The Annals of Applied Statistics
2023, Vol. 17, No. 2, 1175–1198
https://doi.org/10.1214/22-AOAS1664

LATENT MULTIVARIATE LOG-GAMMA MODELS FOR HIGH-DIMENSIONAL MULTITYPE RESPONSES WITH APPLICATION TO DAILY FINE PARTICULATE MATTER AND MORTALITY COUNTS

BY ZHIXING XU[a], JONATHAN R. BRADLEY[b] AND DEBAJYOTI SINHA[c]

Department of Statistics, Florida State University, [a]*zhixing.xu2@sanofi.com,* [b]*jrbradley@fsu.edu,* [c]*sinhad@stat.fsu.edu*

Precise estimation of daily fine particulate matter with a diameter ≤ 2.5 microns (PM2.5) and mortality in the U.S. is an important research challenge in public health because high levels of PM2.5 have been linked to several serious health problems, including lung disease, cardiovascular disease, and stroke. This motivates us to develop a joint Bayesian hierarchical model for bivariate spatial data to obtain precise spatial predictions of two types of responses, continuous skewed PM2.5 levels, and discrete mortality counts over U.S. counties. Our novel modeling framework address several challenges in the area of spatial prediction of mortality counts and PM2.5 levels. Specifically, our model allows for spatial variability and dependence of two types of responses, accommodate an unknown nonlinear spatial relationship between mortality and PM2.5 through basis function expansions, improve the precision of predictions at counties with undisclosed/missing observations, and allow for different missing data patterns for mortality and PM2.5. Furthermore, we introduce a new local measure of association for the cross-dependence between mortality and PM2.5 level. To address the burden of Bayesian computation for large databases, we use the dimension reduction tool and the shared conjugate structure between the Weibull distribution, Poisson distribution, and the multivariate log-gamma distribution. We provide a simulation study to illustrate the performance of our method. Our joint spatial model of "multitype responses" (discrete and continuous responses) and associated Bayesian method are used to analyze bivariate spatial data of daily averaged PM2.5 levels in air and mortality counts (due to diseases related to lung, cardiovascular, respiratory, and stroke) from the Centers for Disease Control and Prevention (CDC) database.

1. Introduction. The National Academy of Sciences has consistently labeled daily fine particulate matter with a diameter of ≤ 2.5 microns (PM2.5) as an important quantity to monitor in air for the assessment of U.S. public health (Burnetta et al. (2018)). This is partially due to the fact that PM2.5 level in air has been shown to be highly associated with incidence/mortality of several diseases (Laden et al. (2000), Schwartz and Neas (2000), Valavanidis, Fiotakis and Vlachogianni (2008)), including lung cancer among "never-smokers" (Turner et al. (2011)), stroke (Anderson, Thundiyil and Stolbach (2012)), cardiovascular disease (Brook et al. (2010)), and overall mortality (Zigler, Dominici and Wang (2012)). This literature on mortality and PM2.5 provides an important "call to action" to have a precise prediction and understanding of the spatially-varying association between PM2.5 and mortality so that public policy makers can find geographic regions that need the most attention (Hoek et al. (2013)). The current surge in interests on the relationship between PM2.5 and mortality has lead to a spike in the amount of data that federal agencies provide on PM2.5, including the centers for Disease Control and Prevention (CDC). Each year the CDC provides hundreds of summary statistics regarding cancer incidence and mortality across U.S. counties (https://www.cdc.gov/). Consequently, there is a clear need to develop spatial statistical

Received February 2020; revised July 2022.
Key words and phrases. Bayesian hierarchical model, multitype responses, high-dimensional data, Gibbs sampler.

methodology to handle such large databases on PM2.5 and mortality. Specifically, we aim at spatial statistical methods for large databases to answer the important questions: what are the areas with high PM2.5 levels and mortality counts, how are these two variables associated spatially, and how to predict one of these variables if it is not available?

There are essentially two categories of existing statistical methods to analyze such spatial data. The first set of existing approaches treat one of mortality count and PM2.5 level as a fixed covariate while effectively ignoring the spatial variability of this fixed covariate, and implement either a generalized linear model (GLM) or a generalized linear mixed model (GLMM) for the other response variable. For example, Dockery et al. (1993) and Miller et al. (2007) included air pollution level as a fixed covariate within a Cox's proportional hazard model for mortality. Spatiotemporal random effects have been included in a nonparametric GLMM (Xu et al. (2014)) and Poisson GLMMs (Xu et al. (2016), Zhang et al. (2016)) to analyze ischemic heart disease mortality and respiratory disease while treating air pollution level as a fixed covariate. The second category of existing approaches, based on expressing the joint model as a product of marginal and conditional distribution, enforce a restrictive either log-linear or linear relationship between mortality and PM2.5 in the conditional distribution without allowing for more complex nonlinear relationships. For example, Fuentes et al. (2006) and Choi, Fuentes and Reich (2009) assume a generalized Poisson GLMM for the mortality response, given PM2.5 level, and then also a spatial statistical model for the (marginal) distribution of PM2.5 level. Both of these approaches require PM2.5 and mortality to be paired. That is, if a PM2.5 level is observed at a region, then the mortality count also needs to be available at this region and vice versa.

We acknowledge that the complex relationship between PM2.5 level and mortality may not be modeled via a direct parametric regression effect of PM2.5 level on mortality count or vice versa. The association between PM2.5 level and mortality may have an underlying "errors in covariate" (e.g., see Carroll, Gail and Lubin (1993), Carroll et al. (2006)) because the PM2.5 level in air may not be measured accurately and may be different from actual exposure to PM2.5. It is possible to also have a latent spatially varying variable affecting both responses. To address this complex and unknown nonlinear relationship between these two responses, we use a more flexible framework than existing approaches to model the spatially varying relationship between these two responses via a shared latent process expressed with the basis function expansion approach commonly used in functional analysis (e.g., see Carroll et al. (2006), Chakraborty and Panaretos (2017), Wikle (2010), among others). The use of a shared (across response type) basis function expansion also allow prediction of mortality and PM2.5 when either of them is missing in any geographical location. Motivated by PM2.5 level and mortality count data, the goal of this article is to develop a sophisticated statistical model and associated analysis to simultaneously incorporate spatial variability of PM2.5 level and mortality count, allow for different missing data patterns among PM2.5 levels and mortality, utilize the possibly nonlinear relationship between these two different response types, and allow for possibly high-dimensional data.

A well-known approach for joint modeling of multiple types of non-Gaussian responses is through a shared latent Gaussian process (LGP); see Gelfand and Schliep (2016) for standard references and recent review of LGP. For example, Christensen and Amemiya (2002) considered a general nonparametric model applicable only for multiple continuous responses measured on a spatial grid. Schliep and Hoeting (2013) proposed a general modeling framework that allows for both continuous and discrete responses using a GLMM model with shared LGP. Similarly, Clarke et al. (2017) allow for several types of responses, including continuous, right censored, zero-inflated, and binary responses. However, one major difficulty with Bayesian GLMM methods (with and without LGP) for multitype data is that the corresponding Markov chain Monte Carlo (MCMC) algorithm involves Metropolis–Hastings

(MH) steps with hard to specify proposal densities. This problem becomes exacerbated when the size of the dataset is large (as for our CDC study). However, our method for PM2.5 levels and mortality counts uses a novel extension of the existing multivariate log-gamma (MLG) distribution (Bradley, Holan and Wikle (2018)) that ensures easy-to-sample conjugate updates within a collapsed Gibbs sampler. Directly sampling from a conjugate full-conditional distribution is particularly important because this allows us to avoid selecting proposal distributions for MH steps within MCMC iterations. The MLG distribution has been used to model continuous skewed (Hu and Bradley (2018)) data and count data (Bradley, Holan and Wikle (2018), Bradley, Holan and Wikle (2020)), separately, but has not been used in a joint model of spatially varying counts and continuous responses. While the use of the MLG distribution leads to a Gibbs sampler that mixes well, we also allow for a reduced rank expression of spatially covarying terms to obtain additional computational speed-ups. Reduced rank spatial models have been shown to have high predictive accuracy and computational efficiency (e.g., see Bradley, Cressie and Shi (2015), Bradley, Cressie and Shi (2016), Heaton et al. (2018), for reviews and comparisons); for example, see the fixed rank kriging methodology (Cressie and Johannesson (2006), Cressie and Johannesson (2008), Shi and Cressie (2007)) and modified predictive processes (Finley et al. (2009)) for choices of reduced rank models for spatial only data. Reduced rank dynamic spatiotemporal models have also been shown to perform well in environmental (Katzfuss and Cressie (2011), Wikle and Cressie (1999)) and social science settings (Bradley, Holan and Wikle (2015)). While we use the dimension reduction tool in this paper, there are several recent methods that are full-rank spatial statistics methods that are computationally feasible that make use of sparse parameterizations (Bradley, Wikle and Holan (2020), Gao, Datta and Banerjee (2022), Peruzzi, Banerjee and Finley (2022)), approximate likelihood techniques (Katzfuss and Guinness (2021), Katzfuss et al. (2020), Zhang, Tang and Banerjee (2021)), and the use of subsampling techniques (Bradley (2021)).

There already exist separate univariate analyses of PM2.5 levels (Datta et al. (2016)) and of mortality counts (Quick, Waller and Casper (2018)) in the high-dimensional spatiotemporal context. However, there is no joint spatial analysis method for PM2.5 levels and discrete mortality counts to deal particularly with high-dimensional data from large national/international databases. We refer to our model as the joint Weibull and Poisson (WAP) model. For needed continued development of joint models for multitype responses, especially in the multivariate spatial statistics settings, we develop a new summary statistic suitable for such responses. The statistic is motivated by techniques to bound the correlation function of non-Gaussian responses in De Oliveira (2013), De Oliveira (2003), and Bradley, Holan and Wikle (2018), and it also bares similarity to the local indicator of spatial association (LISA; Anselin (1995)).

The squared error (empirical) of our WAP model based estimates increases at a slow rate, as the proportion of zeros in the dataset increases from 0% to 20%. It is well known that the failure to address the inflation of zero responses/counts may lead to biased estimation, and consequently, a rich literature is available for various zero-inflated models (e.g., Sellers and Raim (2016), for a discussion). Through sensitivity studies we show that our model performs well, even when there are around 20% of zero-inflated values in mortality counts from different areas. This property is particularly important because our motivating CDC database contains a moderate amount of observed zero mortality counts. Another motivation for the proposed model is that it is applicable to many other related studies in several disciplines where we come across association between high-dimensional skewed continuous response and count responses (e.g., association between daily sunlight (hours) and melanoma incidences, Leiter and Garbe (2008)). To aid readers in using WAP, we provide code on Github through the R package `CM` at: https://github.com/JonathanBradley28/CM.

In Section 2 we review the multivariate log-gamma distribution and introduce the WAP model to jointly model high-dimensional multitype survival responses. In Section 3 we conduct a simulation study to compare the performance of WAP against competing models that

ignore the dependence between Weibull and Poisson data. In Section 4 we use the WAP model to analyze the aforementioned motivating CDC dataset. We provide a conclusion in Section 5. All proofs, technical details and additional simulation/sensitivity studies are given in Supplementary Material (Xu, Bradley and Sinha (2023)) for ease of exposition.

