

# BAYESIAN INFERENCE FOR SPATIALLY-TEMPORALLY MISALIGNED DATA USING PREDICTIVE STACKING

SOUMYAKANTI PAN\* AND SUDIPTO BANERJEE\*

\**Department of Biostatistics, University of California Los Angeles*

**ABSTRACT.** Air pollution remains a major environmental risk factor that is often associated with adverse health outcomes. However, quantifying and evaluating its effects on human health is challenging due to the complex nature of exposure data. Recent technological advances have led to the collection of various indicators of air pollution at increasingly high spatial-temporal resolutions (e.g., daily averages of pollutant levels at spatial locations referenced by latitude-longitude). However, health outcomes are typically aggregated over several spatial-temporal coordinates (e.g., annual prevalence for a county) to comply with survey regulations. This article develops a Bayesian hierarchical model to analyze such spatially-temporally misaligned exposure and health outcome data. We introduce Bayesian predictive stacking, which optimally combines multiple predictive spatial-temporal models and avoids iterative estimation algorithms such as Markov chain Monte Carlo that struggle due to convergence issues inflicted by the presence of weakly identified parameters. We apply our proposed method to study the effects of ozone on asthma in the state of California.

**Keywords.** change of support, modular inference, model combination, weak identifiability

## 1. INTRODUCTION

Spatial and temporal misalignment refers to the setting in which different variables are observed over incompatible spatial supports and/or at asynchronous time points or intervals. To be more specific, spatial misalignment arises when variables are measured at different spatial resolutions, i.e., different sets of locations or areas. For example, a variable might be observed at points (e.g. air quality monitoring stations), which we call *point-referenced* data, while another is aggregated over administrative units (e.g., counties or census tracts), which we call *block* data. Similarly, temporal misalignment occurs when variables are recorded on different time scales.

In this article, we devise a Bayesian hierarchical modeling framework to study the effects of ozone on asthma-related health emergencies among California residents. Several studies collectively underscore the significant impact of exposure to ozone on asthma-related emergency department visits (for e.g., [Gharibi et al., 2019](#); [Nassikas et al., 2020](#)). We obtain data on the monthly average concentration of ozone in California measured by air quality monitoring stations located at different locations throughout the state. Thus, the ozone measurements are spatially point-referenced and temporally aggregated at a monthly scale. However, data on asthma-related emergency department

---

*E-mail address:* [span18@ucla.edu](mailto:span18@ucla.edu), [sudipto@ucla.edu](mailto:sudipto@ucla.edu).

*Date:* May 30, 2025.

visits among Californian residents are reported at the county level and aggregated annually. We refer to such disparity in both spatial and temporal resolutions as spatial-temporal misalignment.

Spatially-temporally misaligned data presents significant challenges for coherent modeling, prediction, and inference. This problem is well known as the *change of support* and *modifiable areal unit* problem (Cressie, 1996; Mugglin et al., 2000; Gelfand et al., 2001; Gotway and and, 2002). Most of the literature focuses on estimating variables at unobserved spatial resolutions or integrating data across varying spatial scales (Banerjee et al., 2014; Zhong and Moraga, 2023). However, efforts to estimate the association between an outcome of interest and a spatially temporally misaligned covariate remain rather limited (Zhu et al., 2003; Cameletti et al., 2019), with existing approaches typically overlooking temporal misalignment. Moreover, a key challenge in analyzing spatial-temporal exposure data is missing observations resulting from intermittent monitoring or data removal due to quality issues such as sensor failures. Traditional approaches often use imputation methods that aim to reconstruct the complete data set (for e.g., Zhu et al., 2003; Quick et al., 2013). Our approach forgoes imputation and works directly with the available irregularly spaced data to produce a more robust framework for spatial-temporal analysis.

We contribute in two novel aspects. First, we propose a modular Bayesian inference framework (Bayarri et al., 2009; Jacob et al., 2017) that regresses an outcome on a spatially-temporally misaligned covariate, based on noisy observations of the latter, ensuring fully model-based propagation of inferential uncertainty. Second, we develop predictive stacking for estimation of such models that analyze spatially-temporally misaligned data, thus representing a methodological advancement over previous work that included spatial Gaussian data (Zhang et al., 2024; Wakayama and Banerjee, 2024; Presicce and Banerjee, 2025), and spatial-temporal non-Gaussian data (Pan et al., 2025). Stacking (Wolpert, 1992; Breiman, 1996; Clyde and Iversen, 2013) is conspicuous in machine learning as an effective alternative (Le and Clarke, 2017; Yao et al., 2018) to traditional Bayesian model averaging (Madigan et al., 1996; Hoeting et al., 1999). The underlying idea in predictive stacking is to optimally assimilate posterior distributions on a grid of candidate values corresponding to intractable and weakly identified hyperparameters, such as the spatial-temporal decay, smoothness, and the noise-to-spatial-temporal variance ratio (Zhang, 2004; Zhang and Zimmerman, 2005; Tang et al., 2021), which impede the convergence of iterative algorithms such as Markov chain Monte Carlo (MCMC). Stacking also differs from quadrature-based approaches, such as INLA (Rue et al., 2009), in that we avoid approximating the posterior distribution of weakly identified parameters. Instead, we average, or “stack” individual posterior distributions, using weights obtained by optimizing a proper scoring rule (Gneiting and Raftery, 2007).

The remainder of the manuscript is structured as follows. Section 2 describes the data set that motivates our methodology. Section 3 introduces our Bayesian hierarchical model for analyzing the spatially-temporally misaligned data and states model assumptions that are critical for posterior inference. Section 4 develops predictive stacking and focuses on developing a computationally efficient algorithm for model estimation. Section 6 describes the detailed analysis of our data set, while Section 7 concludes with a brief discussion.

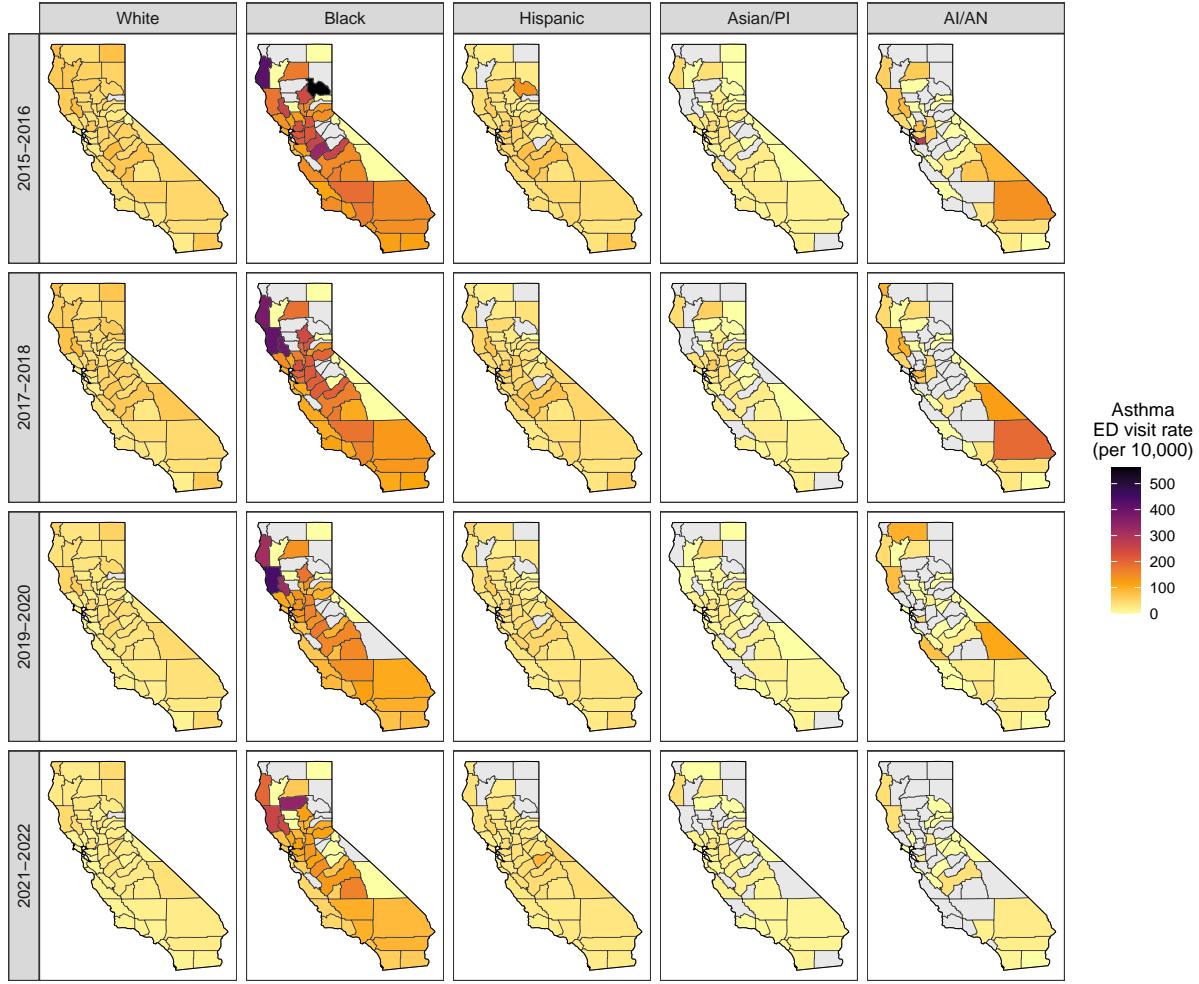


FIGURE 1. Biennial average asthma-related emergency department visit rates (per 10,000) by racial group for each California county from 2015 through 2022. For visualization, rates are averaged over consecutive 2-year periods.

## 2. DATA

The data set on adverse health outcomes comprises annual county-level rates of visits to the emergency department (ED) related to asthma per 10,000 residents of California. These data are obtained from the California Department of Public Health (CDPH) and are derived from the Department of Health Care Access and Information's Emergency Department database, which includes records from all licensed hospitals in the state ([California Breathing Asthma Program, 2024](#)). We analyze data consisting of age-adjusted rates stratified by race/ethnicity (white, black, Hispanic, Asian/Pacific Islander, American Indian/Alaskan Native), derived from annual aggregated counts of asthma-related ED visits from 2015 to 2022 for each of the 58 counties in California. ED visit counts are based only on primary discharge diagnosis codes. The database omits rates based on (1) counts  $< 12$  due to statistical instability, and (2) counts  $\leq 5$  according to the CalHHS Data De-identification Guidelines. This results in approximately 35% missing data that mainly affect the

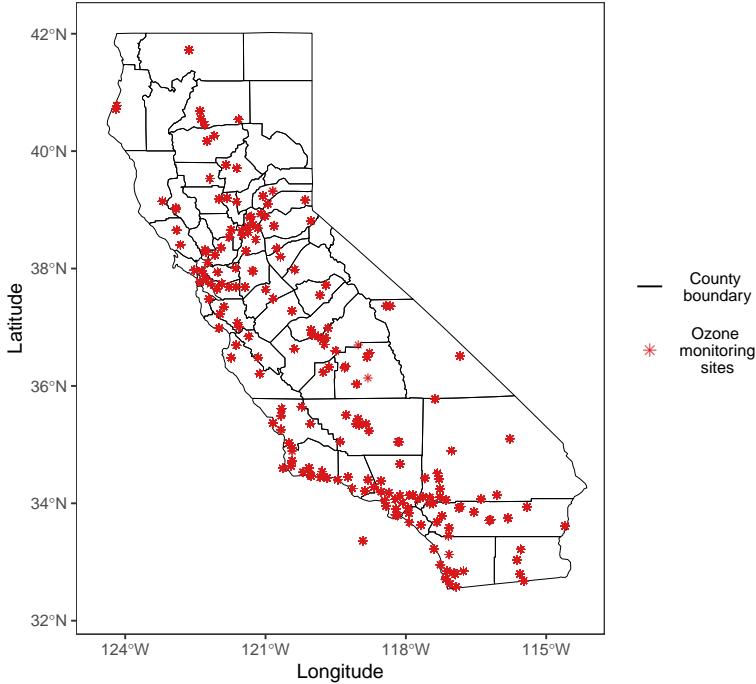


FIGURE 2. County boundaries of California and the geographic locations of 200 ozone monitoring sites active from 2015 to 2022.

Asian/Pacific Islander and American Indian/Alaskan Native groups. Figure 1 maps the aggregated two-year rates of asthma-related emergency department visits observed for each county in California during 2015-2022, revealing a clear pattern of racial disparity, with some counties consistently showing high rates. The age-adjusted rates are highly positive skewed, ranging from 0 to 560.9 per 10,000 residents, with the majority between 15 and 60 (see Figure A3 in the Appendix). Hence we analyze log-transformed rates, as they are more amenable to Gaussian assumptions, that yields analytically tractable posteriors, enabling inference without iterative algorithms.

For exposure data, we extracted hourly measurements of ozone concentration (in parts per million) from the California Air Quality and Meteorological Information System (AQMS) database of the California Air Resources Board for the years 2015-2022. These data were recorded from about 200 air quality monitoring sites in California that were active during this time period. Figure 2 maps the geographic locations of these monitoring sites located in the county boundaries of California, highlighting the heterogeneous spatial distribution of air quality surveillance in counties. The figure also reveals a clear regional clustering of sites near urban and coastal areas in central and southern California, with notable sparsity in the eastern rural inland regions of the state. For our analysis, we aggregate the hourly data into monthly average ozone concentrations. Due to the fact that not all monitoring sites were active throughout the study period and many have missing monthly records, the resulting data set is temporally unbalanced between sites. In total, the data set comprises measurements of ozone concentrations at 15,725 unique spatial-temporal locations, where the temporal component is referenced by monthly intervals rather than exact timestamps. Figure 3 features interpolated spatial surfaces of annual average ozone concentration between 2015

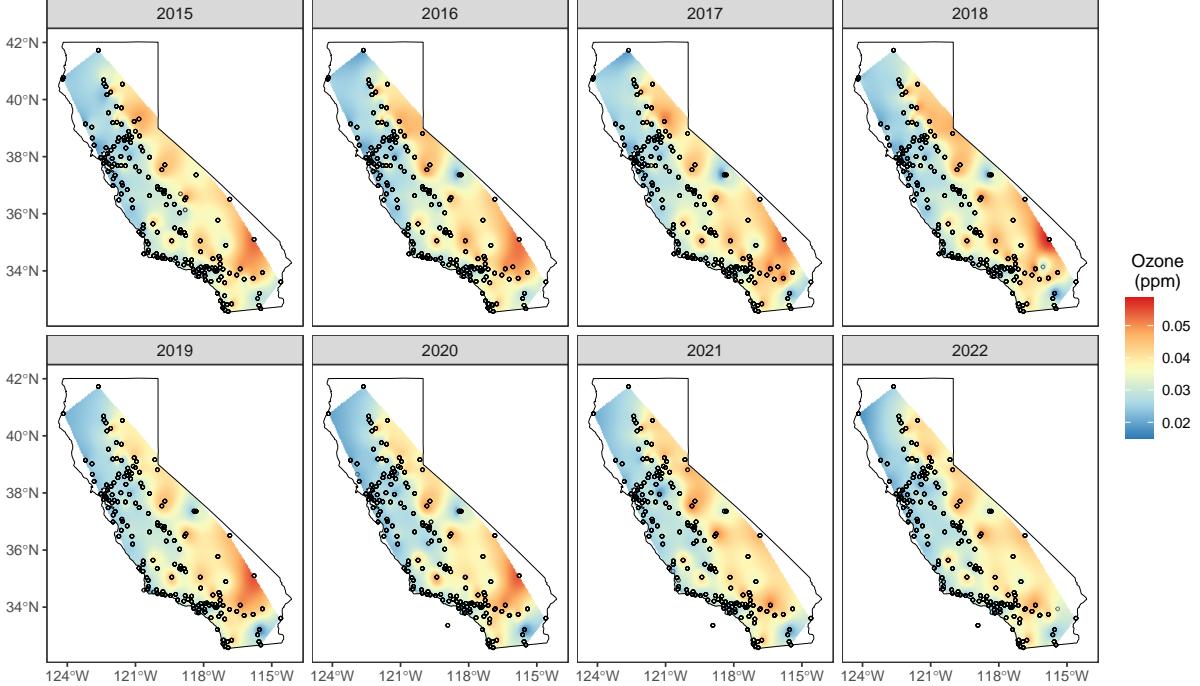


FIGURE 3. Interpolated spatial surface of annual (summed over months) ozone concentration (in parts per million) for California from 2015 to 2022. The geographic coordinates of the air quality monitoring stations are marked by black circles.

and 2022, showing a clear spatial pattern, with the sparsely sampled eastern inland regions recording higher ozone concentrations, demonstrating spatial imbalances in monitoring coverage.

