure 1 is a schematic depiction of the dynamics of the system, where V_N and V_F denote the volumes at the near and far field, respectively.

In this context, the hygienist models the exposure concentrations at the near and far fields based upon observations collected

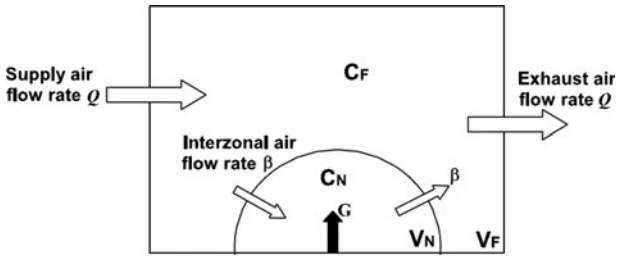

Figure 1. Dynamics of the two-zone model.

over a period of time. Figure 1, along with the assumptions, yields the following two-component model:

$$\frac{d}{dt}\mathbf{c}(\boldsymbol{\theta}_1;\mathbf{x},t) = \mathbf{W}(\boldsymbol{\theta}_1;\mathbf{x})\,\mathbf{c}(\boldsymbol{\theta}_1;\mathbf{x},t) + \mathbf{g}(\boldsymbol{\theta}_1;\mathbf{x}), \tag{1}$$

where

$$\mathbf{c}(\boldsymbol{\theta}_1;\mathbf{x},t) = \begin{bmatrix} c_N(\boldsymbol{\theta}_1;\mathbf{x},t) \\ c_F(\boldsymbol{\theta}_1;\mathbf{x},t) \end{bmatrix},$$

$$\mathbf{W}(\boldsymbol{\theta}_1;\mathbf{x}) = \begin{bmatrix} -\frac{\beta}{V_N} & \frac{\beta}{V_N} \\ \frac{\beta}{V_F} & -\frac{(\beta+Q)}{V_F} \end{bmatrix}, \mathbf{g}(\boldsymbol{\theta}_1;\mathbf{x}) = \begin{bmatrix} \frac{G}{V_N} \\ 0 \end{bmatrix},$$

$$\boldsymbol{\theta}_1 = \{\beta, Q, G\}, \text{ and } \mathbf{x} = \{V_N, V_F\}.$$

The functions $c_N(\boldsymbol{\theta}_1;\mathbf{x},t)$ and $c_F(\boldsymbol{\theta}_1;\mathbf{x},t)$ are the exposure concentrations at time t in the near and far fields, respectively.

The solution of (1) depends upon the eigenvalues of $\mathbf{W}(\boldsymbol{\theta}_1;\mathbf{x})$. When the eigenvalues are real and distinct, we obtain the following solution for (1):

$$\begin{aligned}\mathbf{c}(\boldsymbol{\theta}_1;\mathbf{x},t) = {} & \exp(t\mathbf{W}(\boldsymbol{\theta}_1;\mathbf{x}))\,\mathbf{c}(\boldsymbol{\theta}_1;\mathbf{x},0) + \mathbf{W}^{-1}(\boldsymbol{\theta}_1;\mathbf{x}) \\ & \times [\exp(t\mathbf{W}(\boldsymbol{\theta}_1;\mathbf{x})) - \mathbf{I}_2]\mathbf{g}(\boldsymbol{\theta}_1;\mathbf{x}),\end{aligned} \tag{2}$$

where $\exp(t\mathbf{W}(\boldsymbol{\theta}_1;\mathbf{x}))$ is the matrix exponential (see supplementary material "ODE.pdf"). Assuming that $c_N(\boldsymbol{\theta}_1;\mathbf{x},0) = c_F(\boldsymbol{\theta}_1;\mathbf{x},0) = 0$, Equation (2) can be simplified to produce the following unique solution:

$$\begin{aligned} & c_N(\boldsymbol{\theta}_1;\mathbf{x},t) \\ & = \frac{G}{Q} + \frac{G}{\beta} + G\left(\frac{\beta Q + \lambda_2 V_N(\beta+Q)}{\beta Q V_N(\lambda_1-\lambda_2)}\right)e^{\lambda_1 t} \\ & \quad - G\left(\frac{\beta Q + \lambda_1 V_N(\beta+Q)}{\beta Q V_N(\lambda_1-\lambda_2)}\right)e^{\lambda_2 t}, \\ & c_F(\boldsymbol{\theta}_1;\mathbf{x},t) \\ & = \frac{G}{Q} + G\left(\frac{\lambda_1 V_N + \beta}{\beta}\right)\left(\frac{\beta Q + \lambda_2 V_N(\beta+Q)}{\beta Q V_N(\lambda_1-\lambda_2)}\right)e^{\lambda_1 t} \\ & \quad - G\left(\frac{\lambda_2 V_N + \beta}{\beta}\right)\left(\frac{\beta Q + \lambda_1 V_N(\beta+Q)}{\beta Q V_N(\lambda_1-\lambda_2)}\right)e^{\lambda_2 t}, \end{aligned} \tag{3}$$

where λ_1 and λ_2 are the eigenvalues of $\mathbf{W}(\boldsymbol{\theta}_1;\mathbf{x})$. These are available in closed form and are real and distinct whenever β, Q, V_F, and V_N are all strictly positive (see supplementary material "ODE.pdf"). The latter two quantities are volumes of a chamber and are, hence, positive. Physical considerations ensure that the same is true for β and Q. Assigning priors with positive support ensures a stable system with real solutions.

The exponential terms in (3) decay asymptotically to zero at large values of t. Consequently, the steady-state solutions for the near and far fields are $G/Q + G/\beta$ and G/Q, respectively. Therefore, the model predicts a greater exposure intensity near the emission source compared to the one-compartment model in steady-state conditions. Moreover, when β is less than or equal to Q, the steady-state concentration in the far field is less than twice the steady-state concentration in the near field. In general, Q increases relative to β as the room size increases. Thus, the model draws a distinction between exposures of workers near the source and those farther away from the source.

3. EXPERIMENTAL TWO-ZONE DATA

The experimental two-zone data that we analyze here is a part of a database compiled from a series of designed experiments that were conducted in the industrial hygiene laboratories at the University of Minnesota. The data consist of exposure concentrations of toluene over a period of time, where Q and G were known to be 13.8 m^3/min and 351.5 mg/min, respectively. Measurements at 10 cm and 15 cm from the contamination source represent the exposure concentrations in the near and far fields, respectively. The near field is defined to be a 10 cm high cylinder with a radius of 10 cm around the generation source. Consequently, $V_N = \pi \times 10^{-3}$ m^3. The zone beyond the near field is the far field, which has $V_F = 3.8$ m^3.

A salient feature of this data, and what is not atypical in industrial hygiene, is that measurements at several timepoints are available in only one of the fields, but not simultaneously from both. Given limited resources and other logistics pertaining to setting up the experiment, observations are initially available only from the near field. As the experiment proceeds, we obtain measurements from both the fields. Since taking simultaneous measurements from both fields may be logistically difficult, toward the end of the experiment only the far field is measured to make up for the initial loss of information there.