2. Methodology.

2.1. *Univariate spatial models for Poisson and Weibull responses.* Let $Z(A_i)$ denote the mortality count at region A_i for $i = 1, \ldots, n$ within the spatial domain of interest D (e.g., U.S. counties). The areal units A_i within the domain D are disjoint, that is, $\bigcup_{i=1}^{n} A_i \subset D$, and $A_i \cap A_j$ is empty for $i \neq j$. We observe $Z(A_i)$ at $n_d \leq n$ regions and model the n_d-dimensional vector $\boldsymbol{Z} = \{Z(A_1), Z(A_2), \ldots, Z(A_{n_d})\}'$ of mortality counts as $Z(A_i) \overset{\text{ind}}{\sim} \text{Poisson}[\exp\{Y_d(A_i)\}]$ for $i = 1, \ldots, n_d$ with probability mass function

$$f\big\{Z(A_i)|Y_d(A_i)\big\} = \frac{1}{Z(A_i)!}\exp\big[Z(A_i)Y_d(A_i) - \exp\big\{Y_d(A_i)\big\}\big], \tag{1}$$

where "!" is the factorial symbol. We model both covariate and spatial effects on mean $\mu_d(A) = \exp(Y_d(A))$ of the Poisson distribution of $Z(A)$ in (1) as

$$Y_d(A) = \boldsymbol{x}_d(A)'\boldsymbol{\beta}_d + \boldsymbol{\psi}_d(A)'\boldsymbol{\eta}_d + \gamma_d(A); \quad A \in D,$$

where $\boldsymbol{x}_d(A) = (X_{d1}(A), \ldots, X_{dp}(A))'$ is a p_d-dimensional vector of known covariates observed at region A, $\boldsymbol{\beta}_d$ is a p_d-dimensional parameter vector of the covariate effects, and the subscript "d" is used to indicate that $Y_d(A)$ in (1) is a parameter of a discrete response. The r-dimensional random vector $\boldsymbol{\eta}_d$ with corresponding prespecified r-dimensional spatial basis functions $\boldsymbol{\psi}_d(A)$ model the small-scale spatial variability of mortality counts $Z(A_i)$. The term $\boldsymbol{\psi}_d(A)'\boldsymbol{\eta}_d$ is a spatial basis function expansion which has become a standard tool in modern spatial analysis (Wikle and Hooten (2010)) to model nonstationary and nonlinear processes. The form of $\boldsymbol{\psi}_d(A)$ depends on the class of areal basis function (e.g., aggregations of thin plate spline basis function and bisquare basis function, see Bradley, Wikle and Holan (2017)) chosen to model the moderate spatial variability of moratlity counts. We give examples of choices of $\boldsymbol{\psi}_d(\cdot)$ in Sections 3 and 4. The random error $\gamma_d(A)$ is assumed to capture unknown spatial error not accommodated by $\boldsymbol{x}_d(A)'\boldsymbol{\beta}_d + \boldsymbol{\psi}_d(A)'\boldsymbol{\eta}_d$. Conjugate prior distributions can be specified for the remaining unknown parameters in (1) (Bradley, Holan and Wikle (2020), Diaconis and Ylvisaker (1979)). In particular, $\boldsymbol{\beta}_d$ and $\boldsymbol{\eta}_d$ follow MLG distributions (see Section 2.5) or conditional MLG distributions (see Supplementary Material Appendix A). Classic version of LGP assumes that $\boldsymbol{\beta}_d$ and $\boldsymbol{\eta}_d$ to have multivariate Gaussian priors. The MLG distribution has some advantages over the Gaussian distribution. Specifically, the MLG distribution leads to conjugate full conditional distributions (see Supplementary Material Appendix B) and can be specified as an arbitrarily close approximation of any Gaussian distribution (Bradley, Holan and Wikle (2020)).

A univariate model for PM2.5 can be defined in a similar manner. Suppose the observed vector of PM2.5 levels in air $\boldsymbol{t} = \{T(A_1), T(A_2), \ldots, T(A_{n_c})\}'$ follow Weibull distributions $T(A_i) \overset{\text{ind}}{\sim} \text{Weibull}(\rho(A_i), b_i)$ for $i = 1, \ldots, n$, where $n_c \leq n$, the scale parameter b_i is parameterized as $b_i = \exp(-Y_c(A_i))$ and $\rho(A_i) > 0$ is the shape parameter of the Weibull distribution. The probability density function of the Weibull distribution is

$$\begin{aligned} &f\big\{T(A_i)|\rho(A_i), Y_c(A_i)\big\} \\ &\quad = \rho(A_i)T(A_i)^{\rho(A_i)-1}\exp\big[Y_c(A_i) - T(A_i)^{\rho(A_i)}\exp\big\{Y_c(A_i)\big\}\big], \end{aligned} \tag{2}$$

where the subscript "c" is used to indicate this as a model for continuous response. Further, the parameter $Y_c(A)$ is modeled as

$$Y_c(A) = \boldsymbol{x}_c(A)'\boldsymbol{\beta}_c + \boldsymbol{\psi}_c(A)'\boldsymbol{\eta}_c + \gamma_c(A); \quad A \in D,$$

where $\boldsymbol{x}_c(A)$ is a p_c-dimensional vector of known covariates and $\boldsymbol{\beta}_c$ is a p_c-dimensional parameter vector of the covariate effects. It is assumed that if $T(A)$ is observed at region A, then the vector $\mathbf{x}_c(A)$ is also observed. Similar to model (1) of mortality counts, the small scale spatial variability in PM2.5 levels are accommodated by the r-dimensional vector $\boldsymbol{\eta}_c$ and the corresponding prespecified $\boldsymbol{\psi}_c(A)$ which can be any class of areal basis function (e.g., aggregations of thin plate spline basis function and bisquare basis function, see Bradley, Wikle and Holan (2017)). In general, we can allow $\boldsymbol{\psi}_c$ to be different from $\boldsymbol{\psi}_d$. In Supplementary Material Appendix C we give a complete expression of a model for Weibull responses. Similar to model for mortality counts, we choose conjugate prior and hyper prior specifications.

2.2. *Joint bivariate spatial model for Weibull and Poisson (WAP) responses.* In practice, two different types of responses from a dataset may consist of $\mathscr{D}_c = \{T(B_i) : B_i \in D, i = 1, \ldots, n_c\}$ of continuous responses and $\mathscr{D}_d = \{Z(A_i) : A_i \in D, i = 1, \ldots, n_d\}$ of count-valued observations, where the sets $\{B_i\}$ and $\{A_j\}$ may not be exactly the same as PM2.5 and mortality have different missing/undisclosed regions. The areas $\{B_i\}$ and $\{A_j\}$ are aligned (i.e., $\{B_i\}$ and $\{A_j\}$ both consist of U.S. counties). In our application, $\{B_i\}$ is a different collection of U.S. counties than $\{A_j\}$. For the misaligned setting, one might utilize strategies for spatial change of support (e.g., see Bradley, Wikle and Holan (2016), Gotway and Young (2002), Mugglin, Carlin and Gelfand (2000), among others). As in Section 2.1, we assume continuous-valued and count-valued responses (conditioned on latent processes) are modeled as follows:

$$T(B_i) \overset{\text{ind}}{\sim} \text{Weibull}\big[\rho(B_i), \exp\{-Y_c(B_i)\}\big]; \quad i, \ldots, n_c$$

$$Z(A_j) \overset{\text{ind}}{\sim} \text{Poisson}\big[\exp\{Y_d(A_j)\}\big]; \quad j = 1, \ldots, n_d.$$

Our WAP model makes the following assumption:

$$Y_c(A) = \boldsymbol{x}_c(A)'\boldsymbol{\beta}_c + \boldsymbol{\psi}_{c\eta}(A)'\boldsymbol{\eta} + \boldsymbol{\psi}_c(A)'\boldsymbol{\eta}_c + \gamma_c(A); \quad A \in D, \tag{3}$$

where $Y_c(A)$ is the natural parameter for in Weibull response model and $\boldsymbol{\eta}$ is an r-dimensional random vector. The r-dimensional term $\boldsymbol{\psi}_{c\eta}(A)$ is prespecified and can be any class of areal basis function (e.g., aggregations of thin plate spline basis function and bisquare basis function, see Bradley, Wikle and Holan (2017)). In general, for $e = c, d$, $\boldsymbol{\psi}_{e\eta}$ can be different from $\boldsymbol{\psi}_e$, but it is not required to be different. We make similar assumptions to (3) to incorporate covariate and spatial effects for the Poisson response. That is, the WAP model makes the following assumption:

$$Y_d(A) = \boldsymbol{x}_d(A)'\boldsymbol{\beta}_d + \boldsymbol{\psi}_{d\eta}(A)'\boldsymbol{\eta} + \boldsymbol{\psi}_d(A)'\boldsymbol{\eta}_d + \gamma_d(A); \quad A \in D. \tag{4}$$

As mentioned in the Introduction, the random vector $\boldsymbol{\eta}$ is the same across *all* regions (or is "shared" across all regions) which is crucial for incorporating the dependence between $Y_c(A_i)$ and $Y_d(A_i)$. This is clearly explained by noting that

$$\text{cov}\big\{Y_c(A_i), Y_d(A_i)\big\} = \boldsymbol{\psi}_{c\eta}(A_i)'\,\text{cov}(\boldsymbol{\eta})\boldsymbol{\psi}_{d\eta}(A_i)$$

is nonzero for several i. This random vector $\boldsymbol{\eta}$ is not spatially referenced; however, the basis function expansions $(\boldsymbol{\psi}_{d\eta}(A)'\boldsymbol{\eta}, \boldsymbol{\psi}_{c\eta}(A)'\boldsymbol{\eta}, \boldsymbol{\psi}_d(A)'\boldsymbol{\eta}_d)$ themselves are spatially referenced. Consequently, when performing inference on $\boldsymbol{\eta}$, we discuss elements of a random vector common to all regions as opposed to random variables for a specific region. This use of

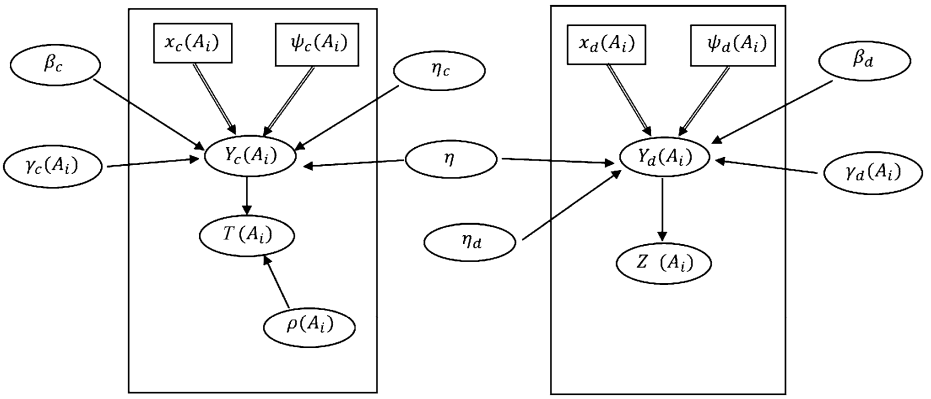

FIG. 1. *A directed graph of the WAP model with $\psi_{e\eta} = \psi_e$ for $e = c, d$. The left box represents Weibull responses, and the right box represents the Poisson response and its priors. We jointly model these two responses with $\boldsymbol{\eta}$ in the middle of the diagram.*

a multivariate random vector within a spatial basis function expansion is currently popular (e.g., see Wikle (2010), for a review). A complete statement of the WAP model is given in Supplementary Material Appendix D. We also provide a directed graph of WAP in Figure 1. The directed graphs for the univariate response models in Sections 2.1 and 2.2 are the same as Figure 1 with $\boldsymbol{\eta}$ removed.

Our method leverages the bivariate spatial dependence between PM2.5 and mortality through the shared random effect $\boldsymbol{\eta}$ to improve the precision of our predictions of both missing/undisclosed PM2.5 and mortality counts in CDC database. Fuentes et al. (2006) and Choi, Fuentes and Reich (2009) also capitalize on the joint dependence between PM2.5 and mortality by setting $T(A)$ to be a covariate in a generalized Poisson GLMM and modeling the marginal distribution of $T(\cdot)$ with a spatial statistical model. However, there are several additional benefits to our specifications in (3)–(4). First, our model does not make the restrictive assumption of a log-linear relationship between mortality and PM2.5., since the spatial basis function expansion $\boldsymbol{\psi}_{e\eta}(A)'\boldsymbol{\eta}$ (for $e = c, d$) allow the shared functional variability between PM2.5 and mortality to be unknown. Furthermore, unlike previous joint models of mortality and PM2.5, when computing our likelihood contribution of an observed mortality count $Z(A)$ for an area A, we do not need to observe or impute $T(A)$. This is because $T(A)$ is not present on the right-hand side of (4), which we avoid doing by instead using a spatial basis function expansion to enforce bivariate spatial dependence between PM2.5 and mortality. Finally, the dimension of $\boldsymbol{\eta}$, $\boldsymbol{\eta}_c$, and $\boldsymbol{\eta}_d$ can be specified to be low dimensional (i.e., dimension reduction tool) to aid with efficient computation of WAP; see Section 2.5 for more discussion on computational considerations.