### 3. BAYESIAN SPATIAL-TEMPORAL HIERARCHICAL MODEL

**3.1. Multi-resolution spatial-temporal process.** We consider a spatial-temporal process as an uncountable set of random variables, say  $\{Z(\ell) : \ell \in \mathcal{D}\}$ , which is endowed with a probability law specifying the joint distribution for any finite sample of locations in  $\mathcal{D} = \mathcal{S} \times \mathcal{T}$ , where  $\mathcal{S} \subset \mathbb{R}^2$  and  $\mathcal{T} \subset [0, \infty)$  are space and time domains, respectively, and  $\ell = (s, t)$  is a spacetime coordinate with  $s \in \mathcal{S}$  and  $t \in \mathcal{T}$  (see, e.g., [Gneiting and Guttorp, 2010](#)). Subsequently, we define two new types of space-time coordinates  $\tilde{\ell} = (\tilde{s}, \tilde{I})$  and  $L = (B, I)$ , where  $\tilde{s} \in \mathcal{S}$ ,  $B \subset \mathcal{S}$  denotes a block or a region within  $\mathcal{S}$ , and  $I, \tilde{I} \subset \mathcal{T}$  denote intervals in  $\mathcal{T}$ . We extend the spatial-temporal process  $Z(\ell)$  to these new coordinate systems as  $\{Z(\tilde{\ell}) : \tilde{\ell} \in \mathcal{S} \times \mathfrak{B}(\mathcal{T})\}$  and  $\{Z(L) : L \in \mathfrak{B}(\mathcal{S} \times \mathcal{T})\}$ , where for any set  $A$ ,  $\mathfrak{B}(A)$  denotes the collection of Borel-measurable subsets of  $A$  ([Billingsley, 1995](#)), and

$$Z(\tilde{\ell}) = Z(\tilde{s}, \tilde{I}) = |\tilde{I}|^{-1} \int_{\tilde{I}} Z(\tilde{s}, t) dt, \quad Z(L) = Z(B, I) = (|B||I|)^{-1} \int_I \int_B Z(s, t) ds dt, \quad (1)$$

where  $|B| = \int_B 1 ds$  denotes the area of  $B$ , and  $|I| = \int_I 1 dt$ ,  $|\tilde{I}| = \int_{\tilde{I}} 1 dt$  denote the lengths of the intervals  $I$  and  $\tilde{I}$ , respectively. This implies that the random variable  $Z(\tilde{\ell})$  denotes a realization of the stochastic process obtained by a temporal averaging of  $Z(\ell)$  over the interval  $\tilde{I}$  at location  $\tilde{s}$ . In our context, this corresponds to the monthly average concentration of ozone at the location  $\tilde{s}$ ,

averaged over the month  $\tilde{I}$ . Similarly,  $Z(L)$  is a realization from the stochastic process obtained by averaging  $Z(\ell)$  over a spatial-temporal block  $L$ , which includes averaging spatially over the region  $B$  and averaging temporally over the interval  $I$ . This may represent ozone concentrations in the county defined by the spatial block  $B$ , averaged over the annual interval  $I$ .

A stationary Gaussian process specification for  $Z(\ell)$  enables us to analytically derive a joint distribution for the spatial-temporal process at each resolution. To be specific, suppose  $Z(\ell)$  is a Gaussian process with mean function  $\mu(\ell; \gamma)$  and covariance function  $\sigma^2 C(\ell, \ell'; \phi)$ , where  $\mu(\cdot; \gamma)$  denotes a trend surface with coefficient vector  $\gamma$ ,  $\sigma^2$  denotes the spatial-temporal variance, and  $\phi$  denotes a generic parameter vector characterizing spatial-temporal decay or smoothness. We collectively call  $\phi$  as ‘‘process parameters’’. For any given spatial-temporal coordinates  $\{\tilde{\ell}_j = (\tilde{s}_j, \tilde{I}_j) : j = 1, \dots, N\}$  and  $\{L_k = (B_k, I_k) : k = 1, \dots, K\}$ , define the  $N \times 1$  vector  $Z_{\tilde{\ell}} = (Z(\tilde{\ell}_1), \dots, Z(\tilde{\ell}_N))^{\top}$ , and the  $K \times 1$  vector  $Z_L = (Z(L_1), \dots, Z(L_K))^{\top}$ . Then, we have

$$p\left(\begin{pmatrix} Z_{\tilde{\ell}} \\ Z_L \end{pmatrix} \middle| \gamma, \sigma^2, \phi\right) = \mathbb{N}\left(\begin{pmatrix} Z_{\tilde{\ell}} \\ Z_L \end{pmatrix} \middle| \begin{pmatrix} \mu_{\tilde{\ell}}(\gamma) \\ \mu_L(\gamma) \end{pmatrix}, \sigma^2 \begin{bmatrix} C_{\tilde{\ell}}(\phi) & C_{\tilde{\ell}, L}(\phi) \\ C_{\tilde{\ell}, L}(\phi)^{\top} & C_L(\phi) \end{bmatrix}\right), \quad (2)$$

where

$$\begin{aligned} (\mu_{\tilde{\ell}}(\gamma))_j &= |\tilde{I}_j|^{-1} \int_{\tilde{I}_j} \mu((\tilde{s}_j, t); \gamma) dt, \quad (\mu_L(\gamma))_k = (|I_k| |B_k|)^{-1} \int_{L_k} \mu(\ell; \gamma) d\ell, \\ (C_{\tilde{\ell}}(\phi))_{j,j'} &= (|\tilde{I}_j| |\tilde{I}_{j'}|)^{-1} \int_{\tilde{I}_{j'}} \int_{\tilde{I}_j} C((\tilde{s}_j, t), (\tilde{s}_{j'}, t'); \phi) dt dt', \\ (C_{\tilde{\ell}, L}(\phi))_{j,k} &= (|\tilde{I}_j| |I_k| |B_k|)^{-1} \int_{\tilde{I}_j} \int_{I_k} \int_{B_k} C((\tilde{s}_j, t'), (s, t); \phi) ds dt dt', \\ (C_L(\phi))_{k,k'} &= (|L_{k'}| |L_k|)^{-1} \int_{L_{k'}} \int_{L_k} C(\ell, \ell'; \phi) d\ell d\ell', \end{aligned} \quad (3)$$

for  $j, j' = 1, \dots, N$ , and  $k, k' = 1, \dots, K$ , and  $|L_k| = |B_k| |I_k|$  for each  $k$ . The joint distribution (2) provides a unified framework that connects the process across the two spatial-temporal resolutions, enabling tractable posterior predictive inference at spatial-temporal blocks  $L_k$  for each  $k$ , conditional on observed realizations at point-referenced temporal blocks  $\tilde{\ell}_j, j = 1, \dots, N$ .

**3.2. Conjugate Bayesian hierarchical model.** Let  $\mathcal{L} = \{L_k : k = 1, \dots, K\}$  be a fixed set of  $K$  spatial-temporal blocks in  $\mathfrak{B}(\mathcal{S} \times \mathcal{T})$ , where  $L_k$  is of the form  $(B_k, I_k)$  with  $B_k$  being a block or region in  $\mathcal{S}$  and  $I_k$  an interval in  $\mathcal{T}$ , for each  $k$ . Let  $Y(\mathcal{L}) = (Y(L_1), \dots, Y(L_K))^{\top}$ , which we simply denote by  $Y$ , be the  $K \times 1$  vector of outcomes observed at  $\mathcal{L}$ . The key challenge of our model is that we do not observe the covariates that are assumed to explain the outcome at the same spatial-temporal resolution at which  $Y$  is observed. Suppose  $\tilde{\mathcal{L}} = \{\tilde{\ell}_j : j = 1, \dots, N\}$ , where  $\tilde{\ell}_j$  is of the form  $(\tilde{s}_j, \tilde{I}_j)$  for each  $j$ , be the  $N$  spatial-temporal coordinates in  $\mathcal{S} \times \mathfrak{B}(\mathcal{T})$ , at which we observe the covariate  $X(\tilde{\mathcal{L}}) = (X(\tilde{\ell}_1), \dots, X(\tilde{\ell}_N))^{\top}$ , which we simply write as the  $N \times 1$  vector  $X$ . We further assume that  $X(\tilde{\mathcal{L}})$  are noisy measurements of a latent spatial-temporal process on  $\mathcal{S} \times \mathfrak{B}(\mathcal{T})$  derived from a parent process  $Z(\ell)$  on  $\mathcal{D}$  following the stochastic integral (1). This setup is particularly motivated by the asthma and ozone datasets described in Section 2, where the

outcome is measured at the county level and aggregated annually, while the covariate is spatially referenced to points, but aggregated on a monthly scale.

We build a hierarchical model by jointly modeling both the outcome and the covariate as conditionally dependent on a shared latent spatial-temporal process,

$$\begin{aligned}
Y(L_k) &= w(L_k)^\top \beta_1 + \beta_2 Z(L_k) + \epsilon_k, \quad \epsilon_k \stackrel{\text{ind}}{\sim} \mathbf{N}(0, (|B_k||I_k|)^{-1}\tau^2), k = 1, \dots, K, \\
(\beta_1^\top, \beta_2)^\top | \tau^2 &\sim \mathbf{N}(\mu_\beta, \tau^2 V_\beta), \quad \tau^2 \sim \mathbf{IG}(a_\tau, b_\tau), \\
X(\tilde{\ell}_j) &= Z(\tilde{\ell}_j) + e_j, \quad e_j \stackrel{\text{ind}}{\sim} \mathbf{N}(0, |\tilde{I}_j|^{-1}\delta^2\sigma^2), \quad j = 1, \dots, N, \\
Z(\ell) | \gamma, \sigma^2, \phi &\sim \mathbf{GP}(\mu(\ell; \gamma), \sigma^2 C(\cdot, \cdot; \phi)), \quad \mu(\ell; \gamma) = \psi(\ell)^\top \gamma, \\
\gamma | \sigma^2 &\sim \mathbf{N}(\mu_\gamma, \sigma^2 V_\gamma), \quad \sigma^2 \sim \mathbf{IG}(a_\sigma, b_\sigma),
\end{aligned} \tag{4}$$

where  $w(L_k) = (w_1(L_k), \dots, w_p(L_k))^\top$  denotes a  $p \times 1$  vector of predictors,  $\beta_1$  is the corresponding  $p \times 1$  vector of fixed effects, and,  $\beta_2$  is a scalar and denotes the slope corresponding to the covariate  $Z(L_k)$ , which denotes the latent spatial-temporal process at  $L_k$ , for each  $k$ . Instead of  $Z(L_K)$ , we observe  $X(\tilde{\ell}_j)$ , which are noisy observations of  $Z(\tilde{\ell}_j)$ , which denotes the latent process at a different spatial-temporal resolution from  $Z(L_k)$ . Note that both  $Z(L_k)$  for each  $k$ , and  $Z(\tilde{\ell}_j)$  for each  $j$ , are completely unobserved, and their joint probability law is specified by the assumption of a Gaussian process (GP)  $Z(\ell)$  on  $\mathcal{D}$ , following (2). The mean function of the GP is characterized by a  $r \times 1$  basis  $\psi(\ell) = (\psi_1(\ell), \dots, \psi_r(\ell))^\top$ , and its corresponding  $r \times 1$  slope vector  $\gamma$ . The purpose of  $\psi(\ell)$  is to model a global trend of the spatial-temporal process which may capture an overall mean, seasonal variations, and so on. The term  $\epsilon_k$  denotes white noise defined only at the coordinates in  $\mathcal{L}$ , and captures the heteroskedasticity arising from inhomogeneity in the volumes of the spatial-temporal blocks at which the outcome is observed. Similarly,  $e_j$  captures the independent measurement error for  $X(\tilde{\ell}_j)$  for each  $j$ , and are defined only at  $\tilde{\mathcal{L}}$ . It is crucial to note that, unlike  $Z(\cdot)$ ,  $\epsilon_j$  and  $e_k$  are two different white noise processes and do not originate from a parent white noise process on  $\mathcal{D}$ . Here,  $\delta^2$  is the noise-to-spatial-temporal variance ratio.

Let  $W$  be the  $K \times p$  matrix with  $(k, u)$ -th element  $w_k(L_u)$  for  $k = 1, \dots, K$  and  $u = 1, \dots, p$ . Collect the regression coefficients into the  $(p+1) \times 1$  vector  $\beta = (\beta_1^\top, \beta_2)^\top$ . Given  $\tau^2$ , we place a multivariate Gaussian prior  $\mathbf{N}(\mu_\beta, \tau^2 V_\beta)$  on  $\beta$ , and subsequently place an inverse-gamma prior  $\mathbf{IG}(a_\tau, b_\tau)$  on  $\tau^2$ . Similarly, we also place a multivariate Gaussian prior  $\mathbf{N}(\mu_\gamma, \sigma^2 V_\gamma)$  on  $\gamma$  conditional on  $\sigma^2$ , and then place an inverse-gamma prior  $\mathbf{IG}(a_\sigma, b_\sigma)$  on  $\sigma^2$ . Define  $K \times 1$  vector  $Z_L = (Z(L_1), \dots, Z(L_K))^\top$ , and  $N \times 1$  vector  $Z_{\tilde{\ell}} = (Z(\tilde{\ell}_1), \dots, Z(\tilde{\ell}_N))^\top$ . Hence, we write the joint distribution of the data, latent process and the model parameters  $p(Y, X, Z_L, Z_{\tilde{\ell}}, \beta, \gamma, \tau^2, \sigma^2)$  as

$$\begin{aligned}
&\mathbf{N}(Y | W\beta_1 + \beta_2 Z_L, \tau^2 D_L) \times \mathbf{N}(\beta | \mu_\beta, \tau^2 V_\beta) \times \mathbf{IG}(\tau^2 | a_\tau, b_\tau) \\
&\times \mathbf{N}(X | Z_{\tilde{\ell}}, \delta^2 \sigma^2 D_{\tilde{\ell}}) \times \mathbf{N}(Z_{\tilde{\ell}} | \tilde{\Psi}\gamma, \sigma^2 C_{\tilde{\ell}}(\phi)) \times \mathbf{N}(Z_L | \mu_{L|\tilde{\ell}}, \sigma^2 C_{L|\tilde{\ell}}) \\
&\times \mathbf{N}(\gamma | \mu_\gamma, \sigma^2 V_\gamma) \times \mathbf{IG}(\sigma^2 | a_\sigma, b_\sigma),
\end{aligned} \tag{5}$$

where  $D_L$  is  $K \times K$  diagonal matrix with the  $k$ th diagonal element  $(|B_k||I_k|)^{-1}$  for each  $k$ , and  $D_{\tilde{\ell}}$  is  $N \times N$  diagonal matrix with the  $j$ th diagonal element  $|\tilde{I}_j|^{-1}$  for each  $j$ . Following (2), we have

$$\mu_{L|\tilde{\ell}} = \bar{\Psi}\gamma + C_{\tilde{\ell}, L}(\phi)^\top C_{\tilde{\ell}}^{-1}(\phi) \left( Z_{\tilde{\ell}} - \tilde{\Psi}\gamma \right) \quad \text{and} \quad C_{L|\tilde{\ell}} = C_L(\phi) - C_{\tilde{\ell}, L}(\phi)^\top C_{\tilde{\ell}}(\phi)^{-1} C_{\tilde{\ell}, L}(\phi), \tag{6}$$

where  $N \times N$  matrix  $C_{\tilde{\ell}}(\phi)$  and  $N \times K$  matrix  $C_{\tilde{\ell},L}(\phi)$  are as defined in (3). Furthermore, the  $N \times r$  basis matrix  $\tilde{\Psi}$  is known and has  $(j,v)$ -th element  $\tilde{\psi}_v(\tilde{\ell}_j) = |\tilde{I}_j|^{-1} \int_{\tilde{I}_j} \psi_v(s,t) dt$  for  $v = 1, \dots, r$  and  $j = 1, \dots, N$ . Similarly, the  $K \times r$  basis matrix  $\bar{\Psi}$  is known, with  $(k,v)$ -th element  $\bar{\psi}_v(L_k) = (|L_k|)^{-1} \int_{L_k} \psi(\ell) d\ell$  for  $k = 1, \dots, K$  and  $v = 1, \dots, r$ . Equation 5 factorizes the joint distribution in a way that reflects the hierarchical structure of the data-generative model, as given by (4).