Figure 2 presents this two-zone experimental dataset. It is divided into three "time zones." In zone I, only the near field provides measurements (83 timepoints), zone II measures both fields (134 timepoints) and, finally, zone III measures only the far field (160 timepoints) to produce $83 + 160 + 2 \times 134 = 511$ measurements. Such data are often referred to as *temporally misaligned* because of the "gaps" at the beginning and the end. It is not necessary that the data be misaligned in a pattern such as in Figure 2. Depending upon the experimental scenario, the missingness can occur completely at random and the resulting observations are not required to be divided into such temporal zones. Our approach applies to any type of temporal misalignment. To avoid complications arising from temporal misalignment, Zhang et al. (2009) analyzed two-zone data using straightforward nonlinear regression restricting themselves to timepoints that yielded measurements in both fields. This incurs loss of information because the partial measurements from either field still carry information on the physical parameters.

A brief exploratory analysis of the data reveals why relying upon the physical model alone for inference and scientific deductions is undesirable. Under the assumption of zero initial concentration, the theoretical implication of the two-zone model is that the concentration in the far field attains steady state after about $351.5/13.8 \approx 25$ mg/m^3. Even a cursory glance at Figure 2 shows that the data suggest a much higher steady-state concentration—perhaps around 250 mg/m^3. Least-square

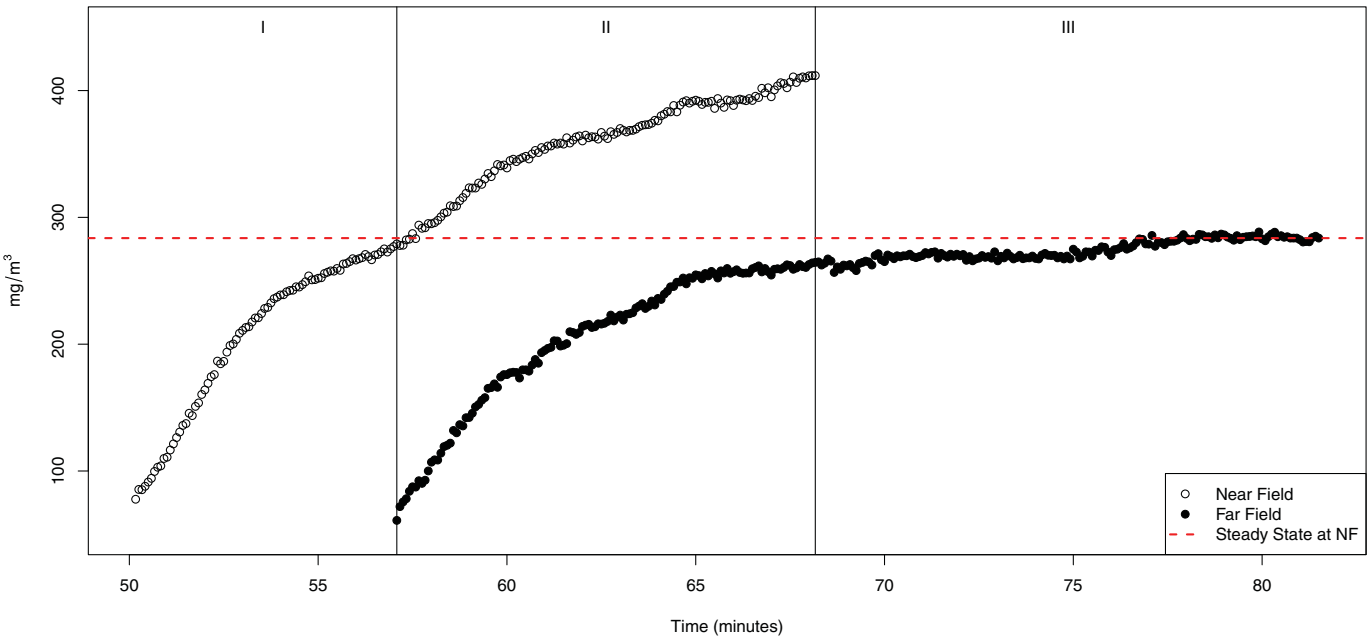

Figure 2. Two-zone experimental data.

analysis and other methods that purely rely upon regressing on the physical model (e.g., Zhang et al. 2009) are also likely to produce grossly biased inference and poor predictions. For instance, the horizontal dashed line in Figure 2 shows the steady-state solution at the near field using a least-square estimate for β. What is required is a stochastic mechanism to capture extraneous variation in the data. Instead of parametric specifications, a richer and more flexible option is to introduce an *unknown* function of time that can be estimated at arbitrary timepoints. Within a Bayesian setting, one needs a prior on this random function, which is provided by a stochastic process over time. We elucidate in subsequent sections.

4. BAYESIAN HIERARCHICAL MODELING

We elucidate our approach using a generic setup that considers the following distinct modeling ingredients: (a) an $m \times 1$ vector of measurements $\mathbf{y}(t) = [y_1(t), \ldots, y_m(t)]^{\mathrm{T}}$ taken at timepoint t, (b) inputs (parameters), denoted as $\boldsymbol{\theta}_1$, in the physical model that are unknown, and (c) variables $\mathbf{x}$ that act as experimental controls and are known. For instance, in the two-zone model $\boldsymbol{\theta}_1 = \{\beta, Q, G\}$, $\mathbf{x} = \{V_N, V_F\}$, and $m = 2$.

Following recent research (see, e.g., Kennedy and O'Hagan 2001; Bayarri et al. 2007), it is beneficial to represent the physical model as a *biased* representation of "reality" resulting in discrepancies between the data and the deterministic physical model. We attribute this discrepancy to two terms: "model bias" and "measurement error." The former results from violations of physical assumptions in the actual workplace, while the latter arises from instrumentation error and other unexplained randomness. A stochastic process to capture this bias is likely to yield better model fit and estimation of underlying variability. It is difficult to pin down the "physical source," or "sources," behind these discrepancies. We do not, therefore, attempt to attach a physical meaning to this process. We model $\mathbf{y}(t)$ as the sum of three components: (a) a systemic component represented by the physical model; (b) a stochastic process to complement the physical model; and (c) a stochastic measurement error process. We write

$$\begin{aligned} y_i(t) &= f_i(\boldsymbol{\theta}_1; \mathbf{x}, t) + \eta_i(t) + \epsilon_i(t), \ i = 1, 2, \ldots, m \\ &\Downarrow \\ \mathbf{y}(t) &= \mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, t) + \boldsymbol{\eta}(t) + \boldsymbol{\epsilon}(t) \ , \end{aligned} \tag{4}$$

where $f_i(\boldsymbol{\theta}_1; \mathbf{x}, t)$ is the physical model, possibly transformed to a scale commensurate with $y_i(t)$, $\eta_i(t)$'s are stochastic processes that capture the bias (extraneous variability), and $\epsilon_i(t)$'s are white-noise processes capturing variation attributable to measurement error and other sources of micro-scale discrepancies. Equation (4) also shows the corresponding vector representation, where $\mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, t)$, $\boldsymbol{\eta}(t)$, and $\boldsymbol{\epsilon}(t)$ are $m \times 1$ vectors whose ith elements are given by $f_i(\boldsymbol{\theta}_1; \mathbf{x}, t)$, $\eta_i(t)$, and $\epsilon_i(t)$, respectively. Our approach does not require $\mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, t)$ be available in closed form, but only that it be computable for any choice of inputs $\{\boldsymbol{\theta}_1, \mathbf{x}, t\}$.