Many of the full-conditional distributions associated with the WAP are conditional MLG distributions, which are very difficult to simulate from, and they either require iterative algorithms (e.g., slice sampling, Neal (2003)) or algorithms with extensive tuning (e.g., Metropolis–Hastings, Chib and Greenberg (1995)). However, we addresses this issue using a data-augmentation strategy similar to Bradley, Holan and Wikle (2018), instead of simulating from an intractable distribution (for details, see Appendices B and D).

2.3. *Properties of the correlation between spatially referenced Poisson and Weibull responses.* A crucial quantity to assess, for any model of multivariate spatial statistical responses, is the spatially varying correlation function. In this section we discuss the properties of the WAP through this correlation function. To do this, we begin with general properties of the correlation $\operatorname{corr}\{Z(A), T(A)|\boldsymbol{\theta}\} = \operatorname{cov}\{Z(A), T(A)|\boldsymbol{\theta}\}/[\operatorname{var}\{Z(A)|\boldsymbol{\theta}\}\operatorname{var}\{T(A)|\boldsymbol{\theta}\}]^{1/2}$

between Weibull response $T(A)$ and Poisson response $Z(A)$ when they *do not* necessarily follow the WAP. Here, $\boldsymbol{\theta}$ is a generic vector of unknown parameters. In this section we show that this $\operatorname{corr}\{Z(A), T(A)|\boldsymbol{\theta}\}$ is related to the efficiency ratios defined as

$$\text{(5)} \qquad \mathrm{ER}_c(A,\boldsymbol{\theta}) = \frac{\operatorname{var}\{\mu_c(A)|\boldsymbol{\theta}\}}{\operatorname{var}\{T(A)|\boldsymbol{\theta}\}} \quad \text{and} \quad \mathrm{ER}_d(A,\boldsymbol{\theta}) = \frac{\operatorname{var}\{\mu_d(A)|\boldsymbol{\theta}\}}{\operatorname{var}\{Z(A)|\boldsymbol{\theta}\}}.$$

Using conditional expectations, we have $\operatorname{var}\{Z(A)|\boldsymbol{\theta}\} = \operatorname{var}\{\mu_d(A)|\boldsymbol{\theta}\} + E[\mu_d(A)|\boldsymbol{\theta}] > \operatorname{var}\{\mu_d(A)|\boldsymbol{\theta}\}$ and, similarly, $\operatorname{var}\{T(A)|\boldsymbol{\theta}\} > \operatorname{var}\{\mu_c(A)|\boldsymbol{\theta}\}$ so that the efficiency ratios in (5) are strictly between zero and one. Proposition 1 below shows that the efficiency ratios have an important relationship with the correlation between Weibull and Poisson data.

PROPOSITION 1. *Suppose that for two given areal units A and B, $T(B)$ follows a Weibull distribution and $Z(A)$ follows a Poisson distribution, then*

$$\text{(6)} \qquad \operatorname{corr}\big\{Z(A), T(B)|\boldsymbol{\theta}\big\} = H(A,B,\boldsymbol{\theta}) \operatorname{corr}\big\{\mu_d(A), \mu_c(A)|\boldsymbol{\theta}\big\},$$

where $H(A,B,\boldsymbol{\theta}) = \mathrm{ER}_d(A,\boldsymbol{\theta})^{1/2}\,\mathrm{ER}_c(B,\boldsymbol{\theta})^{1/2}$ and $\boldsymbol{\theta}$ is a vector of unknown parameters associated with the joint distribution of $\mu_d(A)$ and $\mu_c(A)$.

PROOF. See Supplementary Material Appendix E. □

This result is similar to results connecting the relative dispersion parameter and the correlation for only Poisson counts from two areas (e.g., see Bradley, Holan and Wikle (2018), De Oliveira (2003), De Oliveira (2013)). Equation (6) shows that $\operatorname{corr}\{Z(A), T(A)|\boldsymbol{\theta}\}$ is bounded between $\max(-H(A,A,\boldsymbol{\theta}), -1)$ and $\min(H(A,A,\boldsymbol{\theta}), 1)$. Moreover, as the efficiency ratios decrease, $H(A,A,\boldsymbol{\theta})$ becomes closer to zero. Thus, low efficiency places a very strong restriction on the range of possible values for $\operatorname{corr}\{Z(A), T(A)|\boldsymbol{\theta}\}$. This is true for *any* model that defines the cross-correlation between $\mu_d(A)$ and $\mu_c(A)$; that is we are not assuming that the WAP necessarily holds to derive Proposition 1. Since the correlation function is defined by the second moment, which exists for both the Gaussian distribution and MLG distribution, Proposition 1 holds for both the WAP and a more traditional LGP.

In our application we provide plots (over A) of

$$\text{(7)} \qquad \frac{1}{|\mathcal{N}(A)|} \sum_{B \in \mathcal{N}(A)} E\big\{H(A,B,\boldsymbol{\theta})|\mathscr{D}_d, \mathscr{D}_c\big\},$$

where $\mathcal{N}(A)$ denotes a set of regions determined to be "close" (e.g., nearest neighbor) to region A and $|\mathcal{N}(A)|$ is the number of regions in the set $\mathcal{N}(A)$. By letting $N(A)$ include neighbors, we allow for the possibility that individuals may regularly commute between counties at some point in their lives (e.g., someone works in one county but lives in another). In which case the possibility that the PM2.5 in, say, the county they work in, may aggravate a medical condition eventually leading to mortality. However, one can certainly allow for $N(A) = A$.

This quantity is used as a local indicator (at location A) of the amount of model-based restrictions on the spatial/response-type associations. That is, if (7) is small, this indicates that the cross-correlation between variable types and neighbors (defined by $\mathcal{N}(A)$) is severely restricted. By "restricted" we mean $\operatorname{corr}\{Z(A), T(B)|\boldsymbol{\theta}\}$ is restricted to be between $\max(-H(A,A,\boldsymbol{\theta}), -1)$ and $\min(H(A,A,\boldsymbol{\theta}), 1)$ so that the LICA represents a type of average range statistic. This use of the effective ratio is similar to the well-known local indicator of spatial association (LISA) (Anselin (1995)). As such, we refer to (7) as a *local indicator of restrictions on cross-response-type association* (LICA). In practice, LICA can be used as an exploratory analysis tool to assess regions that have negligible cross-multivariate spatial correlations. For example, if LICA is close to zero at A, then the cross-correlations between

mortality at region A and PM2.5 at regions in $\mathcal{N}(A)$ tend to be small. This is particularly useful for the mortality and PM2.5 dataset, since we are interested in determining whether or not bivariate correlations between PM2.5 level and mortality counts are present for two neighboring areas. Consequently, the LICA may provide important localized information that can be used to guide public health officials. For example, small values of LICA at a region A may suggest that fewer public health interventions for PM2.5 related mortality are needed at region A and vice versa. Since Proposition 1 does not require the WAP to hold, one can compute the LICA statistic for both the WAP and a more traditional LGP.

The LICA can be useful for understanding the second-order (e.g., correlations) relationship between PM2.5 and mortality. However, one should also consider plotting the posterior mean of the shared random effect terms (e.g., $\boldsymbol{\psi}_{d\eta}(A)'\boldsymbol{\eta}$) to investigate the first-order (i.e., means) behavior. The LICA not only gives a sense of where small correlations are present but also a measure of uncertainty, since the correlation is bounded between the range $\max(-H(A, A, \boldsymbol{\theta}), -1)$ and $\min(H(A, A, \boldsymbol{\theta}), 1)$. In general, however, one should be aware that there are several other LISAs that one might consider to assess second order information.

2.4. *Distributional assumptions*: *The Weibull distribution and the multivariate log-gamma distribution*. As stated in Section 2.2, we assume PM2.5 is Weibull distributed. The Weibull distribution is a very flexible class of distributions and is a reasonable choice in terms of picking a distribution with positive support. Of course, one could also consider the log-normal distribution which is a common choice in the PM2.5 literature made for computational convenience. It is well known that both the log-normal and Weibull distributions are flexible model choices in a wide range of settings, and there are several comparisons between the two distributions that suggest both choices are reasonable (e.g., see Kim and Yum (2008), among several others).

We assume that random effects in our model are distributed according to the MLG distribution. Thus, in this section we give a short review on the relevant details on the multivariate log-Gamma distribution. Let $\boldsymbol{w} = (w_1, w_2, \ldots, w_m)'$ be an m-dimensional random vector of m mutually independent log-gamma random variables $w_i = \log(\gamma_i)$, where γ_i is a gamma random variable with shape $\alpha_i > 0$ and rate $\kappa_i > 0$. Then, the multivariate log-gamma random variable $\boldsymbol{q} \sim \mathrm{MLG}(\boldsymbol{c}, \boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ is defined as

$$\boldsymbol{q} = \boldsymbol{c} + \boldsymbol{V}\boldsymbol{w}, \tag{8}$$

where $\boldsymbol{c}$ is a m-dimensional vector, $\boldsymbol{V}$ is a lower-triangular $m \times m$ invertible matrix, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$, and $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_m)'$. The probability density function (pdf) of the m-dimensional random vector $\boldsymbol{q}$ is

$$f(\boldsymbol{q}|\boldsymbol{c}, \boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) = \det(\boldsymbol{V}^{-1})\left(\prod_{i=1}^{m} \frac{\kappa_i^{\alpha_i}}{\Gamma(\alpha_i)}\right) \exp\{\boldsymbol{\alpha}'\boldsymbol{V}^{-1}(\boldsymbol{q} - \boldsymbol{c}) - \boldsymbol{\kappa}' \exp[\boldsymbol{V}^{-1}(\boldsymbol{q} - \boldsymbol{c})]\}.$$

We use the MLG distribution to be priors of the fixed and random effects in equations (3) and (4). We use a gamma prior for each $\rho(A_i) > 0$ which is updated within a Gibbs sampler using a Metropolis–Hasting algorithm. The main advantage of the MLG distribution is that the likelihood contributions from both $\boldsymbol{t}$ and $\boldsymbol{Z}$ have double exponential forms, similar to that of the MLG distribution. This can be exploited to implement a collapsed Gibbs sampler. Further details on the properties of the MLG distribution are in Supplementary Material Appendix B.

2.5. *Computational considerations*. There are two important features of the WAP that lead to efficient computations. First, the use of the MLG distribution may avoid tuning/Metropolis–Hastings steps within a collapsed Gibbs sampler. Consequently, the mixing of the sampler may improve, since we avoid tuning proposal distributions, and the corresponding effective sample will be larger. However, we note that tuning proposal distributions

in the Metropolis algorithm does not necessarily lead to bad mixing, but rather, this is our experience in high-dimensional parameter space settings. In this article we investigate mixing using trace plots and the Gelman–Rubin diagnostic (Gelman and Rubin (1992)).

Second, reduced-rank modeling replaces the distribution of a correlated high-dimensional random vector with a distribution for a low-dimensional random vector. Since the $r \times r$ matrix $\mathbf{V}$ (this matrix valued-parameter is analogous to the Cholesky decomposition of a covariance matrix) is low rank, there is no difficulty involved with inverting a large matrix. There are several models that address the computational problem of inverting covariance matrices for Gaussian processes; for example, see the conditional autoregressive model (Besag (1974), Besag (1986), Besag, York and Mollié (1991)) which parameterize the precision directly. As r increases, the storage of the $r \times r$ matrix $\mathbf{V}$ becomes a computational issue. Reduced-rank methodologies have faced some criticisms recently (Bradley, Wikle and Holan (2020), Stein (2014)). However, these studies focus on the case where there are few to no covariates included in the model. In general, covariates may be able to capture the (spatial) functional complexity of the unknown process which would imply that large values of r are not needed. This point often leads to standard methods to test for the need to model spatial dependence at all (e.g., see Moran (1950), for an early reference). We consider the setting where there are a moderate number of covariates. Hence, it may be reasonable to specify $r \ll \min(n_c, n_d)$.

There are, of course, several other computational algorithms available besides Gibbs sampling. For example, Hamiltonian Markov chains (HMC) allows one to directly simulate from the joint distribution avoiding the need for block updating in a Gibbs sampler (Neal (2011)). However, this approach is not feasible for high-dimensional datasets. Integrated nested Laplace approximations (INLA) (provide approximate) samples from the marginal posterior distributions (Rue, Martino and Chopin (2009)). Since we are specifically interested in joint inference (e.g., see the LICA statistic), INLA does not serve our purposes. Additionally, INLA is really only feasible when there are a few number of parameters (e.g., see a recent review Martino and Riebler (2019), which suggests no more than 15 parameters) which makes it difficult to use for WAP. A method referred to as TMB can handle much larger parameter spaces but, similar to INLA, is restricted to marginal inference (Kristensen et al. (2016)). Of course, one could develop a different model with a different parameterization that could be implemented with INLA, and further computational gains could be made.