**3.3. Model assumptions.** We pursue analytically accessible posterior distribution of all model parameters in (4) conditional on  $\phi$  and  $\delta^2$ . The auxiliary model hyperparameters  $\mu_\beta, V_\beta, \mu_\gamma, V_\gamma, a_\tau, b_\tau, a_\sigma$  and  $b_\sigma$  are assumed fixed. Following (5), since  $\beta$  and  $\tau^2$  are conditionally independent of all model components, given  $(Y, Z_L)$ , we factorize the posterior distribution as

$$p(Z_L, Z_{\tilde{\ell}}, \beta, \gamma, \tau^2, \sigma^2 | Y, X, \phi, \delta^2) = p(\beta, \sigma^2 | Z_L, Y) \times p(Z_L, Z_{\tilde{\ell}}, \gamma, \sigma^2 | Y, X, \phi, \delta^2) \quad (7)$$

We make one additional assumption on  $p(Z_L, Z_{\tilde{\ell}}, \gamma, \sigma^2 | Y, X, \phi, \delta^2)$  as follows.

**Assumption 1.** *Under fixed values of  $\phi$  and  $\delta^2$ , the latent spatial-temporal processes  $(Z_L, Z_{\tilde{\ell}})$  is conditionally independent of the outcome  $Y$ , given noisy measurements of the covariate  $X$ , i.e.,*

$$p(Z_L, Z_{\tilde{\ell}}, \gamma, \sigma^2 | Y, X, \phi, \delta^2) = p(Z_L, Z_{\tilde{\ell}}, \gamma, \sigma^2 | X, \phi, \delta^2).$$

This implies that, instead of estimating a Bayesian full probability model, we assume that the latent spatial-temporal processes  $(Z_L, Z_{\tilde{\ell}})$  are a priori dependent, but once  $X$  is observed,  $Y$  provides no additional information about them. In the context of our analysis, Assumption 1 appears quite natural since it dictates that the estimation of annual county-level ozone concentrations only depends on  $X$ , i.e., the monthly ozone levels measured by the monitoring stations, and does not depend on asthma ED visit rates  $Y$ . In the literature of Bayesian inference for complex hierarchical models, this is a familiar approach, especially when there are multiple data sources that provide information about different parameters in the model (see, for e.g., Bayarri et al., 2009; Jacob et al., 2017). This is commonly known as *modularization*, as Assumption 1 modulates the flow of information from observed data to the latent process. Moreover, our proposed framework can also be viewed as a *cut model* (Plummer, 2014), since it mimics a cut in the directed acyclic graph representing the hierarchical model, separating the graph into two components by logically preventing the “feedback” from one part of the model to the other during inference. We illustrate this in Figure 4. The separate components of the model are often called *modules*. In this context, the hierarchical model (4) is made up of two modules – a linear regression module (Module 1), and a spatial-temporal regression module (Module 2). In a full probability model, the information from the outcome  $Y$  “feeds back” through the graph in Figure 4 to influence the posterior distribution of the latent spatial-temporal processes  $(Z_L, Z_{\tilde{\ell}})$ . Instead, both  $Z_L$  and  $Z_{\tilde{\ell}}$  are estimated using the auxiliary data  $X$ . Cut models are also attractive from a computational viewpoint, as they significantly simplify sampling from the posterior distribution.

**3.4. Posterior distribution.** Following Assumption 1, conditional on  $\phi$  and  $\delta^2$ , we derive that the posterior distribution corresponding to the hierarchical model in (4) as

$$\begin{aligned} & \text{N}(\beta | M_\beta m_\beta, \tau^2 M_\beta) \times \text{IG}(\tau^2 | a_\tau^*, b_\tau^*) \times \text{N}(Z_L | \mu_{L|\tilde{\ell}}, \sigma^2 C_{L|\tilde{\ell}}) \\ & \times \text{N}(Z_{\tilde{\ell}} | M_z m_z, \sigma^2 M_z) \times \text{N}(\gamma | M_\gamma m_\gamma, \sigma^2 M_\gamma) \times \text{IG}(\sigma^2 | a_\sigma^*, b_\sigma^*) , \end{aligned} \quad (8)$$

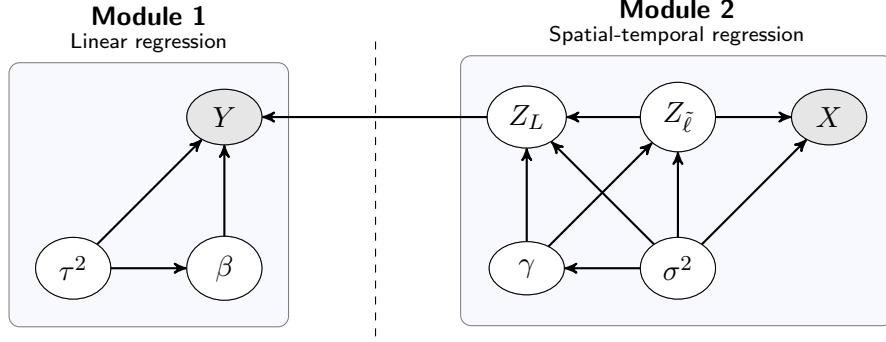


FIGURE 4. Directed acyclic graph (DAG) illustrating the conditional dependence structure of the hierarchical model (4). Nodes shaded in gray correspond to data that are observed; all other nodes represent unobserved (latent or unknown) quantities. The vertical dashed line denotes a “cut” in the DAG, indicating restricted flow of information from  $Y$  to Module 2 during posterior inference.

where  $a_\tau^* = a_\tau + K/2$ ,  $b_\tau^* = b_\tau + (Y^\top D_L^{-1} Y + \mu_\beta^\top V_\beta^{-1} \mu_\beta - m_\beta^\top M_\beta m_\beta)/2$ , and  $a_\sigma^* = a_\sigma + N/2$ ,  $b_\sigma^* = b_\sigma + (X^\top V_X^{-1} X + \mu_\gamma^\top V_\gamma^{-1} \mu_\gamma - m_\gamma^\top M_\gamma m_\gamma)/2$ , with  $V_X = C_{\tilde{\ell}}(\phi) + \delta^2 D_{\tilde{\ell}}$ , and

$$\begin{aligned} M_\beta^{-1} &= \tilde{W}^\top D_L^{-1} \tilde{W} + V_\beta^{-1}, & m_\beta &= \tilde{W}^\top D_L^{-1} Y + V_\beta^{-1} \mu_\beta, \\ M_\gamma^{-1} &= \tilde{\Psi}^\top V_X^{-1} \tilde{\Psi} + V_\gamma^{-1}, & m_\gamma &= \tilde{\Psi}^\top V_X^{-1} X + V_\gamma^{-1} \mu_\gamma, \\ M_z^{-1} &= C_{\tilde{\ell}}(\phi)^{-1} + (1/\delta^2) D_{\tilde{\ell}}^{-1}, & m_z &= (X - \tilde{\Psi} \gamma)/\delta^2, \end{aligned} \quad (9)$$

where  $\tilde{W} = [W, Z_L]$  being the  $K \times (p+1)$  matrix obtained by augmenting  $W$  by the vector  $Z_L$  on the right. The first two terms in (8) denote the conditional posterior distributions  $p(\beta | \tau^2, Z_L, Y)$  and  $p(\tau^2 | Z_L, Y)$ , respectively. The third term denote the posterior predictive distribution  $p(Z_L | Z_{\tilde{\ell}}, \gamma, \sigma^2, \phi, \delta^2)$ , with  $\mu_{L|\tilde{\ell}}$  and  $C_{L|\tilde{\ell}}$  as defined in (6). The fourth through sixth terms represent the conditional posterior distributions  $p(Z_{\tilde{\ell}} | \gamma, \sigma^2, X, \phi, \delta^2)$ ,  $p(\gamma | \sigma^2, X, \phi, \delta^2)$ , and the marginal posterior distribution  $p(\sigma^2 | X, \phi, \delta^2)$ , respectively. This factorization follows from the fact that  $Z_L$  is conditionally independent of the data  $X$ , given  $Z_{\tilde{\ell}}$ ,  $\gamma$ ,  $\sigma^2$  and  $\phi$ . It is important to remark that, the analytical tractability of the posterior distribution arises from the assumption that  $\phi$  and  $\delta^2$  are fixed and due to the condition specified in Assumption 1.

We use composition sampling to draw samples from the posterior distribution in (8). We elaborate the steps in Algorithm 1. Here,  $\text{chol}(\cdot)$  refers to the lower-triangular Cholesky factor of a square matrix. The steps in Algorithm 1, generate  $B$  samples  $\{\sigma^{(b)}, \gamma^{(b)}, Z_{\tilde{\ell}}^{(b)}, Z_L^{(b)}, \tau^{(b)}, \beta^{(b)}\}_{b=1}^B$  from the posterior distribution. The sampling algorithm is dominated by Cholesky decompositions of  $N \times N$  matrices  $V_X$ ,  $M_z$ , and  $C_{\tilde{\ell}}(\phi)$ , which accumulates  $\mathcal{O}(N^3)$  floating-point operations (flops). The matrix  $M_z$  is calculated efficiently using the identity  $M_z = \delta^2 D_{\tilde{\ell}} V_X^{-1} C_{\tilde{\ell}}(\phi)$ , where  $V_X^{-1} C_{\tilde{\ell}}(\phi)$  is calculated using triangular solves of columns of  $C_{\tilde{\ell}}(\phi)$  with respect to the Cholesky factor already calculated  $\text{chol}(V_X)$ . Placing a prior on  $\phi$  and the nugget  $\theta = \delta^2 \sigma^2$  requires iterative algorithms such as MCMC to sample from the posterior distribution, which entails repeated Cholesky decompositions of these  $N \times N$  matrices. Moreover, weak identifiability of  $\phi$ ,  $\sigma^2$  and  $\theta$  impedes convergence

---

**Algorithm 1:** Posterior Sampling

---

**Input:** Data:  $(Y, X, \tilde{\mathcal{L}}, \mathcal{L})$ ; candidate  $(\phi, \delta^2)$ , and hyperparameters.

**Output:** Posterior samples of  $\{\sigma^{2(b)}, \gamma^{(b)}, Z_{\tilde{\ell}}^{(b)}, Z_L^{(b)}, \tau^{2(b)}, \beta^{(b)}\}_{b=1}^B$

Compute  $\text{chol}(V_X)$  given  $\phi, \delta^2$ ;  $\mathcal{O}(N^3)$  flops

Compute  $\text{chol}(M_z)$  using  $\text{chol}(V_X)$ , and  $M_z = \delta^2 D_{\tilde{\ell}} V_X^{-1} C_{\tilde{\ell}}(\phi)$ ;  $\mathcal{O}(N^3)$  flops

Compute  $\text{chol}(C_{\tilde{\ell}}(\phi))$ , then  $\text{chol}(C_{L|\tilde{\ell}})$  using (6);  $\mathcal{O}(KN^2 + K^3)$  flops

Compute  $M_\gamma, m_\gamma$ , then  $a_\sigma^*, b_\sigma^*$ ;  $\mathcal{O}(rN^2 + r^3)$  flops

**for**  $b \leftarrow 1$  **to**  $B$  **do**

- Sample  $\sigma^{2(b)} \sim \text{IG}(a_\sigma^*, b_\sigma^*)$ ;
- Sample  $\gamma^{(b)} \sim \mathcal{N}(M_\gamma m_\gamma, \sigma^{2(b)} M_\gamma)$ ;
- Compute  $m_z^{(b)} = (X - \tilde{\Psi}\gamma^{(b)})/\delta^2$ ;  $\mathcal{O}(N)$  flops
- Sample  $Z_{\tilde{\ell}}^{(b)} \sim \mathcal{N}(M_z m_z^{(b)}, \sigma^{2(b)} M_z)$  using  $\text{chol}(M_z)$ ;  $\mathcal{O}(N^2)$  flops
- Compute  $\mu_{L|\tilde{\ell}}^{(b)}$  using  $Z_{\tilde{\ell}}^{(b)}$  and  $\gamma^{(b)}$  in (6);  $\mathcal{O}(KN^2)$  flops
- Sample  $Z_L^{(b)} \sim \mathcal{N}(\mu_{L|\tilde{\ell}}^{(b)}, \sigma^{2(b)} C_{L|\tilde{\ell}})$ ;  $\mathcal{O}(N)$  flops
- Set  $\tilde{W}^{(b)} = [W, Z_L^{(b)}]$ , compute  $M_\beta^{(b)}, m_\beta^{(b)}, a_\tau^{*(b)}, b_\tau^{*(b)}$  using (9);  $\mathcal{O}(K^3)$  flops
- Sample  $\tau^{2(b)} \sim \text{IG}(a_\tau^{*(b)}, b_\tau^{*(b)})$ ;
- Sample  $\beta^{(b)} \sim \mathcal{N}(M_\beta^{(b)} m_\beta^{(b)}, \tau^{2(b)} M_\beta^{(b)})$ ;

**return**  $B$  samples  $\{\sigma^{2(b)}, \gamma^{(b)}, Z_{\tilde{\ell}}^{(b)}, Z_L^{(b)}, \tau^{2(b)}, \beta^{(b)}\}_{b=1}^B$  from the posterior distribution (8).

---

of the random walk Metropolis steps, which contributes to delayed execution times. Thus, even for moderately sized datasets ( $N \sim 10^3$ ), the computation becomes too onerous for practical use and alternative strategies like low-rank models are used (Finley et al., 2015). In this context, predictive stacking presents itself as an effective alternative by enhancing the practicality of full Gaussian process models for moderately large datasets.

**Remark 1.** Given observations at  $\tilde{\mathcal{L}}$ , let  $\mathcal{L} = \{\ell_i = (s_i, t_i) : i = 1, \dots, N^*\}$  be a collection of  $N^*$  space-time coordinates in  $\mathcal{D}$ , where we wish to predict latent spatial-temporal process. Define  $N^* \times 1$  vector  $Z_\ell = (Z(\ell_1), \dots, Z(\ell_{N^*}))^\top$ . Then, posterior predictive inference at  $\mathcal{L}$  follows from

$$p(Z_\ell | X, \phi, \delta^2) = \int p(Z_\ell | Z_{\tilde{\ell}}, \gamma, \sigma^2, \phi, \delta^2) p(Z_{\tilde{\ell}}, \gamma, \sigma^2 | X, \phi, \delta^2) dZ_{\tilde{\ell}} d\gamma d\sigma^2, \quad (10)$$

where the conditional density  $p(Z_\ell | Z_{\tilde{\ell}}, \gamma, \sigma^2, \phi, \delta^2) = \mathcal{N}(Z_\ell | \mu_{\ell|\tilde{\ell}}, \sigma^2 C_{\ell|\tilde{\ell}})$ , with

$$\mu_{\ell|\tilde{\ell}} = \Psi\gamma + C_{\ell,\tilde{\ell}}(\phi) C_{\tilde{\ell}}^{-1}(\phi) (Z_{\tilde{\ell}} - \tilde{\Psi}\gamma), \quad C_{\ell|\tilde{\ell}} = C_\ell(\phi) - C_{\ell,\tilde{\ell}}(\phi) C_{\tilde{\ell}}(\phi)^{-1} C_{\ell,\tilde{\ell}}(\phi)^\top, \quad (11)$$

where  $(\mu_\ell(\gamma))_i = \mu(\ell_i; \gamma)$ , and  $N^* \times N$  matrix  $C_{\ell,\tilde{\ell}}(\phi)$ , and  $N^* \times N^*$  matrix  $C_\ell(\phi)$  are defined as

$$(C_\ell(\phi))_{i,i'} = C(\ell_i, \ell_{i'}; \phi), \quad (C_{\ell,\tilde{\ell}}(\phi))_{i,j} = |\tilde{I}_j|^{-1} \int_{\tilde{I}_j} C((s_i, t_i), (\tilde{s}_j, t); \phi) dt,$$

for  $i, i' = 1, \dots, N^*$  and  $j = 1, \dots, N$ . We sample from (10) by first drawing  $\{Z_{\tilde{\ell}}^{(b)}, \gamma^{(b)}, \sigma^{2(b)}\}$  from  $p(Z_{\tilde{\ell}}, \gamma, \sigma^2 | X, \phi, \delta^2)$  using Algorithm 1. Then, for each drawn value of  $\{Z_{\tilde{\ell}}^{(b)}, \gamma^{(b)}, \sigma^{2(b)}\}$ , we sample

$Z_\ell$  from  $\mathsf{N}(Z_\ell \mid \mu_{\ell|\tilde{\ell}}^{(b)}, \sigma^2 C_{\ell|\tilde{\ell}})$ , where  $\mu_{\ell|\tilde{\ell}}^{(b)}$  is obtained by substituting  $Z_{\tilde{\ell}}$  and  $\gamma$  by  $Z_{\tilde{\ell}}^{(b)}$  and  $\gamma^{(b)}$ , respectively, in (11). Repeating this for  $b = 1, \dots, B$  yields samples  $\{Z_\ell^{(b)} : b = 1, \dots, B\}$  from (10).