4.1 Multivariate Process Models

The critical element in (4) is $\boldsymbol{\eta}(t)$. We could assume an autoregressive structure (or Kalman filter) treating $\boldsymbol{\eta}(t)$ as discrete, but that precludes estimation of concentrations smoothly at *arbitrary* timepoints and is inconvenient for handling temporal misalignment. An arguably richer option is to treat $\boldsymbol{\eta}(t)$ as an unknown continuous function of time with a Gaussian process prior (see, e.g., Conti and O'Hagan 2010). Assume that we observe $\mathbf{y}(t)$ at n different timepoints $\mathbf{t} = \{t_1, \ldots, t_n\}$ and that $\boldsymbol{\eta}(t) \sim \mathrm{GP}_m\left(\mathbf{0}_m, \mathbf{C}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \cdot, \cdot)\right)$ denotes a zero-centered $m \times 1$ multivariate Gaussian process with $m \times m$ *cross-covariance* matrix $\mathbf{C}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; t, t')$, whose (i, j)th element is the covariance between $\eta_i(t)$ and $\eta_j(t')$ for $i, j = \{1, \ldots, m\}$, and $\boldsymbol{\theta}_2$ is a collection of unknown parameters therein. The Gaussian process implies that $\boldsymbol{\eta} = [\boldsymbol{\eta}^{\mathrm{T}}(t_1), \ldots, \boldsymbol{\eta}^{\mathrm{T}}(t_n)]^{\mathrm{T}}$ is distributed as an $mn \times 1$ multivariate normal distribution: $\boldsymbol{\eta} \mid \Sigma_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \mathbf{t}) \sim$

$\mathrm{N}_{mn}(\mathbf{0}_{mn}, \Sigma_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \mathbf{t}))$, where $\Sigma_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \mathbf{t})$ is the $mn \times mn$ block matrix whose (k, l)th block is $\mathbf{C}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; t_k, t_l)$.

Clearly, care is needed when choosing $\mathbf{C}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \cdot, \cdot)$ so that $\boldsymbol{\Sigma}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \mathbf{t})$ is symmetric and positive definite. To ensure this in a flexible and computationally feasible manner, we adopt a constructive approach that assumes that $\boldsymbol{\eta}(t)$ arises as a linear transformation of a latent $p \times 1$ multivariate process whose components are independent of each other, where $1 \leq p \leq m$. This idea is adopted in the so-called "linear model of coregionalization" in spatial statistics (e.g., Banerjee, Carlin, and Gelfand 2004) but has not, to the best of our knowledge, been used in Bayesian melding applications. To be precise, we assume $\boldsymbol{\eta}(t) = \mathbf{A}\mathbf{w}(t)$, where $\mathbf{A}$ is an $m \times p$ matrix with unknown entries, and $\mathbf{w}(t) \sim \mathrm{GP}_p(\mathbf{0}_p, \mathbf{C}_{\mathbf{w}}(\cdot, \cdot; \boldsymbol{\varphi}))$, where $\mathbf{w}(t) = [w_1(t), \ldots, w_p(t)]^{\mathrm{T}}$ is the $p \times 1$ multivariate latent process and $\boldsymbol{\varphi}$ is a collection of unknown parameters therein.

Now, assume $\mathbf{w}(t)$ has unit variance, that is, $\mathrm{var}\{\mathbf{w}(t)\} = \mathbf{I}_p$. Accordingly, $\mathbf{C}_{\mathbf{w}}(\boldsymbol{\varphi}; t, t') = \mathrm{cov}\{\mathbf{w}(t), \mathbf{w}(t') \,|\, \boldsymbol{\varphi}\}$ is a diagonal matrix with $\rho_i(\varphi_i; t, t')$ as the ith diagonal element, where $\rho_i(\varphi_i; t, t')$ is the correlation between $w_i(t)$ and $w_i(t')$ for all $i = \{1, \ldots, p\}$, and $\boldsymbol{\varphi} = \{\varphi_1, \ldots, \varphi_p\}$. Regardless of how close t and t' are, we assume that there is no correlation between $w_i(t)$ and $w_j(t')$, when $i \neq j$. However, recall that the random process that impacts $\mathbf{y}(t)$ is actually $\boldsymbol{\eta}(t)$, which has the cross-covariance matrix $\mathbf{C}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; t, t') = \mathrm{cov}\{\boldsymbol{\eta}(t), \boldsymbol{\eta}(t') \,|\, \boldsymbol{\theta}_2\} = \mathbf{A}\mathbf{C}_{\mathbf{w}}(\boldsymbol{\varphi}; t, t')\mathbf{A}^{\mathrm{T}}$, where $\boldsymbol{\theta}_2 = \{\mathbf{A}, \boldsymbol{\varphi}\}$. If $p < m$, then $\boldsymbol{\eta}(t)$ is a degenerate Gaussian process (i.e., the covariance matrix for any finite realization is singular) but (4) is still a legitimate model for $\mathbf{y}(t)$ because of the white noise $\boldsymbol{\epsilon}(t)$.

If $\mathbf{A}$ is square (i.e., $p = m$) and nondiagonal, then $\mathbf{C}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; t, t')$ will not be diagonal, which means in this case the model considers correlation between $\eta_i(t)$ and $\eta_j(t')$, even for $i \neq j$. Furthermore, when $t = t'$ we have $\mathbf{C}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; t, t) = \mathbf{A}\mathbf{A}^{\mathrm{T}}$, which means that we can, without loss of generality, set $\mathbf{A}$ to be lower triangular to identify with the Cholesky factor of $\mathbf{C}_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; t, t)$. This, however, is not strictly required and any square root (e.g., from a spectral or singular value decomposition) will yield a valid cross-covariance matrix for $\boldsymbol{\eta}(t)$. In fact, we obtain $\boldsymbol{\eta} \,|\, \Sigma_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \mathbf{t}) \sim \mathrm{N}_{mn}(\mathbf{0}_{mn}, \Sigma_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \mathbf{t}))$, where $\Sigma_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \mathbf{t}) = (\mathbf{I}_n \otimes \mathbf{A})\Sigma_{\mathbf{w}}(\boldsymbol{\varphi}; \mathbf{t})(\mathbf{I}_n \otimes \mathbf{A}^{\mathrm{T}})$ is guaranteed to be symmetric and positive definite as long as $\mathbf{A}$ is nonsingular. Here, $\Sigma_{\mathbf{w}}(\boldsymbol{\varphi}; \mathbf{t})$ denotes the covariance matrix of $\mathbf{w}$, and $\otimes$ represents the Kronecker product.

It remains, then, to choose $\rho_1(\varphi_1; \cdot, \cdot), \ldots, \rho_p(\varphi_p; \cdot, \cdot)$. These will control the smoothness of the underlying process. Had the process been an emulator for the physical model, as is often the case for complex computer models (e.g., Bayarri et al. 2007), we would require the process to be smoother. A common choice is the Gaussian correlation function, $\rho_i(\varphi_i; t, t') = e^{-\varphi_i |t - t'|^2}$. We, however, use the process to model time-varying random effects representing unaccounted structured extraneous variation in the data. Excessive smoothness will lead to poorer fits and is not desirable. For flexibly modeling smoothness as well as strength of association, we opt for the Matérn correlation function

$$\rho(\phi, \nu; t, t') = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{|t - t'|}{\phi}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{|t - t'|}{\phi}\right); \quad \phi > 0, \nu > 0, \tag{5}$$

where $\Gamma(\cdot)$ is the usual gamma function while $\mathcal{K}_{\nu}$ is a modified Bessel function of the second kind. The process is $\lceil \nu - 1 \rceil$ times mean square differentiable, while ϕ determines how quickly the correlation decays over time. In particular, the correlation decays more slowly as ϕ increases (Stein 1999). We assume that $\rho_i(\boldsymbol{\varphi}_i; \cdot, \cdot)$'s are Matérn functions with distinct parameters. Specifically, let $\boldsymbol{\varphi}_i = \{\phi_i, \nu_i\}$ be the Matérn parameters in $\rho_i(\boldsymbol{\varphi}_i; \cdot, \cdot)$, $i = 1, \ldots, p$. Consequently, $\boldsymbol{\theta}_2 = \{\mathbf{A}, \phi_1, \ldots, \phi_p, \nu_1, \ldots, \nu_p\}$. Several simpler choices emerge as special cases, most notably the exponential $\rho(\phi; t, t') = e^{-\phi|t-t'|}$, which results by fixing $\nu = 1/2$. It is worth pointing out that (5), in its full generality, hardly makes a substantive difference in our setting and it is likely that a much simpler distance-based correlation function will suffice for most practical purposes. Nevertheless, we have used (5) for our subsequent analysis as it does not substantially increase the computational burden in our current application.