3. Simulation study. The primary aim of this paper is to jointly analyze PM2.5 (continuous) and mortality (count) to improve the precision of spatial predictions. Thus, in Section 3.1, we illustrate the high predictive performance of WAP compared to independent (unitype) analyses of Weibull and Poisson responses. Additionally, we show the benefits of using the MLG distribution in place of a Gaussian distribution within the GLMM framework in Section 3.2.

3.1. *Illustration of the WAP's performance.* To illustrate the robustness of our model, we choose a simulation model that differs from the model we fit the data. That is, we specify the latent processes to have a specific form that is different from the prior specifications of WAP. Specifically, suppose $T(A_i) \sim \text{Weibull}(\rho(A_i), \exp(-Y_c(A_i))$ and $Z(A_i) \sim \text{Poisson}(\exp(Y_d(A_i)))$, where

$$
\begin{aligned}
&Y_c(A_i) = b_1 + c_1 \sin(A_i) + \epsilon_1(A_i), \qquad \epsilon_1(A_i) \overset{\text{iid}}{\sim} \text{Normal}\big(0, \sigma_c^2\big),\\
&Y_d(A_i) = b_2 + c_2 Y_c(A_i) + \epsilon_2(A_i), \qquad \epsilon_2(A_i) \overset{\text{iid}}{\sim} \text{Normal}\big(0, \sigma_d^2\big), \quad i = 1, \ldots, n.
\end{aligned} \tag{9}
$$

Here, we let $A_i = i$ and $D = \cup A_i$ where $i = 1, \ldots, n$. We provide multiple simulations of multitype spatial fields. The values of b_2 are used to control for the number of zero counts,

and $\sigma_c = 0.2$ and $\sigma_d = 2.4$ are chosen to control the efficiency ratios ER_c and ER_d, given in (5); see Supplementary Material Appendix F for how we specify b_2, σ_c, and σ_d.

To measure the performance of the predictions, we define the sum of squared error (SSE) as

$$\text{SSE}_{ej} = (\hat{Y}_{ej} - Y_{ej})'(\hat{Y}_{ej} - Y_{ej}), \quad e = c, d, \, j = 1, 2,$$

where $\hat{Y}_{cj}(A) = \boldsymbol{x}(A)'\hat{\boldsymbol{\beta}}_c + \boldsymbol{\psi}_{c\eta}(A)'\hat{\boldsymbol{\eta}} + \hat{\gamma}_c(A)$ and $\hat{Y}_{dj}$ has a similar formula. When $j = 1$ ($j = 2$), the posterior means $\hat{\boldsymbol{\eta}}$, $\hat{\boldsymbol{\beta}}_c$, $\hat{\boldsymbol{\eta}}_c$, $\hat{\boldsymbol{\gamma}}_c$, $\hat{\boldsymbol{\beta}}_d$, $\hat{\boldsymbol{\eta}}_d$, and $\hat{\boldsymbol{\gamma}}_d$ are computed using the WAP (univariate models). For this simulation study we find better performance by prespecifying the elements of $\boldsymbol{\psi}_c$ to be identically zero for all A, $\mathbf{V} = \mathbf{I}_r$, and shape parameters of the fixed and random effects set equal to one. Both Y_e and the data are observed at the same regions A_i when computing SSE_e. Hence, SSE_e measures spatial smoothing. In Section 3.2 we assess the performance of predicting Y_e at a missing region relative to competing approaches.

We generate 100 Poisson and 100 Weibull responses so that $n = 200$ observations. We define $\boldsymbol{X}$ to be a $n \times p$ covariate matrix with $p = 2$, and each element in $\boldsymbol{X}$ is generated from a Bernoulli distribution with success probability equals to 0.5. We do this to mimic the application which consists of categorical covariates. We choose thin plate spline basis functions (i.e., Wahba (1990), $\phi(r) = r^2 ln(r)$) to calculate elements in $\boldsymbol{\psi}_{c\eta}$, $\boldsymbol{\psi}_{d\eta}$, and $\boldsymbol{\psi}_d$ and set $r = 5$. The true value of the shape parameter $\rho(\cdot)$, when generating Weibull data, is from a gamma$(10, 0.1)$, and all regions share the same value of the Weibull shape parameter.

For illustration, we simulate data according to (9) with $n = 200$, $b_1 = -3$, $b_2 = 8$, $c_1 = 1.2$, $c_2 = 1.5$, $\rho(A) \equiv \rho$, $\rho \sim \text{Gamma}(10, 0.1)$, and define σ_c and σ_d so that the ER_c and ER_d are moderately small (i.e., ER is roughly one-fifth). Additionally, $b_2 = 8$ leads to roughly (4.45%, 12.5%) zero-valued counts.

Figure 2(a), (b) provides scatterplots of the true values of Y_e vs. their respective posterior means; $e = c, d$. In Figure 2(a) the points fall roughly on the 45° line which suggests we obtain reasonable predictions of the latent process for Weibull data (i.e., Y_e). The points in Figure 2(b) are less variable than that in Figure 2(a) about the line $y = x$ which suggests more precise predictions of the Poisson latent process. However, the predicted value of Y_d are smoother than the actual values of Y_d in that small (large) values of the true value of Y_d are slightly overestimated (underestimated).

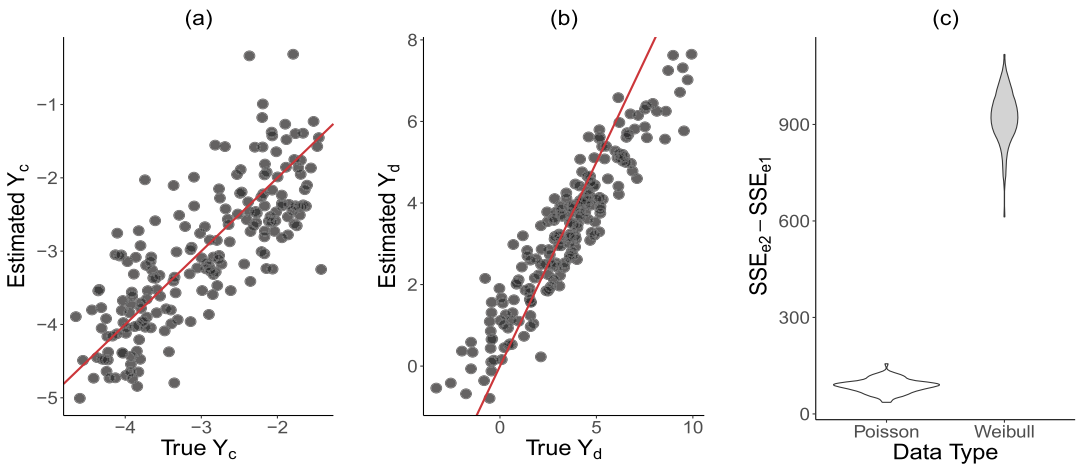

FIG. 2. *Panel (a) shows a scatterplot of the posterior mean of $Y_c(\cdot)$ vs. the true value of $Y_c(\cdot)$. Panel (b) shows a scatterplot of the posterior mean of $Y_d(\cdot)$ vs. the true value of $Y_d(\cdot)$. Panel (c) displays violin plots of $\text{SSE}_{e2} - \text{SSE}_{e1}$ by data type over 100 independent replicates.*

Figure 2(c) provides a violin plot of the difference $\text{SSE}_{e2} - \text{SSE}_{e1}$ over 100 independent replicates of the data $\{Z(A_i)\}$. Positive values indicate preferable performance of WAP over the unitype model. In Figure 2(c) we see that the WAP consistently outperforms the unitype predictions since the violin plots are entirely above zero. Additionally, we see larger improvements for the Weibull predictions. This suggests the cross data-type dependence is successfully leveraged to produce more precise predictions.

The results in Figure 2 are based on simulations with $\text{ER}_c = \text{ER}_d = 1/5$, a moderate number of zeros in the dataset, and five basis functions. However, it is well known that the number of basis functions can greatly effect the performance of spatial mixed effects models (Stein (2014)). Furthermore, the efficiency ratio can be thought of a type of signal to noise ratio which is a well-known factor for assessing the predictive performance of functional data (Wahba (1990)). Finally, it is known that an overwhelming number of zeros can lead to difficulties in prediction for spatial statistical models (De Oliveira (2013)). As such, in Supplementary Material Appendix F we provide several sensitivity studies to observe the predictive performance of WAP (relative to unitype analyses) across different choices of the number of basis functions, ER, and the number of zeros in the observed dataset $\{Z(A_i)\}$. We now briefly summarize the findings in Supplementary Material Appendix F. The WAP model is preferable in terms of SSE to the univariate response model for all number of basis functions considered ($r = 2, 5$, and 10), ER, and proportion of zeros. In general, the gain in predictive performance (i.e., SSE) was preferable for Y_c. Additionally, prediction of Y_c appears invariant to the proportion of zeros in the count-valued dataset.

3.2. *A comparison of WAP and a latent Gaussian process model.* LGP models have been widely used in spatial statistics (e.g., see Gelfand and Schliep (2016), for a recent discussion). In this section we compare WAP to a comparable LGP in terms of both the predictive performance and computational performance. The data models in the LGP remain the same (e.g., the data are assumed to be Weibull and Poisson). We replace the MLG distribution with a Gaussian distribution to define the process and parameter models, assign inverse gamma priors on variance parameters, and set the associated mean parameters equal to zero. We generate data according to (9) and choose the number of basis function to be 10. When implementing the LGP, we fix the value of the shape parameter of the Weibull distribution to its true value, as the performance of the LGP is highly sensitive to this parameter; see Supplementary Material Appendix G for the details of the priors for the LGP model.

The LGP model is implemented using the publicly available software "Just Another Gibbs Sampler" (JAGS; see Depaoli, Clifton and Cobb (2016), for a recent review). The MCMC ran for 10,000 iterations with a burn-in of 5000 iterations which are the same specifications for the WAP model. JAGS software indicates convergence of the LGP model. Convergence was also investigated using the Gelman–Rubin diagnostic (Gelman and Rubin (1992)) which also suggests that the MCMC chains converge in both models. JAGS is an all-purpose coding platform that has become fairly standard for implementing LGPs. However, JAGS is not promoted to be computationally fast, and one should be aware that our conclusions regarding the computational performance of the LGP may differ if other computing resources are considered.

In Table 1, summaries of the Gelman–Rubin (GR) diagnostic and the effective sample size (ESS) (ESS; Gelman et al. (2013), for a standard reference) over 500 simulation replicates are provided for assessing convergence and the efficiency of the two WAP models (i.e., our proposed WAP/MLG and the alternate WAP/LGP specification). Due to the large amount of parameters, we summarize the coefficient parameters that are particularly important in analyses, and we provide the results of coefficient parameters here as example. Results of other parameters are similar to this. The mean ESS in the WAP/MLG model is more than

TABLE 1
Summaries of the Gelman–Rubin (GR) diagnostic and effective sample size (ESS) over 500 independent replicates. Here, Pr(GR diagnostic < 1.1) represents the proportion of simulation replicates where GR is between 1 and 1.1

Parameters	β_{c1}	β_{c2}	β_{d1}	β_{d2}
		WAP/MLG		
Mean ESS	1476.2	1480.6	1513.0	1505.0
Pr(GR diagnostic < 1.1)	100%	100%	100%	100%
		WAP/LGP		
Mean ESS	564.5	554.7	751.0	851.0
Pr(GR diagnostic < 1.1)	75.2%	72.8%	88.0%	85.4%

two times that of the WAP/LGP model. This provides motivation for the use of the MLG distributional assumption and provides evidence that the WAP/MLG model is considerably more computationally efficient than a traditional WAP/LGP model. Additionally, the GR diagnostic from WAP/MLG is consistently between 1 and 1.1 which indicates convergence of parameters in WAP/MLG model. In general, 1 to 1.1 is the desired range of GR (Gelman and Rubin (1992)). However, even in this small data setting, several simulation replicates had GR values for the WAP/LGP model outside of the desired range. This shows the use of MLG improves mixing of the Gibbs sampler.