**3.5. Spatial-temporal correlation function.** As can be seen from (3), evaluation of the elements of  $C_{\tilde{\ell}}(\phi)$  involves numerical integrations over the temporal domain. Popular methods for univariate numerical integration (e.g., trapezoidal/Simpson's rule, quadrature, Monte Carlo) for a smooth function incurs  $\mathcal{O}(M)$  operations, where  $M$  denotes the number of knots/points at which the integrand is evaluated. Hence, evaluation of all elements of  $C_{\tilde{\ell}}(\phi)$  would require  $\mathcal{O}(MN^2)$  flops, which can be computationally expensive for moderately large  $N$ . Therefore, we explore spatial-temporal correlation functions that offer improved tractability, especially over the temporal domain.

In particular, we assume a separable spatial-temporal correlation function (Mardia and Goodall, 1993) of the form  $C(\ell, \ell'; \phi) = C_s(s, s'; \phi_s, \nu) \cdot C_t(t, t'; \phi_t)$ , where  $C_s(s, s'; \phi_s, \nu)$  and  $C_t(t, t'; \phi_t)$  denote the isotropic Matérn and the exponential correlation functions, respectively, given by

$$C_s(s, s'; \phi_s, \nu) = \frac{(\phi_s \|s - s'\|)^\nu}{2^{\nu-1} \Gamma(\nu)} K_\nu(\phi_s \|s - s'\|), \quad (12)$$

$$C_t(t, t'; \phi_t) = \exp(-\phi_t |t - t'|),$$

where  $\|s - s'\|$  is the Euclidean distance between  $s, s' \in \mathcal{S}$ . The function  $\Gamma(\cdot)$  denotes the gamma function, and  $K_\nu$  is the modified Bessel function of the second kind of order  $\nu$  which may be fractional (Abramowitz and Stegun, 1965, Chapter 10). Hence, in this case, the process parameter  $\phi = (\phi_s, \nu, \phi_t)$  is a 3-dimensional parameter. The assumption of this multiplicative form conveniently separates space and time in calculation of covariance matrices  $C_{\tilde{\ell}}(\phi)$ ,  $C_L(\phi)$ , and the cross-covariance matrix  $C_{\tilde{\ell}, L}(\phi)$ . For example, the elements of  $C_{\tilde{\ell}}(\phi)$  can be rewritten as

$$(C_{\tilde{\ell}}(\phi))_{j,j'} = (|\tilde{I}_j| |\tilde{I}_{j'}|)^{-1} C_s(\tilde{s}_j, \tilde{s}_{j'}; \phi_s, \nu) \int_{\tilde{I}_{j'}} \int_{\tilde{I}_j} C_t(t, t'; \phi_t) dt dt'. \quad (13)$$

We make an additional assumption that facilitates closed form expression for the integral in (13).

**Assumption 2.** *The temporal blocks  $\tilde{I}_j$  for  $j = 1, \dots, N$  and  $I_k$  for  $k = 1, \dots, K$ , are of the form  $\tilde{I}_j = (\tilde{a}_j, \tilde{b}_j)$ , and  $I_k = (a_k, b_k)$  where  $\tilde{a}_j < \tilde{b}_j$  for each  $j$  and  $a_k < b_k$  for each  $k$ .*

The distribution theory in Section 3.1 applies to any Borel-measurable subsets  $\tilde{I}_j$  and  $I_k$ . In practice, however, it is natural to assume that observations are aggregated over a single interval.

**Proposition 1.** *For any  $a < b$  and  $c < d$ , suppose  $\tilde{C}_t(a, b, c, d; \phi) = \int_c^d \int_a^b C_t(t, t'; \phi_t) dt dt'$ , then*

(a) *for non-overlapping  $(a, b)$  and  $(c, d)$ , with  $a < b \leq c < d$ ,*

$$\tilde{C}_t(a, b, c, d; \phi_t) = \frac{1}{\phi_t^2} [F(a, d) + F(b, c) - F(a, c) - F(b, d)],$$

(b) *if intervals  $(a, b)$  and  $(c, d)$  overlap, with  $a \leq c < b \leq d$ ,*

$$\tilde{C}_t(a, b, c, d; \phi_t) = \frac{1}{\phi_t^2} [2\phi_t(b - c) + F(a, d) + F(c, b) - F(a, c) - F(b, d)],$$

(c) *if  $(a, b)$  is nested within  $(c, d)$ , i.e. either  $c \leq a < b < d$  or  $c < a < b \leq d$ ,*

$$\tilde{C}_t(a, b, c, d; \phi_t) = \frac{1}{\phi_t^2} [2\phi_t(b - a) + F(a, d) + F(c, b) - F(c, a) - F(b, d)],$$

where  $F(a_1, a_2) = F(a_1, a_2; \phi_t) = \exp(-\phi_t(a_2 - a_1))$  for any  $a_1, a_2 \in \mathcal{T}$ .

*Proof.* See Appendix C for details.  $\square$

We benefit from the customary Assumption 2 that the integral of  $C_t(\cdot, \cdot; \phi_t)$  as appearing in (13) admits closed-form expressions, as detailed in Proposition 1. This facilitates a rapid evaluation of elements of the spatial-temporal covariance matrix  $C_{\tilde{\ell}}(\phi)$  without resorting to approximations using numerical methods. However, Proposition 1 does not extend to similar results for integrations over irregular spatial blocks (e.g., counties of California), and hence numerical methods are essential for computing elements of  $C_L(\phi)$  and  $C_{\tilde{\ell}, L}(\phi)$ . In this context, assuming the separable spatial-temporal covariance function (12) provides additional advantage, as we are able to write the integral as a product of a spatial and a temporal component. Following (3) and Assumption 2, we have

$$\begin{aligned} (C_{\tilde{\ell}, L}(\phi))_{j,k} &= \frac{|B_k|^{-1}}{(\tilde{b}_j - \tilde{a}_j)(b_k - a_k)} \tilde{C}_t(\tilde{a}_j, \tilde{b}_j, a_k, b_k) \int_{B_k} C_s(\tilde{s}_j, s) ds, \\ (C_L(\phi))_{k,k'} &= \frac{(|B_k||B_{k'}|)^{-1}}{(b_k - a_k)(b_{k'} - a_{k'})} \tilde{C}_t(a_k, b_k, a_{k'}, b_{k'}) \int_{B_{k'}} \int_{B_k} C_s(s, s') ds ds', \end{aligned} \quad (14)$$

for each  $j = 1, \dots, N$  and  $k, k' = 1, \dots, K$ . The role of Proposition 1 becomes clear from (14), as it helps to simplify (3) by reducing the necessity of any numerical integration in the time domain. From a computational perspective, due to Proposition 1, evaluation of  $C_{\tilde{\ell}}(\phi)$  requires absolutely no numerical integration,  $C_{\tilde{\ell}, L}(\phi)$  requires  $NK$  numerical integrals and  $C_L(\phi)$  requires  $K^2$  numerical integrals. In practice,  $K$  is generally much smaller compared to  $N$ . For example, California has only 58 counties, and we include observations for 5 racial groups over 8 years from 2015 through 2022, which amounts to  $K = 1510$  after removal of missing records. On the other hand, we have recorded ozone levels at  $N = 15,725$  spatial-temporal coordinates.

#### 4. PREDICTIVE STACKING

**4.1. Choice of candidate models.** To implement stacking, we first fix the values of the hyperparameters of the auxiliary model. In practice, we assume  $\mu_\beta = 0_{p+1}$ ,  $V_\beta = \delta_\beta^2 I_{p+1}$  for a sufficiently large  $\delta_\beta$  to specify a weakly informative prior for the fixed effects  $\beta$ . Here,  $0_{p+1}$  denotes the zero vector of length  $p+1$ . Similarly, we assume  $\mu_\gamma = 0_r$  and  $V_\gamma = \delta_\gamma^2 I_r$  for the Gaussian prior on  $\gamma$ . For  $\tau^2$  and  $\sigma^2$ , we choose the shape parameters  $a_\tau = a_\sigma = 2$  and the scale parameters  $b_\tau = b_\sigma = 0.1$ .

For  $\phi = (\phi_s, \phi_t, \nu)$  and  $\delta^2$ , we choose grids of candidate values given by  $G_{\phi_s}$ ,  $G_{\phi_t}$ ,  $G_\nu$  and  $G_{\delta^2}$ . The grid  $G_{\delta^2}$  is chosen based on the values of the nugget and partial sill, estimated from an empirical semivariogram. [Zhang et al. \(2024\)](#) provides further details on how the quantiles of a Beta distribution specified by the estimated values of the nugget and the partial sill can be used to provide useful information for selecting  $G_{\delta^2}$ . On the other hand, the candidate values of  $\phi_s$  and  $\phi_t$  are chosen so that the “effective range” (distance at which the correlation drops below 5%) corresponding to the candidate values is between 20% and 70% of the maximum distance between the space-time coordinates (see, Chapter 2 [Banerjee et al., 2014](#)). We choose  $G_\nu$  to include some customary values of the Matérn smoothness parameter that are often used in the literature of spatial analysis. For example, a possible choice is  $G_\nu = \{0.5, 1.0, 1.5\}$ .

**4.2. Stacking algorithm.** While we follow the general strategy of stacking predictive densities as proposed in Yao et al. (2018), our development is distinct in that we adapt the approach to our more complex model structure, which differs significantly from the standard spatial-temporal modeling frameworks considered in the previous literature (Zhang et al., 2024; Pan et al., 2025). In this aspect, Assumption 1 plays a key role, as it restricts the inference for the latent spatial-temporal process to Module 2, within which we seek to find an optimal way to combine the inference conditional on candidate values of  $\phi$  and  $\delta^2$ . We elaborate below.

Let  $\mathcal{M} = \{M_1, \dots, M_G\}$  denote a collection of candidate models  $G$ , where  $M_g$  corresponds to fixed values of the parameters  $(\phi_g, \delta_g^2)$ , for  $g = 1, \dots, G$ . Predictive stacking finds a probability distribution  $\tilde{p}$  in the class  $\mathcal{C} = \{\sum_{g=1}^G \alpha_g p(\cdot | X, M_g) : \sum_{g=1}^G \alpha_g = 1, \alpha_g \geq 0\}$ , such that the Kullback-Leibler (KL) divergence between  $\tilde{p}(\cdot | X)$  and  $p_t(\cdot | X)$  is minimized, where  $p_t$  denotes the posterior predictive distribution under the true data-generating model. Here,  $p(\cdot | X, M_g)$  denotes the posterior predictive distribution under the candidate model  $M_g$ , for each  $g$ . We define the stacking weights  $\alpha = (\alpha_1, \dots, \alpha_G)$  as the solution to the optimization problem

$$\max_{\alpha} \quad \frac{1}{N} \sum_{j=1}^N \log \sum_{g=1}^G \alpha_g p(X(\tilde{\ell}_j) | X_{-j}, M_g) \quad \text{subject to} \quad \alpha^\top 1_G = 1, \alpha \in [0, 1]^G, \quad (15)$$

where  $X_{-j}$  denotes the data  $X$  with the  $j$ th observation removed, and  $p(X(\tilde{\ell}_j) | X_{-j}, M_g)$  denotes the leave-one-out predictive density corresponding to the  $j$ th observation. This follows from a result that establishes that minimizing  $\text{KL}(\tilde{p}(\cdot | X), p_t(\cdot | X))$  under the constraint  $\tilde{p} \in \mathcal{C}$  is asymptotically equivalent to the optimization problem in (15) (see, Le and Clarke, 2017; Clyde and Iversen, 2013). The optimization task in (15) falls into the class of convex problems and can be formulated and solved using suitable modeling tools and solvers. Assumption 1 implies that the optimal stacking weights can be computed solely based on Module 2, that is, using only  $X$ . Posterior inference for quantities of interest subsequently proceeds from the “stacked posterior”,

$$\tilde{p}(\cdot | X, Y) = \sum_{g=1}^G \hat{\alpha}_g p(\cdot | X, Y, M_g), \quad (16)$$

where  $\hat{\alpha}_g$  denotes optimal stacking weights obtained by solving the optimization task in (15).

An important prerequisite for evaluating the objective function of interest in (15) is the computation of the leave-one-out predictive densities. Under the hierarchical model (4), the leave-one-out predictive densities admit a closed form, given by

$$p(X(\tilde{\ell}_j) | X_{-j}, M_g) = t_{2a_{\sigma,j}^*} \left( X(\tilde{\ell}_j) \mid \mu_{g,j|-j}(\tilde{\ell}_j), (b_{\sigma,j}^*/a_{\sigma,j}^*) \sigma_{g,j|-j} \right), \quad (17)$$

where  $t_\rho(x; m, v^2)$  denotes the location-scale  $t$ -density with degrees of freedom  $\rho$ , location  $m$  and scale  $v$ , evaluated at  $x$ , and the parameters  $\mu_{g,j|-j}(\tilde{\ell}_j)$  and  $\sigma_{g,j|-j}$  are given by

$$\begin{aligned} \mu_{g,j|-j}(\tilde{\ell}_j) &= R_{g,j}^\top V_{g,X_{-j}} X_{-j} + H_{g,j}^\top M_{\gamma,g,j} m_{\gamma,g,j}, \\ \sigma_{g,j|-j} &= V_{g,X_j|X_{-j}} + H_{g,j}^\top M_{\gamma,g,j} H_{g,j}, \end{aligned}$$

where  $R_{g,j}$  is the  $(N - 1) \times 1$  spatial-temporal cross-correlation matrix between  $\tilde{\mathcal{L}} \setminus \tilde{\ell}_j$  and  $\{\tilde{\ell}_j\}$  under  $M_g$ , and is given by the  $j$ th column of  $C_{\tilde{\ell},-j}(\phi_g)$  which denotes the matrix  $C_{\tilde{\ell}}(\phi_g)$  with both

its  $j$ th row and column removed. Moreover, the matrix  $V_{g,X_{-j}}$  is obtained by removing the  $j$ th row and column of  $V_{g,X} = C_{\tilde{\ell}}(\phi_g) + \delta_g^2 D_{\tilde{\ell}}$ . In addition,  $M_{\gamma,g,j}^{-1} = \tilde{\Psi}_{-j}^\top V_{g,X_{-j}} \tilde{\Psi}_{-j} + V_\gamma^{-1}$ , and  $m_{\gamma,g,j} = \tilde{\Psi}_{-j} V_{g,X_{-j}} X_{-j} + V_\gamma^{-1} \mu_\gamma$  where the  $(N-1) \times r$  matrix  $\tilde{\Psi}_{-j}$  is obtained by deleting the  $j$ th row of  $\tilde{\Psi}$ . The  $r \times 1$  vector  $H_{g,j} = \tilde{\psi}(\tilde{\ell}_j) - \tilde{\Psi}_{-j}^\top V_{g,X_{-j}}^{-1} R_{g,j}$ , and the scalar  $V_{g,X_j|X_{-j}} = (V_{g,X})_{j,j} - R_{g,j}^\top V_{g,X_{-j}} R_{g,j}$  with  $(V_{g,X})_{j,j}$  denoting the  $j$ th diagonal element of  $V_{g,X}$ . Furthermore,  $a_{\sigma,g}^* = a_\sigma + (N-1)/2$ , and  $b_{\sigma,g}^* = b_\sigma + (X_{-j}^\top V_{g,X_{-j}} X_{-j} + \mu_\gamma^\top V_\gamma^{-1} \mu_\gamma - m_{\gamma,g,j}^\top M_{\gamma,g,j} m_{\gamma,g,j})/2$ .