Turning to the measurement error process at any timepoint t, we assume $\boldsymbol{\epsilon}(t) \,|\, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3) \overset{\text{iid}}{\sim} \mathrm{N}_m(\mathbf{0}_m, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3))$. Typically $\boldsymbol{\theta}_3$ is the collection of the $m(m+1)/2$ distinct entries in $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3)$, where $\mathrm{var}\{\epsilon_j(t)\} = \tau_j$ denotes the jth diagonal entry and $\mathrm{cov}\{\epsilon_i(t), \epsilon_j(t)\} = \tau_{ij}$ is the (i, j)th entry, for $i, j = 1, \ldots, m$ and $i < j$.

For n timepoints in $\mathbf{t}$, the $\mathbf{y}(t_i)$'s are conditionally independent and normally distributed, that is, $\mathbf{y}(t_i) \,|\, \boldsymbol{\theta}, \mathbf{w}(t_i), \mathbf{x}, t_i \overset{\text{ind}}{\sim} \mathrm{N}_m(\mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, t) + \mathbf{A}\mathbf{w}(t_i), \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3))$, where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3\}$ (recall $\boldsymbol{\theta}_1 = \{\beta, G, Q\}$). The joint posterior distribution, $p(\boldsymbol{\theta}, \mathbf{w} \,|\, \mathbf{y}, \mathbf{x}, \mathbf{t}, \boldsymbol{\gamma})$, is proportional to

$$\left[\prod_{i=1}^{n} \mathrm{N}_m(\mathbf{y}(t_i) \,|\, \mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, t) + \mathbf{A}\mathbf{w}(t_i), \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3))\right] \times N_{pn}(\mathbf{w} \,|\, \mathbf{0}_{pn}, \boldsymbol{\Sigma}_{\mathbf{w}}(\boldsymbol{\varphi}; \mathbf{t})) \times p(\boldsymbol{\theta} \,|\, \boldsymbol{\gamma}), \tag{6}$$

where $N_k(\mathbf{u} \,|\, \boldsymbol{\mu}, \mathbf{S})$ denotes a k-dimensional multivariate normal density function for $\mathbf{u}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{S}$, and $\boldsymbol{\gamma}$ is the set of all hyperparameters. We suppose that the $\boldsymbol{\theta}_i$'s have independent prior distributions, that is, $p(\boldsymbol{\theta} \,|\, \boldsymbol{\gamma}) = \prod_{i=1}^{3} p(\boldsymbol{\theta}_i \,|\, \boldsymbol{\gamma}_i)$, where $\boldsymbol{\gamma}_i$ is the set of hyperparameters related to the prior distribution of $\boldsymbol{\theta}_i$. Supplementary material "Identifiability.pdf" outlines identifiability conditions for the process parameters in (6).

Estimation of (6) proceeds from a Gibbs sampler with *random-walk* Metropolis steps (Gelman et al. 2003). We implement Markov chain Monte Carlo (MCMC) after integrating out $\mathbf{w}$ from the model to shrink the parameter space and achieve faster convergence for $\boldsymbol{\theta}$. Posterior samples of $\mathbf{w}$ can be obtained subsequently: for each posterior sample of $\boldsymbol{\theta}$, we draw a $\mathbf{w}$ from $\mathbf{w} \,|\, \boldsymbol{\theta}, \mathbf{y} \sim \mathrm{N}_{pn}(\mathbf{m}^*_{\mathbf{w}}, \boldsymbol{\Sigma}^*_{\mathbf{w}})$ with $\mathbf{m}^*_{\mathbf{w}} = \boldsymbol{\Sigma}^*_{\mathbf{w}}[\mathbf{I}_n \otimes (\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3))(\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, \mathbf{t}))]$ and $\boldsymbol{\Sigma}^*_{\mathbf{w}} = [\boldsymbol{\Sigma}^{-1}_{\mathbf{w}}(\boldsymbol{\varphi}, \mathbf{t}) + \mathbf{I}_n \otimes (\mathbf{A}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3)\mathbf{A})]^{-1}$.

4.2 Model Assessment

We will subsequently use the deviance information criterion (DIC) and a modified predictive model choice criteria called the gneiting–raftery scoring rule (GRS) as model comparison metrics. Let $\boldsymbol{\Omega}$ be the collection of parameters. The DIC (Spiegelhalter et al. 2002) is the sum of the posterior expected deviance $\overline{D} = \mathrm{E}_{\boldsymbol{\Omega} \,|\, \mathbf{y}}\left[-2 \log p(\text{data} \,|\, \boldsymbol{\Omega})\right]$ and the effective number of

parameters $p_D = \overline{D} - D(\overline{\boldsymbol{\Omega}})$, where $\overline{\boldsymbol{\Omega}}$ denotes the posterior expectation of $\boldsymbol{\Omega}$. Models with smaller DICs are preferred. Here, we take $\boldsymbol{\Omega}$ as the collection of $\boldsymbol{\mu}_i = \mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, t_i)$'s and $\boldsymbol{\eta}(t_i)$'s, for $i = 1, 2 \ldots, n$, and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3)$. These parameters constitute the "focus" of the DIC.

Gneiting and Raftery (2007) presented a scoring rule (GRS) based upon the predictive distribution of independently *replicated data*. To be precise, let $y_j^{\text{rep}}(t_i)$ denote the replicate for $y_j(t_i)$, $\mathbf{y}^{\text{rep}}(t_i)$ be the $m \times 1$ vector with $y_j^{\text{rep}}(t_i)$ as its jth element and $\mathbf{y}^{\text{rep}}$ be the $mn \times 1$ vector obtained by stacking up the $\mathbf{y}^{\text{rep}}(t_i)$'s. The posterior predictive distribution for $\mathbf{y}^{\text{rep}}$ is

$$p(\mathbf{y}^{\text{rep}} \mid \mathbf{y}) = \int p(\mathbf{y}^{\text{rep}} \mid \boldsymbol{\theta}, \mathbf{w}, \mathbf{y}) p(\boldsymbol{\theta}, \mathbf{w} \mid \mathbf{y}) d\boldsymbol{\theta} d\mathbf{w} = \int p(\mathbf{y}^{\text{rep}} \mid \boldsymbol{\theta}, \mathbf{w}) p(\mathbf{w} \mid \boldsymbol{\theta}, \mathbf{y}) p(\boldsymbol{\theta} \mid \mathbf{y}) d\boldsymbol{\theta} d\mathbf{w} \,. \quad (7)$$