The average (over locations) LICA for WAP is 1.077, while the average LICA value for the LGP model is small at 0.005. This illustrates that the WAP model has less restrictions on the correlation between different response types, while the LGP model implies weak cross-response-type correlations. To assess the relative performance of competing methods when predicting at missing region, we compute the testing squared error (TSE), defined as

$$\text{TSE} = \sum_{A \in \mathcal{D}_{\text{test}}} \big[\log\{\hat{Y}_c(A)\} - \log\{Y_c(A)\}\big]^2 + \big[\log\{\hat{Y}_d(A)\} - \log\{Y_d(A)\}\big]^2, \tag{10}$$

where the regions D_{test} represent 10% randomly selected regions for each variable not included to produce the predictors $\hat{Y}_c$ and $\hat{Y}_d$. We compare on the log-scale for visualization purposes. In Figure 3 we display density plots of TSE and SSE values (for both WAP model and the LGP model) and central processing units (CPU) in seconds (Panel 3b) over 500 independent replicates of the data generating according to (9). The SSE values associated with the WAP model are centered, roughly, around 300, while the SSE values of the LGP model are centered at, roughly, 320. Similarly, the WAP outperforms the LGP in terms of TSE. Moreover, the WAP model appears less variable over the replicates of the data, as the standard deviation of both the TSE and SSE values are much smaller for the WAP model than for the LGP model. In Panel (3b) we see that the CPU time of the LGP model is centered at, roughly, 406 seconds, and the running time for the WAP model is centered at, roughly, 47 seconds. The range of running times of the WAP model is much smaller than the range of the running times of the LGP model.

Another possible comparison with WAP would be a linear model for coregionalization (LMC; Banerjee, Carlin and Gelfand (2015), for a standard reference) implemented via an integrated nested Laplace approximation (INLA). Although the LMC is a model for point referenced spatial data (and we are interested in areal data), it is arguably a reasonable choice for areal data observed on regularly spaced areal units. We are grateful to a referee who provided code which required modification to allow for Poisson/Weibull responses; specifically, see the code from Section 3 in Krainski et al. (2018). A plot of SSE and computation times over the 500 simulation replicates are provided in Figure 3. The LMC/LGP model provides worse

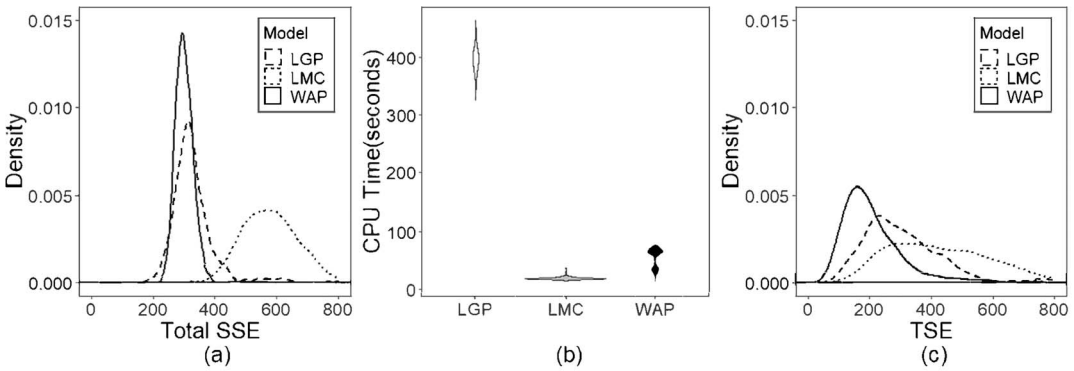

FIG. 3. *The density plots of the total SSE* (*i.e.*, $\mathrm{SSE}_{c1} + \mathrm{SSE}_{d1}$) *of LGP and WAP models* (*panel* (*a*)), *Violin plot of the CPU times* (*seconds*) *of WAP and LGP models* (*panel* (*b*)), *and the density plots of the total TSE of LGP and WAP models* (*panel* (*c*)).

results, as it consistently has larger TSE and total SSE, as compared to WAP as seen in Figure (3). The CPU time of the LMC/LGP model is much less than the WAP/MLG model due to the use of INLA as seen in Panel (3b). However, both methods can be computed quickly for these data. A table summary of Figure 3 is in Supplementary Material Appendix H.

4. Analysis of mortality and PM2.5 data from the CDC. In this section we analyze one of big multitype spatial datasets consisting of air pollution measures and mortality counts of diseases related to air quality provided by the Centers for Disease Control and Prevention (CDC). This particular CDC dataset has mortality counts for 12,760 counties/regions and PM2.5 values for 3111 regions. In line with a large literature developing the relationship between PM2.5 and various diseases (e.g. Dominici et al. (2006), Franklin, Zeka and Schwartz (2007), Kampa and Castanas (2008), Turner et al. (2011)), our goal is to jointly model and analyze averaged PM2.5 in $\mu g/m^3$ and mortality counts of diseases related to lung, cardiovascular, respiratory, and stroke in 2011 to understand the spatial variation in the association between these two responses as well as to predict these responses for all U.S. counties (including the counties where at least one of such responses are missing). WAP explicitly allows researchers to incorporate these types of multitype dependencies, since we incorporate spatial variability of both response types.

For mortality count response we use $p_d = 13$ to incorporate the following available covariates: indicators of gender (male and female), three race groups (Asian or Pacific Islander, Black or African American and White), and eight age groups (eight classes from 15 to 85+). For PM2.5 response there was no immediate covariate information available, and hence, we use $p_c = 1$ with only the intercept. It is well known that gender, race, and age are important predictors for mortality due to lung, cardiovascular, and respitory diseases (e.g., see Anderson, Thundiyil and Stolbach (2012), Brook et al. (2010), Zigler, Dominici and Wang (2012), among several others). We use the Watanabe–Akaike information criterion (WAIC; Watanabe (2010)) to choose the number of basis functions. We consider the number of basis functions to be 20, 30, 40, 50, 100, and, 200, and the value with smallest WAIC is $r = 40$. We assume $\rho(A)$ to be constant across each state, and the remaining prior settings are kept the same as they were in our simulation study. We also use the bisquare basis function from from Cressie and Johannesson (2008). The bisquare basis function is defined as

$$\psi_j(A_i) \equiv \left\{1 - \left(\frac{\|\boldsymbol{u}_i - \boldsymbol{v}_j\|}{r_l}\right)^2\right\}^2, \quad \text{for } \|\boldsymbol{u}_i - \boldsymbol{v}_j\| < r_l; i = 1, \ldots, N, j = 1, \ldots, r, \tag{11}$$

where $\boldsymbol{u}_i$ is the centroid of county A_i, $\boldsymbol{v}_j$ is ith knot location, and r_l is 1.5 times the median of distance between knots in the set $\{\boldsymbol{u}_i : i = 1, \ldots, r\}$. We define $\boldsymbol{v}_j$ to be a centroid as

well, and hence, r_ℓ is also computed based on distances between centroids. We find that using the median between knot distances to define r_l can produce a smaller WAIC for this dataset than the minimum between knot distances, as used in Cressie and Johannesson (2008); however, this may not be true for every dataset. We also consider several choices of classes of radial basis functions and found that the bisquare basis functions gave reasonable results. However, in general there are several other reasonable choices including wavelets, Fourier basis functions, splines, and other radial basis functions. We run the MCMC algorithm for 30,000 iterations and burn-in the first 20,000 iterations. The convergence is verified by trace plots (examples in Supplementary Material Appendix I) and the Gelman–Rubin diagnostic.

4.1. *Prediction, estimation, and goodness-of-fit.* In our data analysis we calculate the quantiles of elements in $\boldsymbol{\eta}$ to investigate the need for multitype dependence. We find that 14 (35%) elements have pointwise credible intervals that are greater than zero, and seven (17.5%) elements have pointwise credible intervals that are completely less than zero. We also found that the posterior probability of the elements being smaller than zero is, roughly, 18%, and the posterior probability of the elements being larger than zero is, roughly, 33%. This provides evidence for incorporating cross-response type dependence. When considering different types of basis functions, the credible intervals consistently support multitype dependence. The mean squared errors (MSE) between the posterior expected value and the log observations are 0.493 and 0.182, respectively. Figure 4 shows both scatterplots of the estimated values (posterior means) vs. the observations (Poisson on log scale) and density plots of the residuals by responses type. Aldaz (2009) includes a bias correction term when interpreting the expected value of the data on the log-scale, which we include in the Panels (b) in Figure 4. The scatterplots for both response types suggest that the WAP model performs well, and the residual plots look symmetric and unimodal. The Poisson residuals have a slight right skewness. This suggests that we have smoothed our Poisson estimates which motivates us to consider the posterior predictive p-value to determine if we have over-smoothed our Poisson estimates.

To assess the goodness-of-fit of WAP model, we use the posterior predictive p-value (Gelman, Meng and Stern (1996), Meng (1994)) with the chi-squared criterion. Here, the Monte Carlo approximation of the posterior predictive p-value is $\frac{1}{B}\sum_{b=1}^{B} I(\chi_b^2 > \chi_o^2)$, where $B = 10{,}000$, $I(\cdot)$ is the indicator function, χ_b^2 is the Chi-statistic between the posterior mean of the responses and the bth replicate of the posterior predictive distribution, and χ_o^2 is the Chi-square statistic between the posterior mean of the responses and the observations. It is desirable for a model to have posterior predictive p-values around 0.5. The posterior predictive p-value of the Weibull responses in WAP model and Univariate Weibull model are 0.670

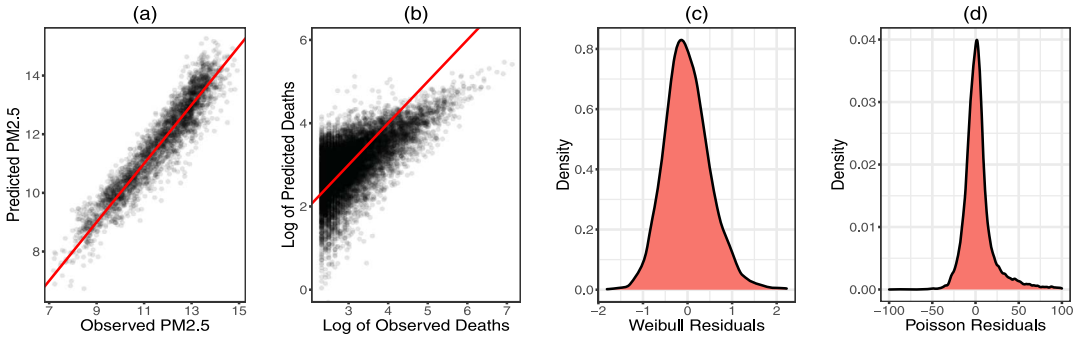

FIG. 4. *Scatterplot of observed vs. predicted PM*2.5 (*panel* (*a*)), *scatterplot of observed vs. predicted log-mortality* (*panel* (*b*)), *and density plots of the residuals of PM*2.5 (*panel* (*c*)) *and the residuals of mortality counts* (*panel* (*d*)) *obtained from the WAP model based analysis with* 40 *basis functions*.

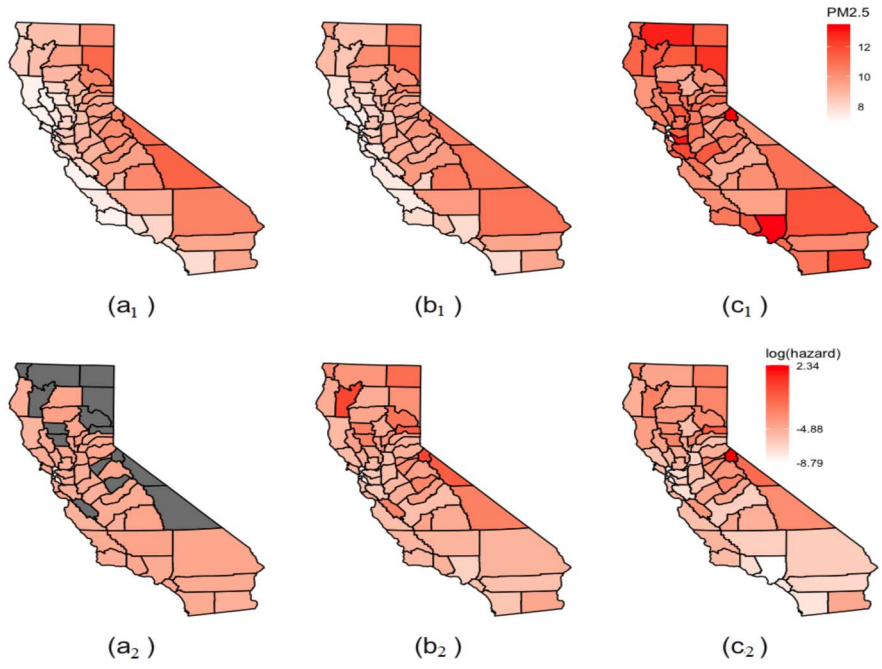

FIG. 5. *Maps of observed PM*2.5 (*panel* (a_1)) *and posterior means of PM*2.5 (*panel* (b_1)) *over different counties in California from WAP model and posterior means of PM*2.5 (*panel* (c_1)) *over different counties in California from LMC model. Maps of observed log risks* (*Panel* (a_2)), *predicted log risks of mortality* (*Panel* (b_2)) *from WAP model, and predicted log risks of mortality* (*Panel* (c_2)) *from LMC model. These predictions are over counties in California for white women over* 85 *suffering from lung cancer. Grey marks the regions with no data available.*

and 0.989, respectively. The posterior predictive p-value of the Poisson responses in WAP model and Univariate Poisson models are 0.510 and 0.568, respectively. These values suggest that the WAP model fits the data well, and the unitype Weibull model and the unitype Poisson model overfit the data.