Evaluation of (17) is dominated by the Cholesky decomposition of the  $(N-1) \times (N-1)$  matrix  $V_{g,X_{-j}}$ , which requires  $\mathcal{O}(N^3)$ , for each  $j$ . Hence, a naive approach to find the leave-one-out predictive densities under each  $M_g$  results in  $\sim \mathcal{O}(N^4)$  flops, which is impractical. We mitigate this issue by reusing the Cholesky factor of  $V_{g,X}$  which has already been computed once while fitting the model  $M_g$  (Kim et al., 2002). Subsequently, we compute the Cholesky factor of  $V_{g,X_{-j}}$  for each  $j$  using an efficient rank-one update algorithm (Krause and Igel, 2015), which ultimately accumulates to  $\mathcal{O}(N^3)$  flops, thereby delivering a significant speedup over the naive approach.

Although rank-one updates are faster than the naive method, they are computationally expensive; hence, an alternative approach is desirable. Here, importance weighting is an attractive option to approximate leave-one-out predictive densities (see, for e.g., Gelfand et al., 1992; Vehtari et al., 2024b). Suppose  $\{Z_{\tilde{\ell},g}^{(b)}, \sigma_g^{2(b)}\}_{b=1}^B$  denotes  $B$  draws from the posterior distribution  $p(Z_{\tilde{\ell}}, \sigma^2 | X, M_g)$ , then for each  $j$ , we approximate the leave-one-out predictive densities by the weighted mean

$$p(X(\tilde{\ell}_j) | X_{-j}, M_g) \approx \frac{1}{\sum_{b=1}^B r_{j,g}^b} \sum_{b=1}^B r_{j,g}^b \mathsf{N}\left(X(\tilde{\ell}_j) | Z_{\tilde{\ell},g,j}^{(b)}, \sigma_g^{2(b)}\right), \quad (18)$$

where  $Z_{\tilde{\ell},g,j}^{(b)}$  denotes the  $j$ th element of  $Z_{\tilde{\ell},g}^{(b)}$ , and  $r_{j,g}^b$  is the important ratio defined as  $1/r_{j,g}^b = \mathsf{N}(X(\tilde{\ell}_j) | Z_{\tilde{\ell},g,j}^{(b)}, \sigma_g^{2(b)})$ . The weights  $r_{j,g}^b$  tend to have a high or infinite variance, introducing instability in the computation (18). To address these issues, Vehtari et al. (2017) proposes stabilizing the weights by fitting a generalized Pareto distribution to the tail of the weight distribution using the empirical Bayes estimation algorithm proposed in Zhang and Stephens (2009). Thus, no additional model fitting is necessary to calculate leave-one-out predictive densities. Compared to  $\mathcal{O}(N^3)$  for the Cholesky factor update algorithm, the computational cost of this approximate method is  $O(N \log N)$  and therefore is much faster. Figure A5 in the Appendix compares exact and approximate leave-one-out predictive densities computed using the closed-form expression (17) and Pareto smoothed importance sampling, respectively, for a spatial regression model. Both methods deliver practically indistinguishable results but differ in computational cost; exact computation requires roughly 3 minutes, whereas importance sampling needs only 0.1 second. We implement this using the R package `loo` (Vehtari et al., 2024a).

## 5. SIMULATION

**5.1. Simulated data.** To address the analytical challenges posed by data that are simultaneously spatially and temporally misaligned, we designed a simulation study that closely replicates the structure and characteristics of the outcome (asthma) and the exposure (ozone) data, as discussed in Section 2. For the exposure data, we consider the unit square, i.e.,  $[0, 1]^2$ , and the interval  $[0, 12]$  as the spatial and temporal domains of interest. The interval  $[0, 12]$  represents the duration of

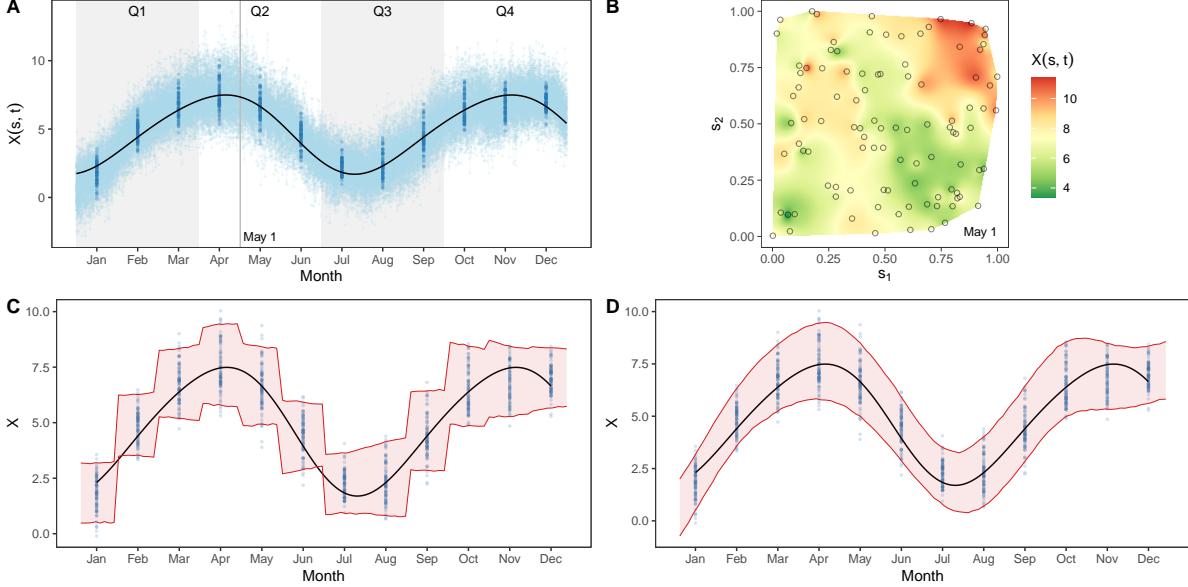


FIGURE 5. A: Light blue dots represents simulated daily measurements at 100 locations over a year, whereas dark blue dots denote monthly averages at each site. B: Interpolated spatial surface of a snapshot of the data (denoted by the vertical line in subfigure A) at May 1. C, D: 95% credible interval of posterior predictive samples drawn from the stacked posterior  $\tilde{p}(Z(\ell) | X)$  with  $\psi(\cdot)$  chosen as monthly factors and Fourier series as basis, respectively, to capture seasonal variations.

twelve months of a year. We simulate spatially-temporally correlated data at  $n_s = 100$  locations sampled uniformly in  $[0, 1]^2$ , and at  $n_t = 360$  equally spaced time points in  $[0, 12]$ , which amounts to spatial-temporal coordinates  $\{\ell_1, \dots, \ell_{n_s n_t}\}$  with  $n_s n_t = 36,000$ . The data is simulated following the model  $X(\ell_i) = Z(\ell_i) + e'_i$  for  $i = 1, \dots, n_s n_t$ , where  $e'_i \stackrel{\text{ind}}{\sim} \mathcal{N}(0, 1)$  denotes the measurement error for each  $i$  and  $Z(\ell) \sim \text{GP}(\mu(\ell), C(\ell, \ell'; \phi))$ . The covariance function  $C(\ell, \ell'; \phi)$  is the same as (12) with  $\phi_s = 4$ ,  $\nu = 0.5$  and  $\phi_t = 0.6$ . We choose the mean function  $\mu(\ell)$  to depend only on time and therefore write it as  $\mu(t)$ . We model  $\mu(t)$  as a random draw from  $\text{GP}(5, 4 \cdot R(t, t'; p, \lambda))$ , where  $R(t, t'; p, \lambda) = \exp(-2\lambda^2 \sin^2(\pi|t - t'|/p))$  denotes a periodic covariance kernel with period  $p = 7$  and decay  $\lambda = 0.1$ . This helps introduce seasonal variations in the simulated data. Thus, the simulated data  $X(\ell_i)$  for  $i = 1, \dots, n_s n_t$  mimic daily measurements of a variable at 100 monitoring sites over a year. From these, we compute monthly averages at each site, which we denote by  $X(\bar{\ell}_j)$  for  $j = 1, \dots, 12n_s$ . To introduce missingness, we randomly remove 10% of the data from the monthly averages and denote the remaining data simply by  $X$ . We proceed to our simulation experiment with these point-referenced monthly averages as the data. See Figures 5A and 5B for a visualization of the simulated data. For subsequent analysis, we consider the grids of candidate values of the spatial decay parameter  $G_{\phi_s} = \{2, 3, 5\}$ ,  $G_{\phi_t} = \{0.3, 0.5, 1\}$ ,  $G_\nu = \{0.5, 1, 1.5\}$ , and  $G_{\delta^2} = \{0.75, 1.5\}$ . Hence, we stack on  $3 \times 3 \times 3 \times 2 = 54$  models.

**5.2. Point-level spatial-temporal prediction.** As a first step, we examine the stacked posterior predictive distribution  $\tilde{p}(X(\ell) | X)$ , which corresponds to predicting the underlying process at a

finer temporal resolution (e.g., daily) from coarser, aggregated observations (see Remark 1). In this context,  $\psi(\cdot)$  in the mean function  $\psi(\cdot)$  of the latent spatial-temporal process plays a key role. Under the assumption that  $\mu(\ell) = \psi(t)^\top \gamma$  depends only on time, we study the two alternative specifications of the mean function

$$\begin{aligned}\mu^{(1)}(t) &= \sum_{v=1}^r \psi_v^{(1)}(t) \gamma_v = \gamma_1 + \sum_{v=2}^r \gamma_v \mathbf{1}(t \in \text{month}_v), \\ \mu^{(2)}(t) &= \sum_{v=1}^r \psi_v^{(2)}(t) \gamma_v = \gamma_1 + \sum_{v=1}^{\lfloor r/2 \rfloor} \gamma_{2v} \sin(2\pi t/p_v) + \gamma_{2v+1} \cos(2\pi t/p_v),\end{aligned}\tag{19}$$

where  $\mathbf{1}(\cdot)$  denotes an indicator function,  $\text{month}_v$  denotes the interval  $(v-1, v)$  for  $v = 2, \dots, 12$ . For example,  $v = 2$  corresponds to the month February. In this case,  $r = 12$  and  $\psi^{(1)}(t)$  corresponds to a simple monthly indicator basis function, which corresponds to monthly factors under monthly aggregated data. On the other hand,  $\mu^{(2)}(t)$  consists of smooth periodic basis functions, such as sine and cosine terms with different periodicity. For our analysis, we choose  $r = 9$  with  $p_v = 3 + v$  for  $v = 1, \dots, 4$ . Figures 5C and 5D illustrate the behavior of the posterior predictive distribution under the two choices of  $\psi(\cdot)$ . We notice that the smooth basis functions lead to superior reconstruction of the latent process compared to discontinuous monthly indicator functions, due to their ability to borrow strength across adjacent time points and capture underlying seasonal trends more effectively.

**5.3. Block-level spatial-temporal prediction.** Next, we study prediction of the latent process at spatial-temporal blocks. Based on monthly data available at the 100 locations, our aim is to predict quarterly aggregated data in a collection of spatial blocks, instead of points. Here, a quarter corresponds to consecutive three-month periods, as shown by alternating shaded regions in Fig. 5A. We divide the spatial domain, the unit square  $[0, 1]^2$ , for the four quarters into 40, 50, 30, and 60 irregular spatial blocks, respectively, obtained by a Voronoi tessellation based on the same number of randomly sampled points, using the R package `deldir` (Turner, 2024). Given the simulated data, we seek the posterior predictive distribution of the aggregated latent process quarterly in these spatial blocks throughout the quarters, based on the stacked posterior  $\tilde{p}(Z_L | X)$ . We assume that the mean function  $\mu(\cdot)$  is characterized by the monthly indicator basis function  $\psi^{(1)}$ , as defined in (19). Figure 6 compares the quarterly aggregated true detrended spatial-temporal surface  $Z(\ell) - \mu(\ell)$ , with which the data were simulated, and its corresponding posterior median in the spatial blocks. We notice that the block-level posterior distributions closely reflect the true spatial-temporal pattern, indicating that our proposed model accurately captures the underlying process. This further demonstrates the flexibility of our proposed hierarchical model (4), as it accommodates complicated survey designs with time-varying spatial blocks.

**5.4. Model comparison.** We evaluate and compare the predictive performance of our proposed model against some alternative approaches. For alternative models, we regress the result  $Y(L_k)$  for each  $k$ , as defined in (4), on the predictors  $w(L_k)$  and  $X(L_k)$ . Here,  $X(L_k)$  denotes the value of the covariate aggregated in the spatial-temporal block  $L_k$  and is distinct from the latent process  $Z(L_k)$ . We use available off-the-shelf spatial interpolation tools to estimate  $X(L_k)$ . More specifically, we consider two competing approaches - multilevel B-spline approximation (MBA), implemented using

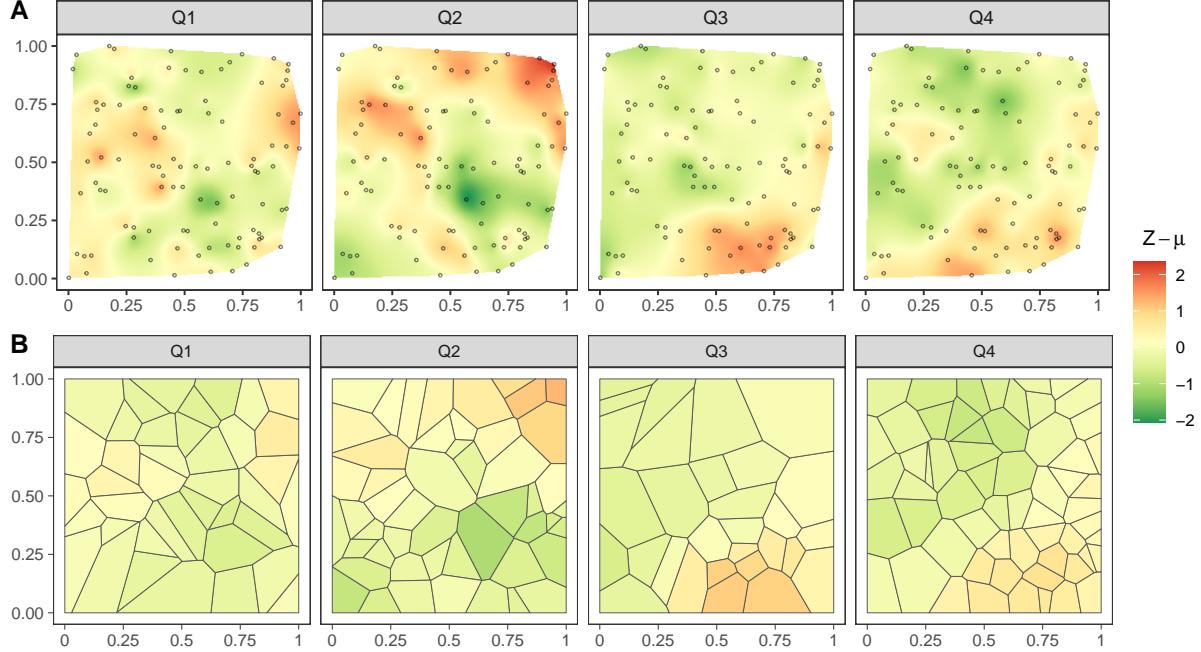


FIGURE 6. A: Interpolated spatial surfaces of the quarterly averaged de-trended true spatial-temporal process which simulated the data; B: Median of the stacked posterior predictive distribution  $\tilde{p}(Z_L - \mu_L | X)$  at the target spatial-temporal blocks.

the R package **MBA** (Finley et al., 2024), and, spatial kriging (**kriging**) from the R package **geoR** (Ribeiro Jr et al., 2024). Unlike our proposed method, **MBA** and **kriging** proceed by first obtaining point estimates in a fine grid in the spatial domain and then calculating averages in each spatial block, completely ignoring the uncertainty surrounding the estimation procedure. In addition, none of the alternative methods account for temporal correlation in the observed data, and temporal aggregation at the quarterly level is achieved by simply averaging over monthly observations.