Each $\mathbf{y}^{\text{rep}}(t_i)$ can be regarded as the "model-predicted" value for the observed $\mathbf{y}(t_i)$; see Gelman et al. (2003). To draw samples from (7), we first sample $\boldsymbol{\theta}$'s from $p(\boldsymbol{\theta} \mid \mathbf{y})$ (using random-walk Metropolis steps). For each sampled $\boldsymbol{\theta}$, we draw $\mathbf{w}$ from $p(\mathbf{w} \mid \boldsymbol{\theta}, \mathbf{y})$, one-for-one given the simulated value of $\boldsymbol{\theta}$. Subsequently, we sample $\mathbf{y}^{\text{rep}}$ from $p(\mathbf{y}^{\text{rep}} \mid \boldsymbol{\theta}, \boldsymbol{\eta})$, one-for-one for each simulated value of $\boldsymbol{\theta}$ and $\mathbf{w}$. The GRS is defined as

$$\text{GRS} = -\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{y_j(t_i) - \mu_{ij}^{\text{rep}}}{\sigma_{ij}^{\text{rep}}}\right)^2 - \sum_{i=1}^{n}\sum_{j=1}^{m}\log\left\{\left(\sigma_{ij}^{\text{rep}}\right)^2\right\},$$

where μ_{ij}^{rep} and σ_{ij}^{rep} are the mean and standard deviation, respectively, of $y_j^{\text{rep}}(t_i)$ in (7). The GRS is easily evaluated from posterior samples. The GRS depends only on the first and second moments of the predictive distribution for the replicated data, and penalizes departure of replicated means from the corresponding observed values (lack of fit), as well as the uncertainty in the replicated data (often reflected by overparameterization). Models having higher GRS are preferred.

4.3 Misaligned Data

Section 4 considers the ideal situation where we observe all the outcomes for every t_i. In practice, however, it is not uncommon to encounter *misaligned* or multivariate missing data in two-zone experimental settings. The means that measurements on some of the outcomes are missing at some timepoints as is the case with us (Section 3).

An advantage of our process-based framework is that inference with misaligned data can be accommodated with some minor tweaks. We elucidate with the two-zone model ($m = 2$) setting, where $\mathbf{A}$ is 2×2, $\mathbf{w}(t) = (w_1(t), w_2(t))^{\text{T}}$, and $w_1(t)$ and $w_2(t)$ are independent Gaussian processes. It helps to distinguish among three sets of timepoints. Let $\mathbf{t}_1$ be the set of timepoints that yield observations only in the near field, $\mathbf{t}_2$ be the timepoints that yield observations only in the far field, and $\mathbf{t}_{12}$ be the timepoints yielding simultaneous measurements from both the fields. The observed data likelihood is now

$$\prod_{t\in\mathbf{t}_{12}} \text{N}_2\left(\mathbf{y}(t) \mid \mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, t) + \mathbf{A}\mathbf{w}(t), \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3)\right) \times \prod_{t\in\mathbf{t}_1} \text{N}_1\left(y_1(t) \mid f_1(\boldsymbol{\theta}_1; \mathbf{x}, t) + \mathbf{a}_{1*}^{\text{T}}\mathbf{w}(t), \tau_1\right) \times \prod_{t\in\mathbf{t}_2} \text{N}_1\left(y_2(t) \mid f_2(\boldsymbol{\theta}_1; \mathbf{x}, t) + \mathbf{a}_{2*}^{\text{T}}\mathbf{w}(t), \tau_2\right), \quad (8)$$

where $\mathbf{a}_{l*}^{\text{T}}$ denotes the lth row vector of $\mathbf{A}$, for $l = 1, 2$. The joint posterior distribution can then be obtained by multiplying (8) with the priors as in (6).

For a more generic setup, some further details on implementation may be useful. Let $\mathbf{y}$ be the $nm \times 1$ vector obtained by stacking up the $\mathbf{y}(t_i)$'s. Suppose that k of its elements are observed and consequently $nm - k$ are missing. Denote by $\mathbf{y}_{\text{o}}$ and $\mathbf{y}_{\text{m}}$ the observed and missing data, respectively. We can write $\mathbf{y}_{\text{o}}$ and $\mathbf{y}_{\text{m}}$ by suitably extracting elements from $\mathbf{y}$. Therefore, there are extraction matrices, $\mathbf{P}_{\text{o}}$ and $\mathbf{P}_{\text{m}}$, such that $\mathbf{y}_{\text{o}} = \mathbf{P}_{\text{o}}\mathbf{y}$ and $\mathbf{y}_{\text{m}} = \mathbf{P}_{\text{m}}\mathbf{y}$. The matrix $\mathbf{P}_{\text{o}}$ is $k \times nm$ and $\mathbf{P}_{\text{m}}$ is $(mn - k) \times mn$. Both these matrices are short and wide and have full row rank.

Bayesian inference evaluates the full posterior predictive distribution $\mathbf{y}_{\text{m}}$,

$$p(\mathbf{y}_{\text{m}} \mid \mathbf{y}_{\text{o}}) = \int p(\mathbf{y}_{\text{m}} \mid \boldsymbol{\theta}, \mathbf{y}_{\text{o}}) p(\boldsymbol{\theta} \mid \mathbf{y}_{\text{o}}) d\boldsymbol{\theta} \,. \quad (9)$$

Obtaining samples from (9) is straightforward and can be performed *after* the posterior samples of $\boldsymbol{\theta}$ have been drawn from $p(\boldsymbol{\theta} \mid \mathbf{y}_{\text{o}})$: for each sampled $\boldsymbol{\theta}$, we draw $\mathbf{y}_{\text{m}}$ from $p(\mathbf{y}_{\text{m}} \mid \boldsymbol{\theta}, \mathbf{y}_{\text{o}})$. Matters are simplified because $\mathbf{y}_{\text{m}} \mid \boldsymbol{\theta}, \mathbf{y}_{\text{o}} \sim \text{N}_{nm-k}(\mathbf{m}(\boldsymbol{\theta}; \mathbf{t}), \mathbf{V}(\boldsymbol{\theta}; \mathbf{t}))$, where

$$\begin{aligned}\mathbf{m}(\boldsymbol{\theta}; \mathbf{t}) &= \mathbf{P}_{\text{m}}\mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, \mathbf{t}) + \mathbf{P}_{\text{m}}\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \\ &\quad \times \mathbf{P}_{\text{o}}^{\text{T}}\left(\mathbf{P}_{\text{o}}\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3)\mathbf{P}_{\text{o}}^{\text{T}}\right)^{-1}(\mathbf{y}_{\text{o}} - \mathbf{P}_{\text{o}}\mathbf{f}(\boldsymbol{\theta}_1; \mathbf{x}, \mathbf{t})); \\ \mathbf{V}(\boldsymbol{\theta}; \mathbf{t}) &= \mathbf{P}_{\text{m}}\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3)\mathbf{P}_{\text{m}}^{\text{T}} - \mathbf{P}_{\text{m}}\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3) \\ &\quad \times \mathbf{P}_{\text{o}}^{\text{T}}\left(\mathbf{P}_{\text{o}}\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3)\mathbf{P}_{\text{o}}^{\text{T}}\right)^{-1}\mathbf{P}_{\text{o}}\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3)\mathbf{P}_{\text{m}}^{\text{T}}.\end{aligned}$$

Here, $\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3) = \Sigma_{\boldsymbol{\eta}}(\boldsymbol{\theta}_2; \mathbf{t}) + \mathbf{I}_n \otimes \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3)$ is the $nm \times nm$ covariance matrix for $\mathbf{y}$ given $\boldsymbol{\theta}$. Crucially, the inverses in $\mathbf{m}(\boldsymbol{\theta}; \mathbf{t})$ and $\mathbf{V}(\boldsymbol{\theta}; \mathbf{t})$ are well defined because $\mathbf{P}_{\text{o}}$ and $\mathbf{P}_{\text{m}}$ have full row rank and $\boldsymbol{\Sigma}_{\mathbf{y}}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_3)$ is nonsingular.