In Figure 5 we present the predictions (i.e., posterior means) for the state of California. We obtain predictions over all U.S. counties, but for the sake of ease of visual illustration we present results over counties in California only. Figure 5(a_1), (b_1) shows maps of observed data and predictions from WAP for PM2.5 over counties in California. We see that both maps show an increasing trend of PM2.5 from west to east. Some counties have slightly higher predicted values than the raw data value.

We present our results for mortality risk, defined by the proportion of deaths out of the known population size, of white females over 85 years suffering from lung cancer (particular group with the largest number of observations). Figure 5(a_2), (b_2) presents the county-level maps of log-observed and log-predicted values (i.e., posterior means) of the mortality risks. We again see an increasing trend with smaller risk of deaths on the western coastal areas than on the eastern areas.

By comparing the predicted maps of PM2.5 and the predicted mortality risk of white females over 85 years old, we see that the mortality risk is increasing from the west to east, in general, which is similar to the trend of PM2.5. This adds additional evidence to the lung disease literature that PM2.5 is related to lung diseases.

4.2. *Comparisons.* As discussed in the Introduction, it is common to directly define a conditional model and to treat PM2.5 as a covariate and mortality as the response. Thus, we compare WAP to the conditionally specified model (the Poisson/MLG model) that uses

mortality as the response and PM2.5 as a covariate. The data model for $Z(A)$ is Poisson with mean $\lambda(A)$, where

$$\log\{\lambda(A)\} = \beta_T T(A) + \mathbf{x}_d(A)'\boldsymbol{\beta}_d + \boldsymbol{\psi}_d(A)'\boldsymbol{\eta}_d + \gamma_d(A); \quad A \in D, \tag{12}$$

where β_T, $\boldsymbol{\beta}_d$, $\boldsymbol{\eta}_d$, and $\gamma_d(A)$ are given MLG priors. The variance, shape, rate, and matrix-valued parameters are the same as in the model for WAP. We choose the number of basis functions to be $r = 40$ which is the same as the WAP model. The Poisson/MLG model is implemented using the R package `CM`.

We also compare to a LMC model via R-INLA which is the same approach we used in Section 3.2. However, the Weibull distribution is replaced by a log-normal distribution in this example. This is because, when using the Weibull distribution, R-INLA produces a calculation error (i.e., an eigenvalue problem in the Hessian matrix). However, this issue does not arise when we using the log-normal distribution. Posterior summaries from the LMC model are provided in Figure 5(c_1), (c_2). Here, we see similar predictions between the two models with the LMC model producing smoother predictions (i.e., closer to a global mean).

In Table 2, we provide the WAIC and the continuous rank probability score (CRPS Gneiting and Raftery (2007)), and we use a loss function formulation of the CRPS where small values are preferable. Calculation is done using the R package `scoringRules` (Jordan, Krüger and Lerch (2017)). The CRPS is a commonly used criterion that assesses the performance of the cumulative distribution function (we choose three quantiles of the data to evaluate the cumulative distribution function in Table 2). In general, these criteria suggest that the WAP model is preferable, followed by the LMC model, and then the Poisson/MLG. This corresponds to intuition considering that both the WAP model and LMC model incorporate bivariate-spatial covariances. The WAIC asseses out-of-sample predictive performance, as it asymptotically approaches a leave-one-out cross-validation (see Watanabe (2010), for more details). We also include the TSE metric (replacing Y_e with the testing data) of (10) for comparing out-of-sample performance of the joint models for Poisson and Weibull data in Table 2, where 10% of the data for each response are treated as testing. The TSE and WAIC suggest that the WAP model outperforms LMC in terms of predictive performance. Additionally, the CRPS provides evidence that, for this dataset, the cumulative distribution function implied by the WAP model is preferable to that of the LMC.

All three models can be computed in a reasonable amount of time. The CPU time of the WAP model with 30,000 iterations is, roughly, 2.3 hours which is longer than the one hour CPU time for the conditionally specified model and longer than the seven minutes to implement the LMC model. To assess convergence of the MCMC for the WAP model, we use trace plots and Gelman–Rubin diagnostic. Due to large number of the parameters, we show example trace plots and summaries of the Gelman–Rubin diagnostic. Example trace plots and summaries of diagnostics for the WAP model are given in Supplementary Material Appendix

TABLE 2

WAIC, CRPS, and CPU time (minutes) by model. The CRPS is computed at 2.5%, 50%, and 97.5% quantiles of each variable. WAP is only implement with MLG specifications since the MCMC for the LGP model did not converge

Model	WAIC	TSE	CRPS PM2.5			CRPS Mortality			CPU
			2.5%	50%	97.5%	2.5%	50%	97.5%	
WAP	221,782	1239	0.046	0.1208	0.6521	0.4867	2.4983	70.7535	138
LMC	293,511	1,244,276	0.0561	1.5443	4.8435	1.0457	10.8531	109.6263	7
Poisson MLG	1,770,657	–	–	–	–	9.8147	17.8127	129.7433	57

I and suggest convergence. We do not provide any inferential comparisons to the alternate WAP/LGP specification, since computationally, it appears that the WAP/LGP specification does not converge (GR diagnostics larger than 1.1). This again provides motivation for the WAP/MLG specification which is computationally feasible for this analysis.

4.3. *Inference on nonlinear relationships, covariate effects, and the dependence between mortality and PM2.5 data from the CDC.* In Sections 4.1 and 4.2, we provided evidence that the LMC and the more standard linear model assumuption in (12) perform worse in terms of WAIC and CRPS than the WAP; the WAP provided predictions for this dataset in a computationally efficient manner, obtained reasonable goodness-of-fit diagnostics, has small in-sample and out-of-sample error, and cross-dependence appears present between mortality and PM2.5. However, we have yet to describe the patterns of the nonlinearity/cross-dependence between mortality and PM2.5 and investigate covariate effects to obtain a deeper understanding on the relationship between PM2.5 and mortality. We now provide these results for counties in California.

In Figure 6 we plot the posterior means of $\boldsymbol{\psi}_{d\eta}(A)'\boldsymbol{\eta}$ (Panel a), $\boldsymbol{\psi}_{d}(A)'\boldsymbol{\eta}_{d}$ (Panel b), $\boldsymbol{\psi}_{c\eta}(A)'\boldsymbol{\eta}$ (Panel c), $\boldsymbol{\psi}_{c}(A)'\boldsymbol{\eta}_{c}$ (Panel d) over counties in California. In general, the shared random effects term for both Poisson and Weibull data show large-scale spatial patterns. In particular, the shared random effect term for Poisson data (i.e., $\boldsymbol{\psi}_{d\eta}(A)'\boldsymbol{\eta}$) identifies a large value in Trinity county (in the northwest) and somewhat constant values elsewhere. The shared random effect term for Weibull data (i.e., $\boldsymbol{\psi}_{c\eta}(A)'\boldsymbol{\eta}$) shows a smooth decreasing pattern in the southeast direction. Both of the respective individual random effects terms (i.e., $\boldsymbol{\psi}_{d}(A)'\boldsymbol{\eta}$ and $\boldsymbol{\psi}_{c}(A)'\boldsymbol{\eta}$) show the inverse patterns. Notice that the posterior means of the shared random effects terms $\boldsymbol{\psi}_{d\eta}(A)'\boldsymbol{\eta}$ and $\boldsymbol{\psi}_{c\eta}(A)'\boldsymbol{\eta}$ do not appear to have a linear relationship.

Also, in Table 3 we provide the posterior mean and credible intervals for the covariate effects. With 15–25-year-olds being the baseline group, there is a clear effect of age, as the risk of mortality appears to be first decreasing and then increasing with age (with credible intervals for all age-group except 35–44 years being away from zero). The county level mortality for black/African Americans does not appear to be noticeably different from the county level mortality for whites (their posterior means are similar in value), but, there are strong posterior evidence that both of these groups have higher mortality than Asian and Pacific Islanders (baseline group). The effect of gender suggests that females appear to have lower mortality, but the credible intervals suggest that the gender difference does not have strong posterior support (i.e., zero is in the credible interval). These results are similar to existing studies that

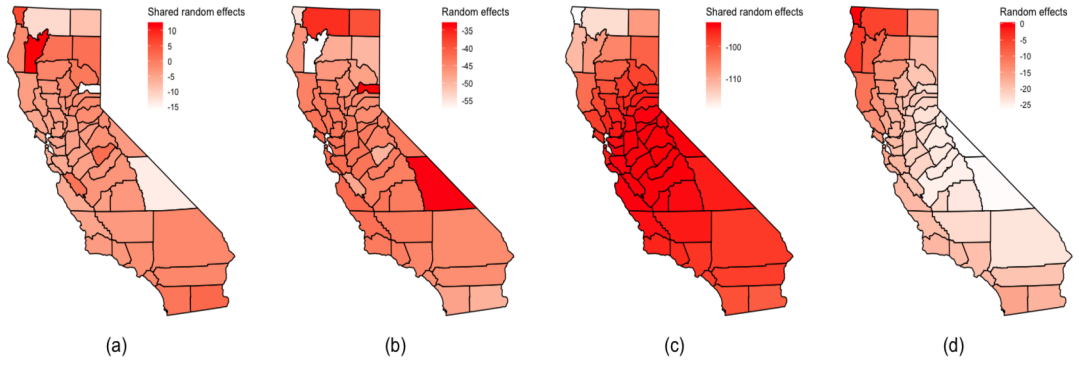

FIG. 6. *This figure shows the posterior means of* $\boldsymbol{\psi}_{d\eta}(A)'\boldsymbol{\eta}$ *(Panel a),* $\boldsymbol{\psi}_{d}(A)'\boldsymbol{\eta}_{d}$ *(Panel b),* $\boldsymbol{\psi}_{c\eta}(A)'\boldsymbol{\eta}$ *(Panel c),* $\boldsymbol{\psi}_{c}(A)'\boldsymbol{\eta}_{c}$ *(Panel d) over counties in California.*

TABLE 3
Posterior Mean and 95% credible intervals for the covariate effects of mortality

	Posterior Mean	2.5%	97.5%
Intercept	−4.9196	−8.1683	1.9281
Female	−0.0011	−0.0122	0.0097
Black/African American	0.1336	0.0565	0.2381
White	0.2617	0.1870	0.3655
25–34 years	−0.2516	−0.4171	−0.0505
35–44 years	−0.0683	−0.1761	0.0875
45–54 years	0.1728	0.0761	0.3247
55–64 years	0.3897	0.2936	0.5407
65–74 years	0.5021	0.4069	0.6546
75–84 years	0.5540	0.4588	0.7063
85+ years	0.5380	0.4425	0.6909

focus on measures of air quality vs. mortality after adjusting for different populations (e.g., see Knuiman et al. (1999), Meng and Lu (2007), among others).

In Figure 7 we plot LICA values for the WAP model to better understand the second-order relationships between PM2.5 and mortality. Most of the counties have LICA values near one, suggesting that there are not strong restrictions on the cross-response-type restrictions. However, there are several counties (e.g., north/northeast California) where the LICA values are small. This suggests that the cross-correlations between mortality and PM2.5 in regions in north/northeast California is small. Notice that the posterior mean of $\boldsymbol{\psi}_{c\eta}(A)'\boldsymbol{\eta}$ was the largest in the north, where LICA also suggests more restrictions on the cross-response dependence.