The predictive accuracy for each model is measured using the widely applicable information criterion, WAIC (Watanabe, 2010; Gelman et al., 2013). We evaluated WAIC for each model on synthetic data, which are simulated following the steps described in Section 5.1 with the exception that we now also sample the latent process corresponding to the quarterly averages in spatial blocks, using (2), where integrated covariance kernels are approximated through Monte Carlo integration with 500 within-polygon samples. Subsequently, we sample an outcome  $Y$  following the model described in (4) with  $p = 2$ , and the covariate  $w(\cdot)$  containing an intercept and a predictor sampled from the standard normal distribution,  $\beta_1 = (5, 1)^\top$ ,  $\beta_2 = -1$  and  $\tau^2 = 5$ . From Table 1, we notice that all methods deliver similar predictive performance, with our proposed method slightly better. Moreover, the execution times for both **MBA** and **kriging** depends on the grid resolution. Also, contrary to common assumption, increasing the grid resolution has little impact on predictive accuracy, probably because finer resolution spatial interpolation contributes little when they are aggregated to obtain block-level estimates. Executing the stacking algorithm corresponds to fitting 54 candidate models in parallel across 6 cores. Hence, in addition to offering competitive predictive

Method	Model-based UQ	Temporal Dependence	WAIC	$n_{\text{grid}}$	Run Time (in secs)
MBA	$\times$	$\times$	1714.80	200	281
			1714.81	100	71
			1714.85	50	17
kriging	$\times$	$\times$	1714.75	200	326
			1714.73	100	81
			1714.75	50	20
Stacking	$\checkmark$	$\checkmark$	1714.53	—	250

TABLE 1. Comparison of predictive performance of our proposed method based on stacking of predictive densities with the alternative approaches. Here,  $n_{\text{grid}}$  refers to the number of points along each axis at which spatial interpolation is performed to estimate block-level observations; UQ: Uncertainty quantification.

performance with reasonable run-time, a crucial advantage of our framework over other methods is the ability to deliver fully model-based uncertainty quantification for the latent spatial-temporal process at any arbitrary location and time.

## 6. DATA ANALYSIS

We study the effects of ozone concentration on asthma-related ED visits by first estimating the spatial-temporal regression module based on available monthly ozone concentration data at 15,725 space-time coordinates. We use the periodic Fourier basis function, as given by  $\psi^{(2)}(\cdot)$  in (19). Guided by visual inspection of the temporal trends in ozone measurements (see Figure A4 in the Appendix), we consider periods of 5, 6, 7, and 12. We consider the separable spatial-temporal covariance kernel (12). For candidate values of  $\phi$ , we choose  $G_{\phi_s} = \{0.3, 0.5, 1\}$ ,  $G_\nu = \{0.5, 1, 1.5\}$ , and  $G_{\delta^2} = \{1.5, 2\}$ . Hence, we stack on the 54 models obtained by the Cartesian product of each of these grids. Once we obtain optimal stacking weights based on 1000 posterior samples of  $\{Z_{\ell}, \gamma, \sigma^2\}$ , we perform posterior predictive inference for ozone concentrations at 20,000 unique space-time coordinates, comprising 100 randomly sampled locations within the convex hull of monitoring sites at 200 time points spanning 2015–2022. Figure 7 displays the 95% posterior predictive credible intervals for the latent spatial-temporal process at these space-time locations, overlaid on the observed ozone measurements. The uncertainty band is estimated from the stacked posterior  $\tilde{p}(Z(\ell) | X)$ . We notice that the temporal trend is captured reasonably well.

Next, we obtain annual county-level ozone concentration predictions based on the stacked posterior  $\tilde{p}(Z(L) | X)$ . Figure 8 shows the posterior median of annual ozone concentration predictions in each county. The spatial patterns closely resemble those in Figure 3, which shows interpolated surfaces from aggregated monthly monitoring data. However, the pronounced ozone hotspots (in the counties of Riverside, San Bernadino, Inyo, Sierra, Plumas, El Dorado) and the low-concentration regions (e.g., Humboldt, Mendocino, Sonoma, Marin, San Francisco, San Mateo, Santa Cruz) in

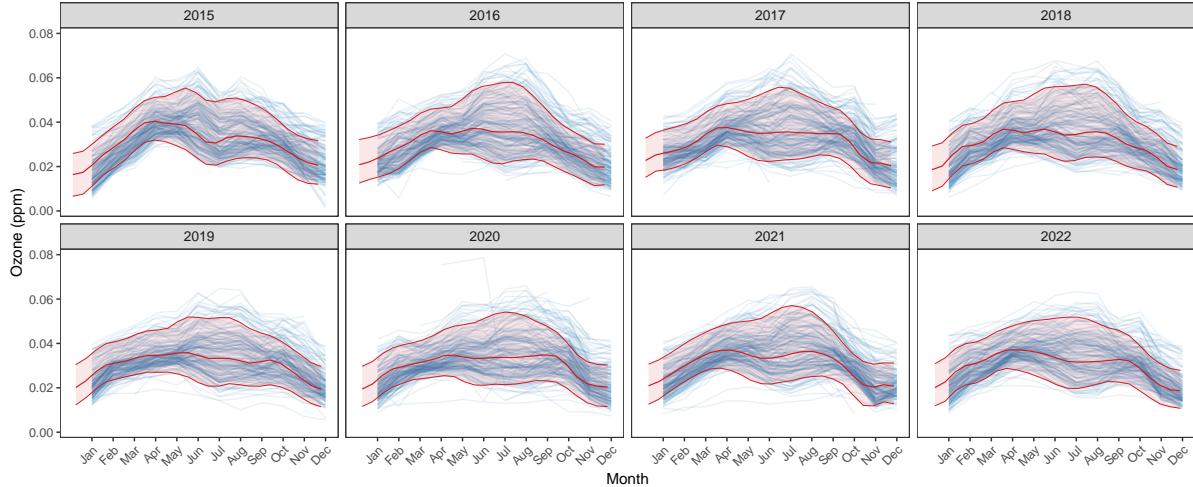


FIGURE 7. 95% credible intervals (in red) of ozone concentration predictions obtained from the stacked posterior  $\tilde{p}(Z(\ell) | X)$ , using a periodic Fourier basis mean.

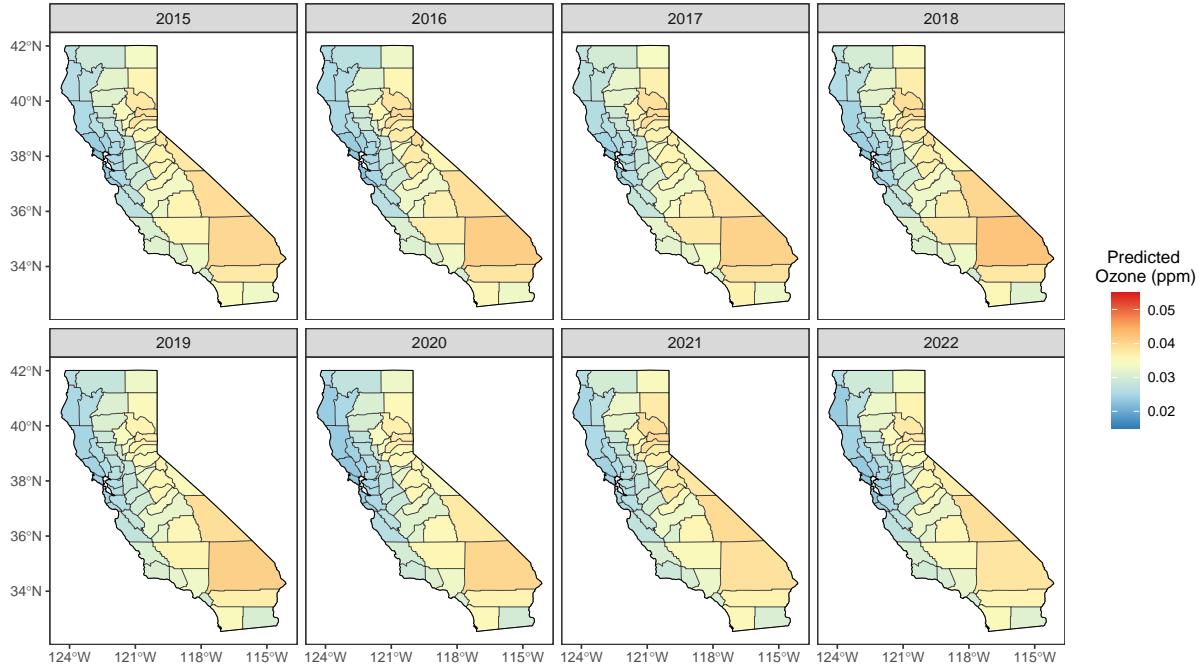


FIGURE 8. Posterior median of annual ozone concentration predictions at counties of California, obtained from monthly point-referenced observations.

Figure 3 are somewhat less pronounced in their corresponding counties in Figure 8. As anticipated, spatial averaging dampens local fluctuations, resulting in smoother estimates that may blur sharp variations in the observed data; nevertheless, the overall pattern of higher inland regions remains largely preserved. We use these posterior predictive samples of annual county-level ozone concentrations to estimate its effect on the asthma-related ED visits.

Parameter	Effect	Posterior median	95% credible interval	Details
$\beta_0$	Intercept	3.53	(3.49, 3.58)	
	<i>Reference group:</i> White			
$\beta_{11}$	American Indian/ Alaskan native	0.24	(0.16, 0.31)	
$\beta_{12}$	Asian/ Pacific Islander	-0.69	(-0.76, -0.62)	
$\beta_{13}$	Black	1.30	(1.24, 1.36)	
$\beta_{14}$	Hispanic	0.004	(-0.05, 0.06)	
$\beta_2$	Ozone (per 0.005 ppm)	-0.025	(-0.05, 0.00)	<i>baseline:</i> 0.03 ppm
$\beta_3$	Year	-0.10	(-0.11, -0.09)	<i>baseline:</i> 2015
$\sigma^2$	Error variance	0.13	(0.12, 0.14)	

TABLE 2. Asthma and ozone data 2015-22: posterior summary of model parameters.

Subsequently, we estimate a linear regression module with the objective of studying the effect of ozone, race/ethnicity, and time, on age-adjusted asthma-related emergency department visit rates. For clarity, we reformulate the linear model in (4) using symbolic notation as follows. Suppose  $Y_{ijt}$  denotes the age-adjusted asthma-related ED visit rate (per 10,000) for county  $i$ , race/ethnicity  $j$  and year  $t$ , for  $i = 1, \dots, 58$ ,  $j = 1, \dots, 5$  and  $t = 1, \dots, 8$ . We consider the log-linear model

$$\log(Y_{ijt}) = \beta_0 + \beta_1 \cdot \text{Race}_j + \beta_2 \cdot \text{Ozone}_{it} + \beta_3 \cdot \text{Year}_t + \epsilon_{ijt}, \quad \epsilon_{ijt} \stackrel{\text{ind}}{\sim} N(0, |B_i|^{-1}\tau^2) \quad (20)$$

where  $\beta_0$  is the intercept,  $\text{Race}_j$  is a  $4 \times 1$  dummy-encoded vector representing the  $j$ th racial group,  $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13}, \beta_{14})^\top$  is the corresponding  $4 \times 1$  vector of regression coefficients,  $\text{Ozone}_{it}$  is the estimated ozone concentration for county  $i$  in year  $t$ ,  $\text{Year}_t = t$  for each  $t$ ,  $|B_i|$  is the area of county  $i$ , and  $\epsilon_{ijt}$  denote independent measurement errors. We assume a hierarchical model surrounding (20) as given by (4). We assume the Gaussian prior  $\beta \mid \sigma^2 \sim N(0, \sigma^2 V_\beta)$  with  $V_\beta = 10^3 I_7$ , where  $\beta = (\beta_0, \beta_1^\top, \beta_2, \beta_3)^\top$ , and place an inverse gamma prior  $\sigma^2 \sim \text{IG}(0.01, 0.01)$ .

We present a comprehensive summary of the posterior distributions of the regression coefficients in Table 2. We find that all regression coefficients except  $\beta_{14}$  have 95% posterior credible intervals that exclude zero, suggesting strong evidence of meaningful differences in contributions specific to each racial group, except that the Hispanic and White groups do not differ significantly in their effects on asthma-related health emergencies, for a given year and ozone level. For better understanding, we examine the relative effect sizes corresponding to the racial groups of white ( $e^{\beta_0}$ ), American Indian/ Alaskan native ( $e^{\beta_0 + \beta_{11}}$ ), Asian/ Pacific Islander ( $e^{\beta_0 + \beta_{12}}$ ), black ( $e^{\beta_0 + \beta_{13}}$ ), and Hispanic ( $e^{\beta_0 + \beta_{14}}$ ) racial groups, which reflects the expected rates of asthma-related ED visits for each racial group in the reference year 2015, when exposed to the baseline level of ozone of 0.03 ppm. Figure 9A illustrates the posterior distributions of the relative effects sizes of each racial group. Our analysis reveals that the average asthma-related ED visit rates of the black group are approximately 3.6 times the average rates of the white (CI: 3.4, 3.8) and Hispanic (CI: 3.4, 3.9) groups, 2.9 (CI: 2.7, 3.1) times the average rate of the American Indian/Alaskan native group, and around 7.3 (CI:

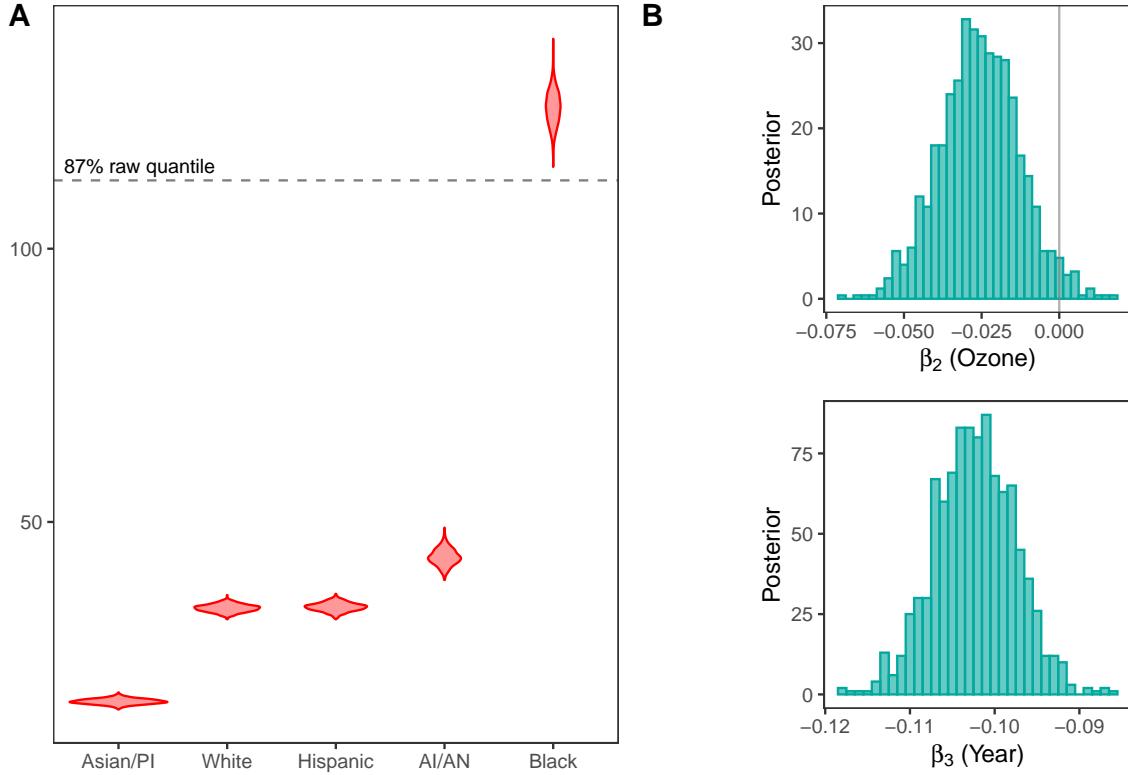


FIGURE 9. A: Posterior distributions of relative effect sizes at constant ozone concentration and year; AI/AN: American Indian/ Alaskan native, PI: Pacific Islanders. B: Posterior distributions of the coefficients corresponding to ozone and year.

6.8, 7.9) times the average rate of the Asian / Pacific Islander group, when all are exposed to the same levels of ozone concentration, during the year 2015. Here, the reported values represent posterior medians, and “CI” denotes the corresponding 95% posterior credible intervals. Moreover, as seen from Figure 9A, the relative effect size of the black group lies above the 87th percentile of the observed rates. These substantial differences highlight the persistent and disproportionate burden of asthma-related emergency department visits among the black population in California.