5. DATA ANALYSIS

We now apply our PBBM approach to datasets simulated from two-zone experiments as well as the experimental data described in Section 3. The simulation studies we perform here demonstrate the identifiability of model parameters and the flexibility and effectiveness of the PBBM approach under diverse structural specifications. In particular, we compare the performance of different association structures using three distinct specifications for $\mathbf{A}$: vector (V), diagonal (D), and lower triangular (LT). See Table 1. Moreover, we compare the PBBM

Table 1. Matrix structures for $\mathbf{A}$

(a) V	(b) D	(c) LT
$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$	$\begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix}$	$\begin{bmatrix} a_1 & 0 \\ a_3 & a_2 \end{bmatrix}$

Table 2. Parameter values used to simulate the synthetic two-zone datasets

		Parameters												
		$\boldsymbol{\theta}_1$			$\boldsymbol{\theta}_2$							$\boldsymbol{\theta}_3$		
Model	**A**	β	Q	G	ϕ_1	ϕ_2	ν_1	ν_2	a_1	a_2	a_3	τ_1	τ_2	τ_{12}
PBBM	V	7.25	15	105	8	–	2.5	-	0.032	0.141	–	0.0005	0.0100	0.0020
	D	7.25	15	105	15	8	0.5	2.5	0.032	0.141	–	0.0005	0.0100	0.0020
	LT	7.25	15	105	15	8	0.5	2.5	0.032	0.062	0.127	0.0005	0.0100	0.0020
BNLR	–	7.25	15	105	–	–	–	–	–	–	–	0.0010	0.0200	0.0040

with the simpler Bayesian nonlinear regression model (BNLR), which is essentially the PBBM without the random process (i.e., $\boldsymbol{\eta}(t) = \mathbf{0}$).

Specifications for $\mathbf{A}$ depend upon the dimension of $\mathbf{w}(t)$ and the parameters in $\mathbf{C}_{\mathbf{w}}(\boldsymbol{\varphi}, \cdot, \cdot)$. For V, $\mathbf{w}(t)$ is a univariate Gaussian process ($p = 1$) and, therefore, $\boldsymbol{\varphi} = \{\phi_1, \nu_1\}$. For D and LT, $\mathbf{w}(t)$ is a bivariate Gaussian process ($p = 2$) and $\boldsymbol{\varphi} = \{\phi_1, \phi_2, \nu_1, \nu_2\}$. Finally, $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}(\boldsymbol{\theta}_3) = \left[\begin{smallmatrix} \tau_1 & \tau_{12} \\ \tau_{12} & \tau_2 \end{smallmatrix}\right]$ accommodating *correlated* measurement errors.

5.1 Simulation Study

We first compare the performance of the models using synthetic two-zone datasets that were generated according to the PBBM and BNLR frameworks. Specifically, we simulate 100 independent datasets from the BNLR model and from each of the three PBBM specifications in Table 1. Each dataset is composed of exposure concentrations (log-scale) in the two fields observed at 100 equally spaced timepoints between 1 and 100 min. The model parameters used to generate the data are presented in Table 2.

As seen in Table 2, the parameters associated with the marginal variance of $\mathbf{y}(t)$ (i.e., a_i's and τ_i's) were chosen to be relatively small, which is typical in actual experimental scenarios. This also offers the BNLR a fairer platform to perform effectively because larger values for the a_i's can produce more variable concentration curves that would be more congruous to models with random effects. The input parameters for the two-zone model (i.e., $\boldsymbol{\theta}_1 = \{\beta, Q, G\}$ and $\mathbf{x} = \{V_N, V_F\}$) were taken from physical considerations deemed plausible by industrial hygienists (e.g., Ramachandran 2005). These values simulate a workplace where the volume of the near field is equal to half of the volume of a sphere with radius 0.8 m, that is, $V_N = 1.1$ m^3. Moreover, we assume $V_F = 240$ m^3 and zero initial concentrations in both fields. In this scenario, the theoretical steady-state concentration at the near field ($G/Q + G/\beta \approx 21.5$ mg/m^3) is roughly three times higher than that at the far field ($G/Q = 7$ mg/m^3). (Information regarding the prior settings are available in the supplementary material "PriorSettings.pdf.")

We divide each simulated dataset into a training set and a test set. The training set consists of exposure concentrations in both fields at 70 timepoints randomly selected between 1 and 100 min. The testing set is composed of the exposure concentrations at the remaining timepoints. For each model, inference was based on 5000 posterior samples obtained from our MCMC algorithm after discarding the first 5000 iterations as burn-in. For random-walk Metropolis steps, we transformed parameters, if necessary, to have support on the real line so that normal proposals could be used and then transformed them back to the original scale. For the Gaussian process covariance functions, the substantive inference from the Matérn and the exponential were essentially indistinguishable. Subsequently, we present only the results for the Matérn.

Table 3 presents the DIC and the GRS, averaged over the 100 independently generated datasets, for the PBBMs and BNLR. Here, the row labels represent the model generating the data (i.e., the "true" model), while the column labels represent the model used to fit the datasets. The numbers in the parenthesis are the standard errors.

Table 3 reveals that, in general, both comparison metrics suggest that the "true" model (i.e., the one from which the data was generated) seems to excel. A noticeable exception occurs when the data are generated from the BNLR. In this case, the DIC score suggests that all models fit the data equally well, while the GRS shows that PBBM with the LT structure outperforms the others. In fact, both the DIC and GRS metrics

Table 3. DIC and GRS metrics for the simulation study assuming no temporal misalignment. The standard errors, from the 100 simulations, are shown in parenthesis

GOF	Model	BNLR	D	LT	V
DIC	BNLR	−470.35 (14.68)	−470.38 (14.89)	−470.31 (15.11)	−470.43 (14.82)
	D	−356.83 (41.36)	−510.78 (18.44)	−507.11 (18.88)	−478.19 (25.45)
	LT	−459.62 (33.95)	−522.40 (16.39)	−535.27 (16.60)	−522.71 (18.20)
	V	−513.16 (29.60)	−546.63 (20.22)	−560.89 (17.23)	−560.09 (17.68)
GRS	BNLR	619.81 (22.10)	621.98 (22.53)	627.36 (22.84)	623.34 (22.65)
	D	606.75 (35.52)	723.08 (24.23)	717.94 (28.15)	661.50 (38.50)
	LT	597.21 (39.79)	657.23 (25.98)	703.07 (34.45)	673.43 (37.04)
	V	626.01 (48.19)	698.64 (34.52)	735.82 (27.51)	732.71 (29.44)

Table 4. Multivariate potential scale reduction factor (Brooks and Gelman 1998)

BNLR	D	LT
1.07	1.01	1.01

indicate that PBBM with LT structure is always very competitive and usually excels irrespective of the underlying generating mechanism. The standard errors indicated in parenthesis show that there seem to be no significant differences between the DICs of the true model and the PBBM with structure LT. On the other hand, the PBBM models seem to have significantly lower DIC and GRS scores than the BNLR when the data are generated from the PBBM models. In summary, PBBM performs better than the BNLR, except when the data come from the BNLR model, in which case they show similar performance. Some additional analyses are available in the supplementary material "AdditionalAnalysis.pdf."