5. Discussion. In this article we are motivated by data made available by the CDC. In particular, monitoring PM2.5 is important because it helps to assess public health and provides an avenue to do the clinical inference related to mortality. These variables are known to be dependent (Laden et al. (2000), Schwartz and Neas (2000), Valavanidis, Fiotakis and Vlachogianni (2008)). However, several existing methods to model such data assume a linear-type relationship (e.g., log-linear), do not model the variability of both variables, and do not allow for different missing data patterns among the variables. As a result, we developed the WAP to leverage multitype dependence to improve estimation of PM2.5. We introduce the joint Weibull and Poisson (WAP) model which is a framework that can be used to model high-

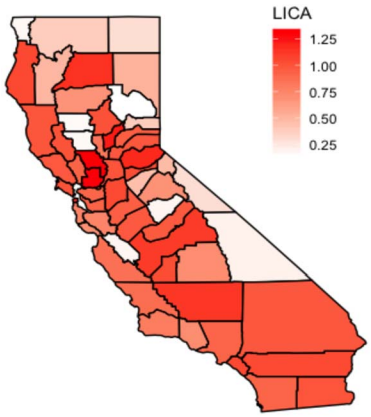

FIG. 7. *The LICA map over counties in California from the WAP model.*

dimensional continuous and count-valued (or multitype) responses. Most of the work modeling multitype responses are in machine learning and nonparametric settings, and hence, our WAP adds to an important growing literature within the Bayesian analysis literature. Another important contribution is that we use the multivariate log-gamma distribution as a conjugate prior. This allows the WAP to be easily implemented, using a collapsed Gibbs sampler, which avoids complicated tuning and other approaches used in other standard Bayesian algorithms. Using a reduced rank set of basis functions, our WAP can be applied to high-dimensional datasets with less cost than full-rank methods.

The WAP addresses four limitations of existing analyses of mortality incidence and PM2.5: (1) much of the current methodology applies a type of regression model and treats one of the variables as a fixed covariate. However, variability is present in both variables and should be included in the model; (2) we allow for possibly nonlinear relationships between mortality and PM2.5; (3) we are able to leverage dependence in both mortality and PM2.5, and allow both variables to have different missing data patterns, and (4) the pressing need to develop precise predictions of both PM2.5 and mortality suggests that developments are needed to handle large datasets. This motivates the use of the MLG (to improve mixing and have a larger effective sample size) and dimension reduction.

In addition to these important developments in using WAP to analyze PM2.5 and mortality incidences, there are several worthwhile methodological contributions. In particular, we have provided the following five general methodological contributions: (1) A statistical model that allows for and leverages cross-dependence between Weibull data and Poisson data, in addition to spatial dependence, (2) a statistical model that simultaneously incorporates the variability associated with Poisson and Weibull data, (3) extension of the use of the multivariate log-gamma (MLG) distribution to jointly model Weibull responses—a model that simultaneously allow for the variability of PM2.5, mortality, cross-dependence, and the high-dimensional nature of the datasets, (4) an efficiency ratio result for joint Weibull and Poisson spatial data, and (5) the introduction of an exploratory analysis tool to assess the amount of cross-response type spatial dependence that is present at each location (i.e., the LICA statistic). These results show that the cross-correlation between Weibull and Poisson data are more flexibly modeled using the multivariate log-gamma distribution, instead of the multivariate normal distribution.

A key feature of our model is the use of spatial basis function expansions. One important consideration when using spatial basis function expansions is the presence of possible spatial confounding. In general, one can check to see if the covariate matrix and basis function matrix are perfectly collinear by finding the reduced row echelon form of the covariate matrix concatenated with the basis matrix (e.g., see Sengupta and Cressie (2013)). In our analysis the basis matrix is not perfectly collinear with the covariate matrix, and consequently, confounding is mitigated. There are basis function matrices that more directly avoid spatial confounding (e.g., see Bradley, Holan and Wikle (2015), Griffith and Tiefelsdorf (2007), Hughes and Haran (2013), among others). We considered a version of these basis set in our analysis of PM2.5 and mortality and found smaller values of information criteria using bisquare basis functions.

In the simulations study we illustrate the high-predictive performance of WAP relative to univariate models and a joint bivariate LGP model. In the Supplementary Material we also provide several sensitivities of WAP to several different factors, including the choice of the number of basis functions, the efficiency ratio, and the proportion of zero Poisson counts. We generate data that is different from our model, and the WAP was able to accurately estimate this signal, even though the data were not generated from WAP. The sensitivity study suggested that the performance of WAP will not be impacted largely by the proportion of zeros, similar to that of our dataset, but the results show a potential risk of decrease in performance when faced with a dataset with a larger proportion of zeros. This result is expected

since there is a rather large literature on zero-inflated Poisson models which are motivated by similar empirical results. Consequently a zero-inflated WAP is an important topic of future research.

Funding. Debajyoti Sinha and Zhixing Xu's research was partially funded by Pfeiffer Cancer Research Foundation, Grant R03CA205018 from the National Cancer Institute, and Hobbs Foundation. Jonathan R. Bradley's research was partially supported by the U.S. National Science Foundation (NSF) under NSF grant SES-1853099 and the National Institutes of Health (NIH) under grant 1R03AG070669-01.

SUPPLEMENTARY MATERIAL

Supplement to "Latent multivariate log-gamma models for high-dimensional Multi-Type responses with application to daily fine particulate matter and mortality counts" (DOI: 10.1214/22-AOAS1664SUPP; .pdf). Supplementary Appendix A: The Conditional MLG Distribution. A review of the conditional MLG Distribution. Supplementary Appendix B: Data Augmentation for conditional MLG Random Vectors. Additional details on the data augmentation strategy used for computational purposes. Supplementary Appendix C: Uni-Type Model for Weibull Responses. A description of a univariate spatio-temporal model for Weibull distributed data. Supplementary Appendix D: WAP Model Prior Specifications and the Derivation of the Full-Conditional Distributions. Additional model details on prior specifications of the WAP along with the derivations of the full conditional distributions for the collapsed Gibbs sampler. Supplementary Appendix E: Proof of Proposition 1. Proof of Proposition 1. Supplementary Appendix F: Additional sensitivity simulation studies. Several sensitivity studies on various specifications/settings to implement WAP. Supplementary Appendix G: Priors for LGP in Section 3.2. The specification of each prior used in Section 3.2. Supplementary Appendix H: Additional simulation Results. Additional table with metrics comparing WAP to LMC. Supplementary Appendix I: MCMC Diagnostics for Section 4.2. Additional details on MCMC diagnostics.

REFERENCES

ALDAZ, J. M. (2009). Self-improvement of the inequality between arithmetic and geometric means. *J. Math. Inequal.* **3** 213–216. MR2542299 https://doi.org/10.7153/jmi-03-21

ANDERSON, J. O., THUNDIYIL, J. G. and STOLBACH, A. (2012). Clearing the air: A review of the effects of particulate matter air pollution on human health. *J. Med. Toxicol.* **8** 166–175.

ANSELIN, L. (1995). Local indicators of spatial association. *Geogr. Anal.* **27** 93–115.

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2015). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. MR3362184

BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc. Ser. B* **36** 192–236. MR0373208

BESAG, J. (1986). On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B* **48** 259–302. MR0876840

BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–59. MR1105822 https://doi.org/10.1007/BF00116466

BRADLEY, J. R. (2021). An approach to incorporate subsampling into a generic Bayesian hierarchical model. *J. Comput. Graph. Statist.* **30** 889–905. MR4356593 https://doi.org/10.1080/10618600.2021.1923518

BRADLEY, J. R., CRESSIE, N. and SHI, T. (2015). Comparing and selecting spatial predictors using local criteria. *TEST* **24** 1–28. MR3314567 https://doi.org/10.1007/s11749-014-0415-1

BRADLEY, J. R., CRESSIE, N. and SHI, T. (2016). A comparison of spatial predictors when datasets could be very large. *Stat. Surv.* **10** 100–131. MR3527662 https://doi.org/10.1214/16-SS115

BRADLEY, J. R., HOLAN, S. H. and WIKLE, C. K. (2015). Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. *Ann. Appl. Stat.* **9** 1761–1791. MR3456353 https://doi.org/10.1214/15-AOAS862

BRADLEY, J. R., HOLAN, S. H. and WIKLE, C. K. (2018). Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion). *Bayesian Anal.* **13** 253–310. MR3773410 https://doi.org/10.1214/17-BA1069

BRADLEY, J. R., HOLAN, S. H. and WIKLE, C. K. (2020). Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. *J. Amer. Statist. Assoc.* **115** 2037–2052. MR4189775 https://doi.org/10.1080/01621459.2019.1677471

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2016). Bayesian spatial change of support for count-valued survey data with application to the American community survey. *J. Amer. Statist. Assoc.* **111** 472–487. MR3538680 https://doi.org/10.1080/01621459.2015.1117471

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2017). Regionalization of multiscale spatial processes by using a criterion for spatial aggregation error. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 815–832. MR3641409 https://doi.org/10.1111/rssb.12179

BRADLEY, J. R., WIKLE, C. K. and HOLAN, S. H. (2020). Hierarchical models for spatial data with errors that are correlated with the latent process. *Statist. Sinica* **30** 81–109. MR4285486 https://doi.org/10.5705/ss.202016.0230

BROOK, R. D., RAJAGOPALAN, S., POPE III, C. A., BROOK, J. R., BHATNAGAR, A., DIEZ-ROUX, A. V., HOLGUIN, F., HONG, Y., LUEPKER, R. V. et al. (2010). Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association. *Circulation* **121** 2331–2378.

BURNETTA, R., HONG, C., SZYSZKOWICZA, M., FANN, N., HUBBELLD, B., POPE, C. A., APTEF, J. S., BRAUERG, M., COHENH, A. et al. (2018). Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci. USA* **115** 9592–9597.

CARROLL, R. J., GAIL, M. H. and LUBIN, J. H. (1993). Case-control studies with errors in covariates. *J. Amer. Statist. Assoc.* **88** 185–199. MR1212486

CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. and CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models*: *A Modern Perspective*, 2nd ed. *Monographs on Statistics and Applied Probability* **105**. CRC Press/CRC, Boca Raton, FL. MR2243417 https://doi.org/10.1201/9781420010138

CHAKRABORTY, A. and PANARETOS, V. M. (2017). Regression with genuinely functional errors-in-covariates. ArXiv preprint. Available at arXiv:1712.04290.

CHIB, S. and GREENBERG, E. (1995). Understanding the Metropolis-Hastings algorithm. *Amer. Statist.* **49** 327–335.

CHOI, J., FUENTES, M. and REICH, B. J. (2009). Spatial-temporal association between fine particulate matter and daily mortality. *Comput. Statist. Data Anal.* **53** 2989–3000. MR2667605 https://doi.org/10.1016/j.csda.2008.05.018

CHRISTENSEN, W. F. and AMEMIYA, Y. (2002). Latent variable analysis of multivariate spatial data. *J. Amer. Statist. Assoc.* **97** 302–317. MR1947288 https://doi.org/10.1198/016214502753479437

CLARKE, J. S., NEMERGUT, D., SEYEDNASROLLAH, B., TURNER, P. and ZHANG, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecol. Monogr.* **87** 34–56.

CRESSIE, N. and JOHANNESSON, G. (2006). Spatial prediction for massive datasets. In *Mastering the Data Explosion in the Earth and Environmental Sciences*: *Australian Academy of Science*, *Elizabeth and Frederick White Conference* 1–11.

CRESSIE, N. and JOHANNESSON, G. (2008). Fixed rank Kriging for very large spatial data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 209–226. MR2412639 https://doi.org/10.1111/j.1467-9868.2007.00633.x

DATTA, A., BANERJEE, S., FINLEY, A. O., HAMM, N. A. S. and SCHAAP, M. (2016). Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *Ann. Appl. Stat.* **10** 1286–1316. MR3553225 https://doi.org/10.1214/16-AOAS931

DE OLIVEIRA, V. (2003). A note on the correlation structure of transformed Gaussian random fields. *Aust. N. Z. J. Stat.* **45** 353–366. MR1999517 https://doi.org/10.1111/1467-842X.00289

DE OLIVEIRA, V. (2013). Hierarchical Poisson models for spatial count data. *J. Multivariate Anal.* **122** 393–408. MR3189330 https://doi.org/10.1016/j.jmva.2013.08.015

DEPAOLI, S., CLIFTON, J. P. and COBB, P. R. (2016). Just another Gibbs sampler (JAGS) flexible software for MCMC implementation. *J. Educ. Behav. Stat.* **41** 628–649.

DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist.* **7** 269–281. MR0520238

DOCKERY, D. W., POPE, C. A., XU, X., SPENGLER, J. D., WARE, J. H., FAY, M., FERRIS JR, G. B. and SPEIZER, F. E. (1993). An association between air pollution and mortality in six US cities. *N. Engl. J. Med.* **329** 1753–1759.

DOMINICI, F., PENG, R. D., BELL, M. L., PHAM, L., MCDERMOTT, A., ZEGER, S. L. and SAMET, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *J. Am. Med. Assoc.* **295** 1127–1134.

FINLEY, A. O., SANG, H., BANERJEE, S. and GELFAND, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Comput. Statist. Data Anal.* **53** 2873–2884. MR2667597 https://doi.org/10.1016/j.csda.2008.09.008

FRANKLIN, M., ZEKA, A. and SCHWARTZ, J. (2007). Association between PM2.5 and all-cause and specific-cause mortality in 27 US communities. *J. Expo. Sci. Environ. Epidemiol.* **17** 279–287. https://doi.org/10.1038/sj.jes.7500530

FUENTES, M., SONG, H.-R., GHOSH, S. K., HOLLAND, D. M. and DAVIS, J. M. (2006). Spatial association between speciated fine particles and mortality. *Biometrics* **62** 855–863. MR2247215 https://doi.org/10.1111/j.1541-0420.2006.00526.x

GAO, L., DATTA, A. and BANERJEE, S. (2022). Hierarchical multivariate directed acyclic graph autoregressive models for spatial diseases mapping. *Stat. Med.* **41** 3057–3075. MR4444886 https://doi.org/10.1002/sim.9404

GELFAND, A. E. and SCHLIEP, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spat. Stat.* **18** 86–104. MR3573271 https://doi.org/10.1016/j.spasta.2016.03.006

GELMAN, A., MENG, X.-L. and STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6** 733–807. MR1422404

GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.

GELMAN, A., CARLIN, J. B., STERN, H. S., VEHTARI, A. and RUBIN, D. B. (2013). *Bayesian Data Analysis*. CRC Press/CRC, Boca Raton, FL.

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 https://doi.org/10.1198/016214506000001437

GOTWAY, C. A. and YOUNG, L. J. (2002). Combining incompatible spatial data. *J. Amer. Statist. Assoc.* **97** 632–648. MR1951636 https://doi.org/10.1198/016214502760047140

GRIFFITH, D. A. and TIEFELSDORF, M. (2007). Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environ. Plan. A.* **39** 1193–1221.

HEATON, M. J., DATTA, A., FINLEY, A. O., FURRER, R., GUINNESS, J., GUHANIYOGI, R., GERBER, F., GRAMACY, R. B., HAMMERLING, D. et al. (2018). A case study competition among methods for analyzing large spatial data. *J. Agric. Biol. Environ. Stat.* **24** 398–425.

HOEK, G., KRISHNAN, R. M., BEELEN, R., PETERS, A., OSTRO, B., BRUNEKREEF, B. and KAUFMAN, J. D. (2013). Long-term air pollution exposure and cardio-respiratory mortality: A review. *Environ. Health* **12** 43.

HU, G. and BRADLEY, J. (2018). A Bayesian spatial-temporal model with latent multivariate log-gamma random effects with application to earthquake magnitudes. *Stat* **7** e179. MR3796093 https://doi.org/10.1002/sta4.179

HUGHES, J. and HARAN, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 139–159. MR3008275 https://doi.org/10.1111/j.1467-9868.2012.01041.x

JORDAN, A., KRÜGER, F. and LERCH, S. (2017). Evaluating probabilistic forecasts with scoringRules. ArXiv preprint. Available at arXiv:1709.04743.

KAMPA, M. and CASTANAS, E. (2008). Human health effects of air pollution. *Environ. Pollut.* **151** 362–367.

KATZFUSS, M. and CRESSIE, N. (2011). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *J. Time Series Anal.* **32** 430–446. MR2841794 https://doi.org/10.1111/j.1467-9892.2011.00732.x

KATZFUSS, M. and GUINNESS, J. (2021). A general framework for Vecchia approximations of Gaussian processes. *Statist. Sci.* **36** 124–141. MR4194207 https://doi.org/10.1214/19-STS755

KATZFUSS, M., GUINNESS, J., GONG, W. and ZILBER, D. (2020). Vecchia approximations of Gaussian-process predictions. *J. Agric. Biol. Environ. Stat.* **25** 383–414. MR4139037 https://doi.org/10.1007/s13253-020-00401-7

KIM, J. S. and YUM, B.-J. (2008). Selection between Weibull and lognormal distributions: A comparative simulation study. *Comput. Statist. Data Anal.* **53** 477–485. MR2649102 https://doi.org/10.1016/j.csda.2008.08.012

KNUIMAN, M. W., JAMES, A. L., DIVITINI, M. L., RYAN, G., BARTHOLOMEW, H. C. and MUSK, A. W. (1999). Lung function, respiratory symptoms, and mortality: Results from the busselton health study. *Ann. Epidemiol.* **9** 297–306. https://doi.org/10.1016/s1047-2797(98)00066-0

KRAINSKI, E., GÓMEZ-RUBIO, V., BAKKA, H., LENZI, A., CASTRO-CAMILO, D., SIMPSON, D., LINDGREN, F. and RUE, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. CRC Press/CRC, Boca Raton.

KRISTENSEN, K., NIELSEN, A., BERG, C. W., SKAUG, H. and BELL, B. (2016). TMB: Automatic differentiation and Laplace approximation. *J. Stat. Softw.* **60** 1–21.

LADEN, F., NEAS, L. M., DOCKERY, D. W. and SCHWARTZ, J. (2000). Association of fine particulate matter from different sources with daily mortality in six US cities. *Environ. Health Perspect.* **108** 941.

LEITER, U. and GARBE, C. (2008). Epidemiology of melanoma and nonmelanoma skin cancer the role of sunlight. In *Sunlight, Vitamin D and Skin Cancer* 89–103. Springer, New York, NY.

MARTINO, S. and RIEBLER, A. (2019). Integrated Nested Laplace Approximations (INLA). ArXiv preprint. Available at arXiv:1907.01248.

MENG, X.-L. (1994). Posterior predictive p-values. *Ann. Statist.* **22** 1142–1160. MR1311969 https://doi.org/10.1214/aos/1176325622

MENG, Z. and LU, B. (2007). Dust events as a risk factor for daily hospitalization for respiratory and cardiovascular diseases in Minqin, China. *Atmos. Environ.* **41** 7048–7058.

MILLER, K. A., SISCOVICK, D. S., SHEPPARD, L., SHEPHERD, K., SULLIVAN, J. H., ANDERSON, G. L. and KAUFMAN, D. J. (2007). Long-term exposure to air pollution and incidence of cardiovascular events in women. *N. Engl. J. Med.* **356** 447–458.

MORAN, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika* **37** 17–23. MR0035933 https://doi.org/10.1093/biomet/37.1-2.17

MUGGLIN, A. S., CARLIN, B. P. and GELFAND (2000). Fully model-based approaches for spatially misaligned data. *J. Amer. Statist. Assoc.* **95** 877–887.

NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31** 705–767. MR1994729 https://doi.org/10.1214/aos/1056562461

NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 113–162. CRC Press, Boca Raton, FL. MR2858447

PERUZZI, M., BANERJEE, S. and FINLEY, A. O. (2022). Highly Scalable Bayesian Geostatistical Modeling via Meshed Gaussian Processes on Partitioned Domains. *J. Amer. Statist. Assoc.* **117** 969–982. MR4436326 https://doi.org/10.1080/01621459.2020.1833889

QUICK, H., WALLER, L. A. and CASPER, M. (2018). A multivariate space-time model for analysing county level heart disease death rates by race and sex. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **67** 291–304. MR3758767 https://doi.org/10.1111/rssc.12215

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

SCHLIEP, E. M. and HOETING, J. A. (2013). Multilevel latent Gaussian process model for mixed discrete and continuous multivariate response data. *J. Agric. Biol. Environ. Stat.* **18** 492–513. MR3142597 https://doi.org/10.1007/s13253-013-0136-z

SCHWARTZ, J. and NEAS, L. M. (2000). Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren. *Epidemiology* **11** 6–10. https://doi.org/10.1097/00001648-200001000-00004

SELLERS, K. F. and RAIM, A. (2016). A flexible zero-inflated model to address data dispersion. *Comput. Statist. Data Anal.* **99** 68–80. MR3473082 https://doi.org/10.1016/j.csda.2016.01.007

SENGUPTA, A. and CRESSIE, N. (2013). Hierarchical statistical modeling of big spatial datasets using the exponential family of distributions. *Spat. Stat.* **4** 14–44.

SHI, T. and CRESSIE, N. (2007). Global statistical analysis of MISR aerosol data: A massive data product from NASA's Terra satellite. *Environmetrics* **18** 665–680. MR2408937 https://doi.org/10.1002/env.864

STEIN, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spat. Stat.* **8** 1–19. MR3326818 https://doi.org/10.1016/j.spasta.2013.06.003

TURNER, M. C., KREWSKI, D., POPE III, C. A., CHEN, Y., GAPSTUR, S. M. and THUN, M. J. (2011). Long-term ambient fine particulate matter air pollution and lung cancer in a large cohort of never-smokers. *Am. J. Respir. Crit. Care Med.* **184** 1374–1381.

VALAVANIDIS, A., FIOTAKIS, K. and VLACHOGIANNI, T. (2008). Airborne particulate matter and human health: Toxicological assessment and importance of size and composition of particles for oxidative damage and carcinogenic mechanisms. *J. Environ. Sci. Health, Part C* **26** 339–362.

WAHBA, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia, PA. MR1045442 https://doi.org/10.1137/1.9781611970128

WATANABE, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11** 3571–3594. MR2756194

WIKLE, C. K. (2010). Low-rank representations for spatial processes. In *Handbook of Spatial Statistics. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 107–118. CRC Press, Boca Raton, FL. MR2730946 https://doi.org/10.1201/9781420072884-c8

WIKLE, C. K. and CRESSIE, N. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika* **86** 815–829. MR1741979 https://doi.org/10.1093/biomet/86.4.815

WIKLE, C. K. and HOOTEN, M. B. (2010). A general science-based framework for dynamical spatio-temporal models. *TEST* **19** 417–451. MR2745992 https://doi.org/10.1007/s11749-010-0209-z

XU, Z., BRADLEY, J. R. and SINHA, D. (2023). Supplement to "Latent multivariate log-gamma models for high-dimensional multitype responses with application to daily fine particulate matter and mortality counts." https://doi.org/10.1214/22-AOAS1664SUPP

XU, M., GUO, Y., ZHANG, Y., WESTERDAHL, D., MO, Y., LIANG, F. and PAN, X. (2014). Spatiotemporal analysis of particulate air pollution and ischemic heart disease mortality in Beijing, China. *Environ. Health* **13** 109.

XU, Q., LI, X., WANG, S., WANG, C., HUANG, F., GAO, Q., WU, L., TAO, L., GUO, J. et al. (2016). Fine particulate air pollution and hospital emergency room visits for respiratory disease in urban areas in Beijing, China, in 2013. *PLoS ONE* **11** e0153099.

ZHANG, L., TANG, W. and BANERJEE, S. (2021). Fixed-domain asymptotics under Vecchia's approximation of spatial process likelihoods. ArXiv preprint. Available at arXiv:2101.08861.

ZHANG, F., LIU, X., ZHOU, L., YU, Y., WANG, L., LU, J., WANG, W. and KRAFFT, T. (2016). Spatiotemporal patterns of particulate matter (PM) and associations between PM and mortality in Shenzhen, China. *BMC Public Health* **16** 215.

ZIGLER, C. M., DOMINICI, F. and WANG, Y. (2012). Estimating causal effects of air quality regulations using principal stratification for spatially correlated multivariate intermediate outcomes. *Biostatistics* **13** 289–302.