Figure 9B displays the posterior distributions of the regression coefficients corresponding to ozone concentration and year. We observe a marginally negative association between ozone and asthma-related ED visit rates. This suggests a weak inverse relationship, with higher ozone levels associated with slightly lower asthma-related ED visit rates. Although the association is not statistically strongly supported, it is consistent with patterns reported in previous studies (Zhu et al., 2003). This also contrasts with the commonly held assumption that a higher concentration of ozone increases the risk of asthma-related symptoms. More specifically, we find that an increase of 0.005 ppm of ozone concentration leads to a drop in asthma-related ED visit rates by a factor of 0.97 (CI: 0.94, 1). This relationship suggests seasonal interactions of ozone (Quick et al., 2015). Ozone levels also tend to be higher in rural inland areas (e.g., Central Valley), population density is lower, and ED utilization is lower due to access barriers, which does not necessarily imply a lower incidence of

asthma. Urban coastal areas (e.g., San Francisco Bay area, Los Angeles) may have lower ozone but higher asthma rates due to other pollutants, higher population density and reporting, and different healthcare-seeking behaviors. Our analysis shows that each year is associated with a 10% annual decrease in asthma-related ED visit rates, with the expected rate decreasing by a factor of 0.9 (CI: 0.89–0.91). This is consistent with the downward trend in Figure A2 of the Appendix.

We fit the model on 15,725 space-time coordinates using a stacked posterior that combines 54 candidate models. The entire procedure took roughly 90 minutes, a considerable improvement over traditional full Bayesian MCMC approaches, where a single iteration could take over 15 minutes. Furthermore, our regression inference framework accounts for the uncertainty in the estimated ozone concentrations by integrating the samples from their posterior distribution. This ensures that both the latent process and its downstream effects are quantified in a fully probabilistic and computationally efficient manner. Moreover, the use of periodic Fourier basis functions for modeling temporal trends enables smooth interpolation and prediction at any time point within the study window. This flexibility would be difficult to achieve using conventional discrete-time models. In addition, a model relying on monthly basis functions would be unable to provide predictions for months with no available data, which is a likely scenario given the irregularity frequently encountered in environmental monitoring records.

## 7. DISCUSSION

A key strength of our modular Bayesian framework is its capacity to jointly estimate ozone concentrations at arbitrary spatial and temporal resolutions, along with their association with an outcome of interest. Stacking of predictive densities enables us to obtain fully model-based uncertainty quantification for all model parameters by averaging over a collection of candidate models, each representing different specifications of process parameters. Future methodological directions may involve developing a multivariate areal time series model in place of the current linear regression framework, with the goal of capturing additional sources of variability and complex temporal-spatial dependencies inherent in the data. The data analysis presented in this article not only underscores existing racial disparities in health outcomes but also points to systemic inequities in environmental exposure, access to healthcare, and underlying social determinants of health. It emphasizes the need for targeted public health interventions and policies that address the structural drivers of asthma morbidity and improve health equity in racial and ethnic communities. Future data analysis may consider jointly estimating multiple exposure variables, such as NO<sub>2</sub>, PM<sub>2.5</sub>, and others, using multivariate spatial-temporal regression models to better account for potential correlations among pollutants and their combined effects on health outcomes.

## ACKNOWLEDGEMENT

This work used computational and storage services of the Hoffman2 Shared Cluster provided by UCLA Office of Advanced Research Computing's Research Technology Group. The authors were supported by research grants R01ES027027 from the National Institute of Environmental Health Sciences (NIEHS), R01GM148761 from the National Institute of General Medical Science (NIGMS)

and DMS2113778 from the Division of Mathematical Sciences of the National Science Foundation (NSF-DMS).

#### DATA AND CODE AVAILABILITY

The asthma data is openly available at the [CalHHS](#) website hosted by the California Department of Public Health (CDPH). The data on ozone measurements and geographic locations of the air quality monitoring stations are collected using the Air Quality and Meteorological Information System ([AQMIS](#)) database query tool of the California Air Resource Board (CARB). The code is provided as an R package, `spStackCOS`, available for download from the GitHub repository at <https://github.com/SPan-18/spStackCOS-dev>. Supplementary code to reproduce the results and findings presented in this article can be found at <https://github.com/SPan-18/AsthmaOzoneCA>.

#### CONFLICTS OF INTEREST

All authors declare that they have no conflicts of interest.

#### REFERENCES

- Milton Abramowitz and Irene A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*. Dover Publications, Inc., New York, 1965.
- Sudipto Banerjee, Brad P. Carlin, and Alan E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, 2nd edition, 2014. doi: 10.1201/b17115.
- M. J. Bayarri, J. O. Berger, and F. Liu. Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119 – 150, 2009. doi: 10.1214/09-BA404.
- Patrick Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 3rd edition, 1995.
- Leo Breiman. Stacked regressions. *Machine learning*, 24(1):49–64, 1996.
- California Breathing Asthma Program. Asthma emergency department visit rates, 2024. URL <https://www.cdph.ca.gov/Programs/CCDPHP/DEODC/EHIB/CPE/Pages/CaliforniaBreathing.aspx>. Accessed on 2025-04-18.
- Michela Cameletti, Virgilio Gómez-Rubio, and Marta Blangiardo. Bayesian modelling for spatially misaligned health and air pollution data through the INLA-SPDE approach. *Spatial Statistics*, 31:100353, 2019. doi: 10.1016/j.spasta.2019.04.001.
- Merlise A. Clyde and Edwin S. Iversen. Bayesian model averaging in the M-open framework. In *Bayesian Theory and Applications*, pages 483–498. Oxford University Press, 2013. ISBN 9780199695607. doi: 10.1093/acprof:oso/9780199695607.003.0024.
- Noel A.C. Cressie. Change of support and the modifiable areal unit problem. *Geographical Systems*, 3:159–180, 1996.
- Andrew Finley, Sudipto Banerjee, and Øyvind Hjelle. *MBA: Multilevel B-Spline Approximation*, 2024. URL <https://CRAN.R-project.org/package=MBA>. R package version 0.1-2.
- Andrew O. Finley, Sudipto Banerjee, and Alan E. Gelfand. spbayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical Software*, 63(13): 1–28, 2015. doi: 10.18637/jss.v063.i13.

- A E Gelfand, D K Dey, and H Chang. Model Determination using Predictive Distributions with Implementation via Sampling-Based Methods. In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, Dedicated to the memory of Morris H. DeGroot, 1931–1989*. Oxford University Press, 1992. ISBN 9780198522669. doi: 10.1093/oso/9780198522669.003.0009.
- Alan E. Gelfand, Li Zhu, and Bradley P. Carlin. On the change of support problem for spatio-temporal data. *Biostatistics*, 2(1):31–45, 2001. ISSN 1465-4644. doi: 10.1093/biostatistics/2.1.31.
- Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24:997–1016, 2013. doi: 10.1007/s11222-013-9416-2.
- Hamed Gharibi, Marcela R. Entwistle, Sandie Ha, Mariaelena Gonzalez, Paul Brown, Donald Schweizer, and Ricardo Cisneros. Ozone pollution and asthma emergency department visits in the central valley, California, USA, during June to September of 2015: a time-stratified case-crossover analysis. *Journal of Asthma*, 56(10):1037–1048, 2019. doi: 10.1080/02770903.2018.1523930.
- T. Gneiting and P. Guttorp. Continuous-parameter spatio-temporal processes. In A.E. Gelfand, P.J. Diggle, M. Fuentes, and P Guttorp, editors, *Handbook of Spatial Statistics*, Chapman & Hall CRC Handbooks of Modern Statistical Methods, pages 427–436. Taylor and Francis, 2010.
- Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437.
- Carol A Gotway and Linda J Young and. Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648, 2002. doi: 10.1198/016214502760047140.
- Jennifer A. Hoeting, David Madigan, Adrian E. Raftery, and Chris T. Volinsky. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. *Statistical Science*, 14(4):382 – 417, 1999. doi: 10.1214/ss/1009212519.
- Pierre E. Jacob, Lawrence M. Murray, Chris C. Holmes, and Christian P. Robert. Better together? statistical learning in models made of modules, 2017. URL <https://doi.org/10.48550/arXiv.1708.08719>.
- Tae Yoon Kim, Jeong Soo Park, and Dennis D Cox and. Fast algorithm for cross-validation of the best linear unbiased predictor. *Journal of Computational and Graphical Statistics*, 11(4): 823–835, 2002. doi: 10.1198/106186002826.
- Oswin Krause and Christian Igel. A more efficient rank-one covariance matrix update for evolution strategies. In *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*, FOGA ’15, page 129–136, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450334341. doi: 10.1145/2725494.2725496.
- Tri Le and Bertrand Clarke. A Bayes Interpretation of Stacking for  $\mathcal{M}$ -Complete and  $\mathcal{M}$ -Open Settings. *Bayesian Analysis*, 12(3):807 – 829, 2017. doi: 10.1214/16-BA1023.
- David Madigan, Adrian E Raftery, C Volinsky, and Jennifer Hoeting. Bayesian model averaging. In *Proceedings of the AAAI Workshop on Integrating Multiple Learned Models, Portland, OR*, pages 77–83, 1996.
- K. V. Mardia and C. R. Goodall. Spatial-temporal analysis of multivariate environmental monitoring data. *Multivariate Environmental Statistics*, 1993.

- Andrew S. Mugglin, Bradley P. Carlin, and Alan E. Gelfand. Fully model-based approaches for spatially misaligned data. *Journal of the American Statistical Association*, 95(451):877–887, 2000. doi: 10.1080/01621459.2000.10474279.
- Nicholas Nassikas, Keith Spangler, Neal Fann, Christopher G. Nolte, Patrick Dolwick, Tanya L. Spero, Perry Sheffield, and Gregory A. Wellenius. Ozone-related asthma emergency department visits in the us in a warming climate. *Environmental Research*, 183:109206, 2020. ISSN 0013-9351. doi: 10.1016/j.envres.2020.109206.
- Soumyakanti Pan and Sudipto Banerjee. *spStack: Bayesian Geostatistics Using Predictive Stacking*, 2024. URL <https://CRAN.R-project.org/package=spStack>. R package version 1.0.1.
- Soumyakanti Pan, Lu Zhang, Jonathan R. Bradley, and Sudipto Banerjee. Bayesian inference for spatial-temporal non-gaussian data using predictive stacking, 2025. URL <https://doi.org/10.48550/arXiv.2406.04655>.
- Martyn Plummer. Cuts in bayesian graphical models. *Statistics and Computing*, 25:37–43, 2014. doi: 10.1007/s11222-014-9503-z.
- Luca Presicce and Sudipto Banerjee. Bayesian transfer learning for artificially intelligent geospatial systems: A predictive stacking approach, 2025. URL <https://doi.org/10.48550/arXiv.2410.09504>.
- Harrison Quick, Sudipto Banerjee, and Bradley P. Carlin. Modeling temporal gradients in regionally aggregated California asthma hospitalization data. *The Annals of Applied Statistics*, 7(1):154 – 176, 2013. doi: 10.1214/12-AOAS600.
- Harrison Quick, Bradley P. Carlin, and Sudipto Banerjee. Heteroscedastic conditional auto-regression models for areally referenced temporal processes for analysing california asthma hospitalization data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 64(5):799–813, 2015. doi: 10.1111/rssc.12106.
- Paulo Justiniano Ribeiro Jr, Peter Diggle, Ole Christensen, Martin Schlather, Roger Bivand, and Brian Ripley. *geoR: Analysis of Geostatistical Data*, 2024. URL <https://CRAN.R-project.org/package=geoR>. R package version 1.9-4.
- Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian Inference for Latent Gaussian models by using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):319–392, 2009. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2008.00700.x.
- Wenpin Tang, Lu Zhang, and Sudipto Banerjee. On identifiability and consistency of the nugget in Gaussian spatial process models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(5):1044–1070, 2021. doi: <https://doi.org/10.1111/rssb.12472>. URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12472>.
- Rolf Turner. *deldir: Delaunay Triangulation and Dirichlet (Voronoi) Tessellation*, 2024. URL <https://CRAN.R-project.org/package=deldir>. R package version 2.0-4.
- Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing*, 27(5):1413–1432, 2017. ISSN 0960-3174. doi: 10.1007/s11222-016-9696-4.

- Aki Vehtari, Jonah Gabry, Måns Magnusson, Yuling Yao, Paul-Christian Bürkner, Topi Paananen, and Andrew Gelman. loo: Efficient leave-one-out cross-validation and waic for bayesian models, 2024a. URL <https://mc-stan.org/loo/>. R package version 2.8.0.
- Aki Vehtari, Daniel Simpson, Andrew Gelman, Yuling Yao, and Jonah Gabry. Pareto smoothed importance sampling. *Journal of Machine Learning Research*, 25(72):1–58, 2024b. URL <http://jmlr.org/papers/v25/19-556.html>.
- Tomoya Wakayama and Sudipto Banerjee. Process-based inference for spatial energetics using Bayesian predictive stacking, 2024. URL <https://doi.org/10.48550/arXiv.2405.09906>.
- Sumio Watanabe. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(116):3571–3594, 2010. URL <http://jmlr.org/papers/v11/watanabe10a.html>.
- David H Wolpert. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.
- Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman. Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, 13(3):917–1007, 2018.
- Hao Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261, 2004. doi: 10.1198/016214504000000241.
- Hao Zhang and Dale L. Zimmerman. Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, 92(4):921–936, 2005. ISSN 0006-3444. doi: 10.1093/biomet/92.4.921.
- Jin Zhang and Michael A. Stephens. A new and efficient estimation method for the generalized pareto distribution. *Technometrics*, 51(3):316–325, 2009. doi: 10.1198/tech.2009.08017.
- Lu Zhang, Wenpin Tang, and Sudipto Banerjee. Bayesian geostatistics using predictive stacking, 2024. URL <https://doi.org/10.48550/arXiv.2304.12414>.
- Ruiman Zhong and Paula Moraga. Bayesian hierarchical models for the combination of spatially misaligned data: A comparison of melding and downscaler approaches using INLA and SPDE. *Journal of Agricultural, Biological and Environmental Statistics*, 29:110–129, 2023. doi: 10.1007/s13253-023-00559-w.
- Li Zhu, Bradley P. Carlin, and Alan E. Gelfand. Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma er visits in atlanta. *Environmetrics*, 14(5):537–557, 2003. doi: <https://doi.org/10.1002/env.614>.

## Appendix to “Bayesian inference for spatial-temporal non-Gaussian data using predictive stacking”

### APPENDIX A. DESCRIPTIVE OVERVIEW OF THE DATASET

In this section, we present an exploratory data analysis of age-adjusted emergency department (ED) visit rates per 10,000 residents in California, across counties and racial groups. Figure A1 display annual county-level ED visit rates for each racial group throughout the study period 2015–22, revealing clear spatial and racial patterns. Notably, the black population consistently reports higher ED visit rates across the study period, with elevated rates concentrated in the coastal areas of northern California. In terms of data completeness, the white racial group has 2.1% missing data and the black group has 10.8% missing data, while the Asian/Pacific Islander and American Indian/Alaska Native groups exhibit high levels of missingness, with 34.1% and 59.3% missing respectively. On the other hand, at the county level, Los Angeles, Riverside, Sacramento and San Diego counties have complete ED visit data. The counties with the most missing data are Lassen, Siskiyou, Calaveras, Del Norte, Plumas, Mariposa, Modoc, Glenn, Nevada, and Trinity, each exhibiting approximately 50–60% missingness. These counties are predominantly rural and located in Northern California and the Sierra Nevada region. These areas are characterized by mountainous terrain and lower population density, which may contribute to challenges in healthcare access and data reporting completeness. These patterns underscore the importance of considering both spatial and demographic dimensions in analyzing the data.

Next, we plot the ED visit rates over time to examine temporal trends within each racial group. As shown in Figure A2, most groups exhibit a gradual decline in rates over the study period, suggesting potential improvements in underlying health conditions, access to preventive care, or changes in reporting and healthcare utilization. A strong declining trend is particularly evident for the white and the Asian/Pacific Islander groups while other groups show more variability. These patterns provide important context for interpreting cross-sectional differences and support the inclusion of temporal components in our modeling framework.

Next, we provide justification for the Gaussian modeling assumptions on the log-transformed ED visit rates. As shown in Figure A3A, the distribution of the raw ED visit rates exhibits strong positive skewness across all racial groups, with a long right tail driven by a small number of counties reporting exceptionally high rates. This pronounced skewness is inconsistent with the symmetry assumed under Gaussian models. Applying a log transformation helps stabilize the variance and mitigate the effect of extreme values, resulting in a distribution that is more symmetric and approximately Gaussian (see Figure A3B). The log-transformed rates exhibit reduced variability across the range of values and are more amenable to modeling under Gaussian assumptions.