5.2 Analysis of Misaligned Experimental Data

We now analyze the misaligned experimental data described in Section 3. We consider the BNLR and also the PBBM approach with structures LT and D. We omit V because its performance was found to be very similar to that of D. For each model, three parallel MCMC chains were run for 30,000 iterations. Table 4 shows the multivariate potential scale reduction factor (Brooks and Gelman 1998) to check MCMC convergence. Customarily, values less than 1.2 are deemed satisfactory. For posterior analysis, we discarded the first 15,000 iterations of each chain and took every 30th sample, thus obtaining a final "thinned" MCMC sample of 2250 for each model.

Table 5 shows that the PBBM significantly outperforms the BNLR. This confirms what we suspected in Section 3. The assumptions underlying the deterministic physical model do not hold in actual workplace environments. Consequently, a statistical model that only has physical model and measurement error components will fit the data poorly. These facts are also confirmed by Figure 3(a)–3(c), which plot the means for the replicated data against the observed log-exposure concentrations

Table 5. DIC and GRS scores for the actual workplace data

Model	DIC	p_D	$\overline{D}$	GRS
BNLR	768.56138	0.9767	767.58468	−721.9574
D	−2857.18172	36.46883	−2893.65056	3819.8092
LT	−2856.37636	37.21186	−2893.58822	3824.98638

(dots). The solid line (with slope 1) represents equality between the model replicated means and the observations. Figure 3(a) shows the miserably poor fit of the simple BNLR. Returning to Table 5, we see that between the LT and D, the GRS indicates that the LT performs slightly better than the D, although the DIC seems to suggest they are quite similar. This, too, is reaffirmed by Figure 3(a) and 3(c).

Table 6 presents the estimated posterior means, 95% credible intervals, and Monte Carlo standard errors (MCSE) for the main parameters of the competing models. The estimates for τ_1 and τ_2 are noticeably higher for the BNLR than under PBBMs. This is unsurprising because the BNLR attributes the entire variation in the data to measurement errors, while PBBM attributes part of the variation to the underlying latent process as well.

We also see a substantial bias in the airflow (β) estimates from the BNLR, which is largely attributable to the poor fit of simple nonlinear regression models. This is not entirely surprising as models with random effects tend to estimate "fixed" effects differently from models without random effects, a problem that has been investigated in linear mixed model contexts with findings not dissimilar to ours (see, e.g., Reich, Hodges, and Zadnik 2006; Hodges and Reich 2010). What is particularly disconcerting is the narrow credible interval for the airflow as a result of the BNLR's inability to adequately capture variability. The substantive implications of the above can be serious. For example, the BNLR estimates suggest, very precisely, that the near field attains theoretical steady state after about 196.2 min, which, on the basis of the exploratory data analysis, seems to be a gross overestimation. Straightforward least-square analysis, often used by hygienists, also suffers from such biases. The PBBM approach, on the other hand, suggests this to be at around 79 min but is more cautious and yields much wider credible intervals, between 53.4 and 309 min, suggesting that

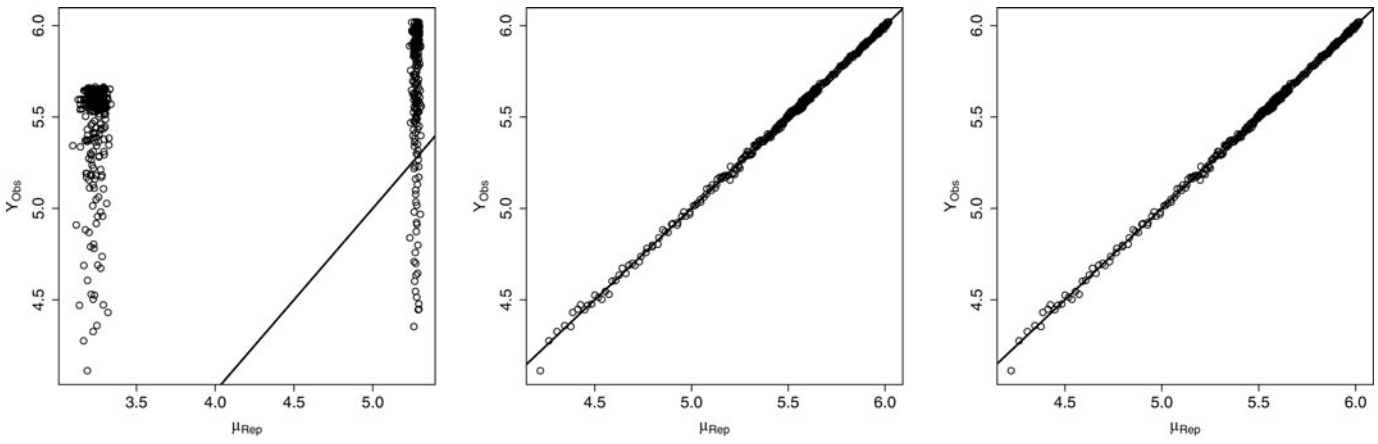

Figure 3. Posterior predictive means for the replicated data plotted against the observed log exposure concentrations for the workplace data.

Table 6. Posterior summaries for the main parameters in the BNLR and PBBM in the workplace data

BNLR				LT			
Par	Mean	95% CI	MCSE	Par	Mean	95% CI	MCSE
β	2.059	(1.999, 2.111)	0.004	β	6.570	(1.240, 12.587)	0.117
τ_1	0.371	(0.317, 0.434)	0.002	a_1	1.283	(0.776, 2.003)	0.017
τ_2	4.207	(3.635, 4.906)	0.012	a_2	1.391	(0.740, 2.227)	0.018
τ_{12}	1.249	(1.079, 1.453)	0.003	a_3	0.488	(−0.453, 1.576)	0.021
		D		τ_1	1.2e-04	(1e-04, 1.41e-04)	3.2e-07
β	6.437	(1.127, 12.732)	0.125	τ_2	0.001	(8.5e-04, 1.1e-03)	3.4e-06
a_1	1.245	(0.716, 1.926)	0.018	τ_{12}	2.55e-04	(2.1e-04, 3e-04)	8.4e-07
a_2	1.513	(0.977, 2.292)	0.011				
τ_1	1.2e-04	(1e-04, 1.44e-04)	4.6e-07				
τ_2	0.001	(9e-04, 1.2e-03)	4e-06				
τ_{12}	2.5e-04	(2.1e-04, 3e-04)	1.1e-06				

airflow cannot be precisely estimated from the data. The PBBM sufficiently protects the hygienist against spurious exposure assessment and management. (More results are available in the supplementary material "AdditionalAnalysis.pdf".)

6. DISCUSSION

We proffered a PBBM approach for predicting exposure concentrations over time in industrial workplaces. We believe our current application to be the first serious venture of Bayesian melding in the domain of industrial hygiene, a field that has a strong Bayesian presence in the form of subjective judgment but still relies largely upon least squares and straightforward Bayesian regressions (BNLR) (see, e.g., Zhang et al. 2009).

The PBBM is applicable whenever full inference on physical parameters and subsequent predictions are sought. We show that the PBBM delivers substantial, sometimes dramatic, improvements in inference than straightforward nonlinear regression. The PBBM approach reflects the variability much better, provides far superior fits to the data, yields better predictions, and, perhaps most importantly from the hygienist's perspective, provides a much more realistic assessment of the uncertainties involved in the estimation of the model.

Based upon our current findings, we advocate estimating inputs to the physical model whenever possible. We recognize that full inference here will require solving the physical model, which may be infeasible in certain settings. However, a very large number of physical processes can be formulated as general systems of linear ODEs, whose solutions closely depend upon the eigenvalues of the coefficient matrix (see "ODE.pdf" in supplementary materials). Assigning reasonable priors to the eigenvalues will yield tractable solutions to such systems making the PBBM framework widely applicable. We note that estimation of parameters embedded within the systems of ODEs is a relevant and active area of research in diverse settings (see, e.g., Johnson, Pecquerie, and Nisbet 2013, and references therein).

The rich association structures permissible within PBBM are noteworthy. Since this appears after regressing on the posited physical model, these structures can be applied even if a posited physical model were computationally prohibitive. In such cases, a distinct and smoother Gaussian process on the space of inputs can be deployed as a fast interpolator or emulator for the physical model (e.g., Kennedy and O'Hagan 2001; Bayarri et al. 2007; Conti and O'Hagan 2010) while our specifications for $\boldsymbol{\eta}(t)$ can be used exactly as here. Also, this easily adapts to physical models with high-dimensional output (e.g., Higdon et al. 2008).

A few extensions are worth noting. Clearly we have only skimmed the surface in our choice of physical models. The experimental design can be modified to collect measurements in different spatial locations at each timepoint. Space–time physical models based upon diffusion principles can then be combined with spatio-temporal stochastic processes to create highly flexible melding frameworks. In fact, enrichments such as allowing the inputs to vary over space and time can be envisioned as well. Future possibilities may also include space–time dynamical specifications for $\boldsymbol{\eta}(t)$. Finally, we recognize that a drawback of the PBBM is the somewhat limited interpretability of the stochastic process. On the other hand, the intrinsic parameters of the physical model have clearer scientific meaning. There is a tradeoff between the physical model and the additional stochastic process. It may be instructive to develop a likelihood-based model directly from the physical model (e.g., by discretizing the ODE), and then absorbing the resulting systematic pattern into the stochastic process. We identify these as future directions of research on Bayesian melding.

SUPPLEMENTARY MATERIALS

Prior settings: All information regarding which prior distributions we used in the simulation study and in the experimental data sections are in the file "PriorSettings.pdf."

Additional analysis: Extra statistical analysis for both simulation study and experimental data are provided in the file "AdditionalAnalysis.pdf."

Identifiability: The conditions for the identifiability of process parameters in (6) are outlined in the file "Identifiability.pdf."

ODE: The derivation of the solution of (1) is presented in the file "ODE.pdf."

Rcode: R programs to implement the modeling and data analysis presented in the manuscript.

[Received January 2013. Revised August 2013.]

REFERENCES

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman & Hall/CRC. [242]

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C. H., and Tu, J. (2007), "A Framework for Validation of Computer Models," *Technometrics*, 49, 138–154. [239,241,242,246]

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2011), "Space-Time Data Fusion Under Error in Computer Model Output: An Application to Modeling Air Quality," *Biometrics*, 68, 837–848. [239]

Brooks, S. P., and Gelman, A. (1998), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455. [245]

Conti, S., and O'Hagan, A. (2010), "Bayesian Emulation of Complex Multi-Output and Dynamic Computer Models," *Journal of Statistical Planning and Inference*, 140, 640–651. [239,241,246]

Fuentes, M., and Raftery, A. E. (2005), "Model Evaluation and Spatial Interpolation by Bayesian Combination of Observations With Outputs From Numerical Models," *Biometrics*, 61, 36–45. [239]

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003), *Bayesian Data Analysis* (2nd ed.), Boca Raton, FL: Chapman & Hall/CRC. [242,243]

Gneiting, T., and Raftery, A. E. (2007), "Strictly Proper Scoring Rules, Prediction and Estimation," *Journal of the American Statistical Association*, 102, 359–378. [243]

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008), "Computer Model Calibration Using High-Dimensional Output," *Journal of the American Statistical Association*, 103, 570–583. [246]

Hodges, J. S., and Reich, B. J. (2010), "Adding Spatially Correlated Errors Can Mess Up the Fixed Effect You Love," *The American Statistician*, 64, 335–344. [245]

Johnson, L., Pecquerie, L., and Nisbet, R. (2013), "Bayesian Inference for Bioenergetic Models," *Ecology*, 94, 882–894. [246]

Keil, C. B., Ten Berge, W. F., Fehenbacher, M. C., Jayjock, M. A., Nicas, M., and Reinke, P. (2009), *Mathematical Models for Estimating Occupational Exposure to Chemicals*, Fairfax, Virginia: American Industrial Hygiene Association. [238]

Kennedy, M. C., and O'Hagan, A. (2001), "Bayesian Calibration of Computer Models," *Journal of the Royal Statistical Society,* Series B, 63, 425–464. [239,241,246]

Nicas, M. (1996), "Estimating Exposure Intensity in an Imperfectly Mixed Room," *American Industrial Hygiene Association Journal*, 57, 542–550. [238]

Nicas, M., and Jayjock, M. (2002), "Uncertainty in Exposure Estimates Made by Modeling Versus Monitoring," *American Industrial Hygiene Association Journal*, 63, 275–283. [238]

Poole, D., and Raftery, A. E. (2000), "Inference for Deterministic Simulation Models: The Bayesian Melding Approach," *Journal of the American Statistical Association*, 95, 1244–1255. [239]

Raftery, A. E., and Bao, L. (2010), "Estimating and Projecting Trends in HIV/AIDS Generalized Epidemics Using Incremental Mixture Importance Sampling," *Biometrics*, 66, 1162–1173. [239]

Raftery, A. E., Givens, G. H., and Zeh, J. E. (1995), "Inference From a Deterministic Population Dynamics Model for Bowhead Whales," *Journal of the American Statistical Association*, 90, 402–416. [239]

Ramachandran, G. (2005), *Occupational Exposure Assessment for Air Contaminants*, Boca Raton, FL: CRC Press. [238,244]

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006), "Effects of Residual Smoothing on the Posterior of the Fixed Effects in Disease-Mapping Models," *Biometrics*, 62, 1197–1206. [245]

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer-Verlag. [239]

Ševčíková, H., Raftery, A. E., and Waddell, P. A. (2007), "Assessing Uncertainty in Urban Simulations Using Bayesian Melding," *Transportation Research Part B*, 41, 652–669. [239]

——— (2011), "Uncertain Benefits: Application of Bayesian Melding to the Alaskan Way Viaduct in Seattle," *Transportation Research Part A*, 45, 540–553. [239]

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002), "Bayesian Measures of Model Complexity and Fit," *Journal of the Royal Statistical Society,* Series B, 64, 583–639. [242]

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer-Verlag. [242]

Zhang, Y., Banerjee, S., Yang, R., Lungu, C., and Ramachandran, G. (2009), "Bayesian Modeling of Exposure and Air Flow Using Two-Zone Models," *The Annals of Occupational Hygiene*, 53, 409–424. [239,240,246]