Further, we plot the recorded measurements of ozone levels across time and observe clear seasonal fluctuations and periodic patterns, indicative of strong temporal structure (see Figure A4). These recurring trends suggest the need to account for seasonality when modeling the temporal dynamics of ozone exposure. In addition to temporal variation, we also find substantial spatial heterogeneity in ozone concentrations, highlighting the importance of incorporating spatial effects into the modeling framework. Together, these observations motivate the use of flexible spatial-temporal models

that can effectively capture both the periodic temporal behavior and the variability across different geographic locations.

## APPENDIX B. ADDITIONAL SIMULATION RESULTS

Calculation of leave-one-out predictive densities is central to evaluating the objective function used to determine optimal stacking weights. In the main article, we present two approaches for computing LOO predictive densities. The first is an exact method based on closed-form expressions, while the second is an approximate method known as Pareto Smoothed Importance Sampling (PSIS). To assess the accuracy of PSIS relative to the exact method, we conduct a simulation study using a spatial regression model. Suppose  $\{s_1, \dots, s_n\}$  denote a collection of  $n$  locations in the unit square, where we simulate spatially point-referenced responses  $X(s_i)$  for  $i = 1, \dots, n$ , using

$$X(s_i) = W\beta + Z(s_i) + \epsilon(s_i), \quad \epsilon(s_i) \stackrel{\text{ind}}{\sim} \mathcal{N}(0, \delta^2\sigma^2), \quad (\text{A1})$$

where  $W$  is  $n \times 2$  matrix comprising of an intercept and a predictor sampled from a standard normal distribution,  $\beta = (2, 5)^\top$ ,  $\epsilon(s_i)$  for each  $i$  denote independent measurement error,  $Z(s_i)$  are realizations of central spatial Gaussian process, given by  $Z(s) \sim \text{GP}(0, C_s(s, s'; \phi_s, \nu))$ , with covariance function  $C_s(\cdot, \cdot)$  as defined in (12). We take  $\sigma^2 = 0.4$ ,  $\phi_s = 2$ ,  $\nu = 0.5$  and  $\delta^2 = 1.5$ .

We assign priors to  $\beta$  and  $\sigma^2$  as discussed in Section 4.1 of the main article. We fix the values of  $\phi_s = 3.5$ ,  $\nu = 0.75$ , and  $\delta^2 = 1$ . Under this model and prior specification, we compute the exact leave-one-out predictive densities using closed-form expressions. To obtain the leave-one-out predictive densities via PSIS, we use posterior samples of  $\beta$  and  $Z = (Z(s_1), \dots, Z(s_n))^\top$  obtained using the R package `spStack` (Pan and Banerjee, 2024), and calculate the log-pointwise predictive densities (lppd). We then apply the `psis()` function from the R package `loo` (Vehtari et al., 2024a) to compute the leave-one-out predictive densities using stabilized importance weights. Figure A5 presents a comparison of the two methods, showing no visible differences in the resulting predictive densities. However, the exact method is considerably more computationally intensive than PSIS, demonstrating the latter's efficiency for large datasets.

## APPENDIX C. TECHNICAL DETAILS

**Proposition 1.** *For any  $a < b$  and  $c < d$ , suppose  $\tilde{C}_t(a, b, c, d; \phi) = \int_c^d \int_a^b C_t(t, t'; \phi_t) dt dt'$ , then*

(a) *for non-overlapping  $(a, b)$  and  $(c, d)$ , with  $a < b \leq c < d$ ,*

$$\tilde{C}_t(a, b, c, d; \phi_t) = \frac{1}{\phi_t^2} [F(a, d) + F(b, c) - F(a, c) - F(b, d)],$$

(b) *if  $c = a$  and  $d = b$ , then*

$$\tilde{C}_t(a, b, a, b; \phi_t) = \frac{2}{\phi_t^2} [\phi_t(b - a) + F(a, b) - 1],$$

(c) *if intervals  $(a, b)$  and  $(c, d)$  overlap, with  $a \leq c < b < d$  or  $a < c < b \leq d$ ,*

$$\tilde{C}_t(a, b, c, d; \phi_t) = \frac{1}{\phi_t^2} [2\phi_t(b - c) + F(a, d) + F(c, b) - F(a, c) - F(b, d)],$$

(d) if  $(a, b)$  is nested within  $(c, d)$ , i.e. either  $c \leq a < b < d$  or  $c < a < b \leq d$ ,

$$\tilde{C}_t(a, b, c, d; \phi_t) = \frac{1}{\phi_t^2} [2\phi_t(b-a) + F(a, d) + F(c, b) - F(c, a) - F(b, d)] ,$$

where  $F(a_1, a_2) = F(a_1, a_2; \phi_t) = \exp(-\phi_t(a_2 - a_1))$  for any  $a_1, a_2 \in \mathcal{T}$ .

*Proof.* (a) Without loss of generality, we assume  $a < b < c < d$ . The condition  $a < b < c < d$  denotes that the intervals  $(a, b)$  and  $(c, d)$  are disjoint and  $(a, b)$  lies on the left of  $(c, d)$ . Since,  $t \in (a, b)$  and  $t' \in (c, d)$ . This means  $t' \geq t$  holds true always, and hence,  $|t-t'| = t'-t$ . So, we evaluate the integral  $I_1$  as follows.

$$\begin{aligned} I_1 &= \int_c^d \int_a^b \exp(-\phi_t|t-t'|) dt dt' = \int_c^d \int_a^b \exp(-\phi_t(t'-t)) dt dt' \\ &= \frac{1}{\phi_t} \left( e^{\phi_t b} - e^{\phi_t a} \right) \int_c^d \exp(-\phi_t t') dt' = \frac{1}{\phi_t^2} \left( e^{\phi_t b} - e^{\phi_t a} \right) \left( e^{-\phi_t c} - e^{-\phi_t d} \right) \\ &= \frac{1}{\phi_t^2} \left( e^{\phi_t(b-c)} - e^{\phi_t(b-d)} - e^{\phi_t(a-c)} + e^{\phi_t(a-d)} \right) \\ &= \frac{1}{\phi_t^2} [F(b, c) - F(a, c) - F(b, d) + F(a, d)] \end{aligned}$$

The equality case  $b = c$  simply implies  $F(c, b) = 0$ .

(b1) If  $c = a$  and  $d = b$ , then we are dealing with the integration

$$\begin{aligned} \int_a^b \int_a^b e^{-\phi_t|t-t'|} dt dt' &= \int_a^b \left( \int_a^{t'} e^{-\phi_t|t-t'|} dt + \int_{t'}^b e^{-\phi_t|t-t'|} dt \right) dt' \\ &= \int_a^b \left( \int_a^{t'} e^{-\phi_t(t'-t)} dt + \int_{t'}^b e^{-\phi_t(t-t')} dt \right) dt' \\ &= \int_a^b \left( e^{-\phi_t t'} \cdot \frac{1}{\phi_t} \left( e^{\phi_t t'} - e^{\phi_t a} \right) + e^{\phi_t t'} \frac{1}{\phi_t} \left( e^{-\phi_t t'} - e^{-\phi_t b} \right) \right) dt' \\ &= \frac{1}{\phi_t} \int_a^b 1 - e^{\phi_t(a-t')} + 1 - e^{\phi_t(t'-b)} dt' \\ &= \frac{2}{\phi_t} (b-a) - \frac{1}{\phi_t} e^{\phi_t a} \int_a^b e^{-\phi_t t'} dt' - \frac{1}{\phi_t} e^{-\phi_t b} \int_a^b e^{\phi_t t'} dt' \\ &= \frac{2}{\phi_t} (b-a) - \frac{1}{\phi_t^2} e^{\phi_t a} \left( e^{-\phi_t a} - e^{-\phi_t b} \right) - \frac{1}{\phi_t^2} e^{-\phi_t b} \left( e^{\phi_t b} - e^{\phi_t a} \right) \\ &= \frac{2}{\phi_t} (b-a) - \frac{1}{\phi_t^2} \left( 1 - e^{\phi_t(a-b)} + 1 - e^{\phi_t(a-b)} \right) \\ &= \frac{2}{\phi_t^2} \left( \phi_t(b-a) + e^{-\phi_t(b-a)} - 1 \right) \\ &= \frac{2}{\phi_t^2} [\phi_t(b-a) + F(a, b) - 1] . \end{aligned}$$

(b2) In this case, we have  $a < c < b < d$ . The integration relies on splitting the integral into parts depending on the sign of the term  $|t - t'|$ , where  $t \in (a, b)$  and  $t' \in (c, d)$ . We split

the integral into three double integrals  $I_1$ ,  $I_2$  and  $I_3$ , where  $I_1$  is on  $(c, d) \times (a, c)$ ,  $I_2$  is on  $(c, b) \times (c, b)$ , and  $I_3$  is on  $(b, d) \times (c, b)$ .

$$\begin{aligned} \int_c^d \int_a^b e^{-\phi_t|t-t'|} dt dt' &= \int_c^d \int_a^c e^{-\phi_t(t'-t)} dt dt' + \int_c^b \int_c^b e^{-\phi_t|t-t'|} dt dt' + \int_b^d \int_c^b e^{-\phi_t(t'-t)} dt dt' \\ &= I_1 + I_2 + I_3 , \end{aligned}$$

In  $I_1$ ,  $t \leq t'$ , so we have  $|t - t'| = t' - t$ . In  $I_2$ ,  $t - t'$  may be both positive and negative. And, for  $I_3$ ,  $t \leq t'$ , so we have  $|t - t'| = t' - t$ . We tackle each of these integrals separately.

$$\begin{aligned} I_1 &= \int_c^d \int_a^c e^{-\phi_t(t'-t)} dt dt' = \int_c^d e^{-\phi_t t'} \left( \int_a^c e^{\phi_t t} dt \right) dt' \\ &= \frac{1}{\phi_t} \left( e^{\phi_t c} - e^{\phi_t a} \right) \int_c^d e^{-\phi_t t'} dt' = \frac{1}{\phi_t^2} \left( e^{\phi_t c} - e^{\phi_t a} \right) \left( e^{-\phi_t c} - e^{-\phi_t d} \right) \\ &= \frac{1}{\phi_t^2} \left( 1 - e^{\phi_t(c-d)} - e^{\phi_t(a-c)} + e^{\phi_t(a-d)} \right) . \end{aligned}$$

Similarly, we find the integral  $I_3$ , which is very similar to  $I_1$ .

$$\begin{aligned} I_3 &= \int_b^d \int_c^b e^{-\phi_t(t'-t)} dt dt' = \int_b^d e^{-\phi_t t'} \left( \int_c^b e^{\phi_t t} dt \right) dt' \\ &= \frac{1}{\phi_t^2} \left( e^{\phi_t b} - e^{\phi_t c} \right) \left( e^{-\phi_t b} - e^{-\phi_t d} \right) \\ &= \frac{1}{\phi_t^2} \left( 1 - e^{\phi_t(b-d)} - e^{\phi_t(c-b)} + e^{\phi_t(c-d)} \right) \end{aligned}$$

Following part (b) of Proposition 1, we have

$$I_2 = \int_c^b \int_c^b e^{-\phi_t|t-t'|} dt dt' = \frac{2}{\phi_t^2} \left( \phi_t(b-c) + e^{-\phi_t(b-c)} - 1 \right)$$

Combining  $I_1$ ,  $I_2$  and  $I_3$ , we have

$$\begin{aligned} I_1 + I_2 + I_3 &= \frac{1}{\phi_t^2} \left( 1 - e^{\phi_t(c-d)} - e^{\phi_t(a-c)} + e^{\phi_t(a-d)} \right) + \frac{2}{\phi_t^2} \left( \phi_t(b-c) + e^{-\phi_t(b-c)} - 1 \right) \\ &\quad + \frac{1}{\phi_t^2} \left( 1 - e^{\phi_t(b-d)} - e^{\phi_t(c-b)} + e^{\phi_t(c-d)} \right) \\ &= \frac{1}{\phi_t^2} \left( 2 - e^{\phi_t(a-c)} + e^{\phi_t(a-d)} - e^{\phi_t(b-d)} - e^{\phi_t(c-b)} \right) \\ &\quad + \frac{2}{\phi_t^2} \left( \phi_t(b-c) + e^{-\phi_t(b-c)} \right) - \frac{2}{\phi_t^2} \\ &= \frac{2}{\phi_t} (b-c) + \frac{1}{\phi_t^2} \left( e^{\phi_t(c-b)} - e^{\phi_t(a-c)} + e^{\phi_t(a-d)} - e^{\phi_t(b-d)} \right) \\ &= \frac{1}{\phi_t^2} \left[ 2\phi_t(b-c) + e^{-\phi_t(b-c)} - e^{-\phi_t(c-a)} - e^{-\phi_t(d-b)} + e^{-\phi_t(d-a)} \right] \\ &= \frac{1}{\phi_t^2} [2\phi_t(b-c) + F(c,b) - F(a,c) - F(b,d) + F(a,d)] \end{aligned}$$

(c) In this case the interval  $(a, b)$  is nested within the interval  $(c, d)$ . Hence, we have  $c < a < b < d$ . We split the integral with respect to  $t'$  on  $(c, d)$  into three - one over  $(c, a)$ , one over  $(a, b)$  and the last over  $(b, d)$ . Hence, we have

$$\begin{aligned} \int_c^d \int_a^b e^{-\phi_t|t-t'|} dt dt' &= \int_c^a \int_a^b e^{-\phi_t|t-t'|} dt dt' + \int_a^b \int_a^b e^{-\phi_t|t-t'|} dt dt' + \int_b^d \int_a^b e^{-\phi_t|t-t'|} dt dt' \\ &= I_1 + I_2 + I_3 . \end{aligned}$$

We evaluate each of the integrals  $I_1$ ,  $I_2$  and  $I_3$  separately. First, we find

$$\begin{aligned} I_1 &= \int_c^a \int_a^b e^{-\phi_t(t-t')} dt dt' = \int_c^a e^{\phi_t t'} \left( \int_a^b e^{-\phi_t t} dt \right) dt' \\ &= \frac{1}{\phi_t^2} (e^{-\phi_t a} - e^{-\phi_t b}) (e^{\phi_t a} - e^{\phi_t c}) \\ &= \frac{1}{\phi_t^2} (1 - e^{\phi_t(c-a)} - e^{\phi_t(a-b)} + e^{\phi_t(c-b)}) . \end{aligned}$$

Similarly, we find the integral  $I_3$  similar to  $I_1$ .

$$\begin{aligned} I_3 &= \int_b^d \int_a^b e^{-\phi_t(t'-t)} dt dt' = \int_b^d e^{-\phi_t t'} \left( \int_a^b e^{\phi_t t} dt \right) dt' \\ &= \frac{1}{\phi_t^2} (e^{\phi_t b} - e^{\phi_t a}) (e^{-\phi_t b} - e^{-\phi_t d}) \\ &= \frac{1}{\phi_t^2} (1 - e^{\phi_t(b-d)} - e^{\phi_t(a-b)} + e^{\phi_t(a-d)}) \end{aligned}$$

Following part (b) of Proposition 1, we find the integral  $I_2$ . Next, we combine  $I_1$ ,  $I_2$  and  $I_3$ .

$$\begin{aligned} I_1 + I_2 + I_3 &= \frac{1}{\phi_t^2} (1 - e^{\phi_t(c-a)} - e^{\phi_t(a-b)} + e^{\phi_t(c-b)}) + \frac{2}{\phi_t^2} (\phi_t(b-a) + e^{-\phi_t(b-a)} - 1) \\ &\quad + \frac{1}{\phi_t^2} (1 - e^{\phi_t(b-d)} - e^{\phi_t(a-b)} + e^{\phi_t(a-d)}) \\ &= \frac{1}{\phi_t^2} (2 - 2e^{\phi_t(a-b)} - e^{\phi_t(c-a)} + e^{\phi_t(c-b)} - e^{\phi_t(b-d)} + e^{\phi_t(a-d)}) \\ &\quad + \frac{2}{\phi_t} (b-a) + \frac{1}{\phi_t^2} (2e^{\phi_t(a-b)} - 2) \\ &= \frac{1}{\phi_t^2} [2\phi_t(b-a) + e^{-\phi_t(b-c)} - e^{-\phi_t(a-c)} - e^{-\phi_t(d-b)} + e^{-\phi_t(d-a)}] \\ &= \frac{1}{\phi_t^2} [2\phi_t(b-a) + F(c,b) - F(c,a) - F(b,d) + F(a,d)] \end{aligned}$$

□

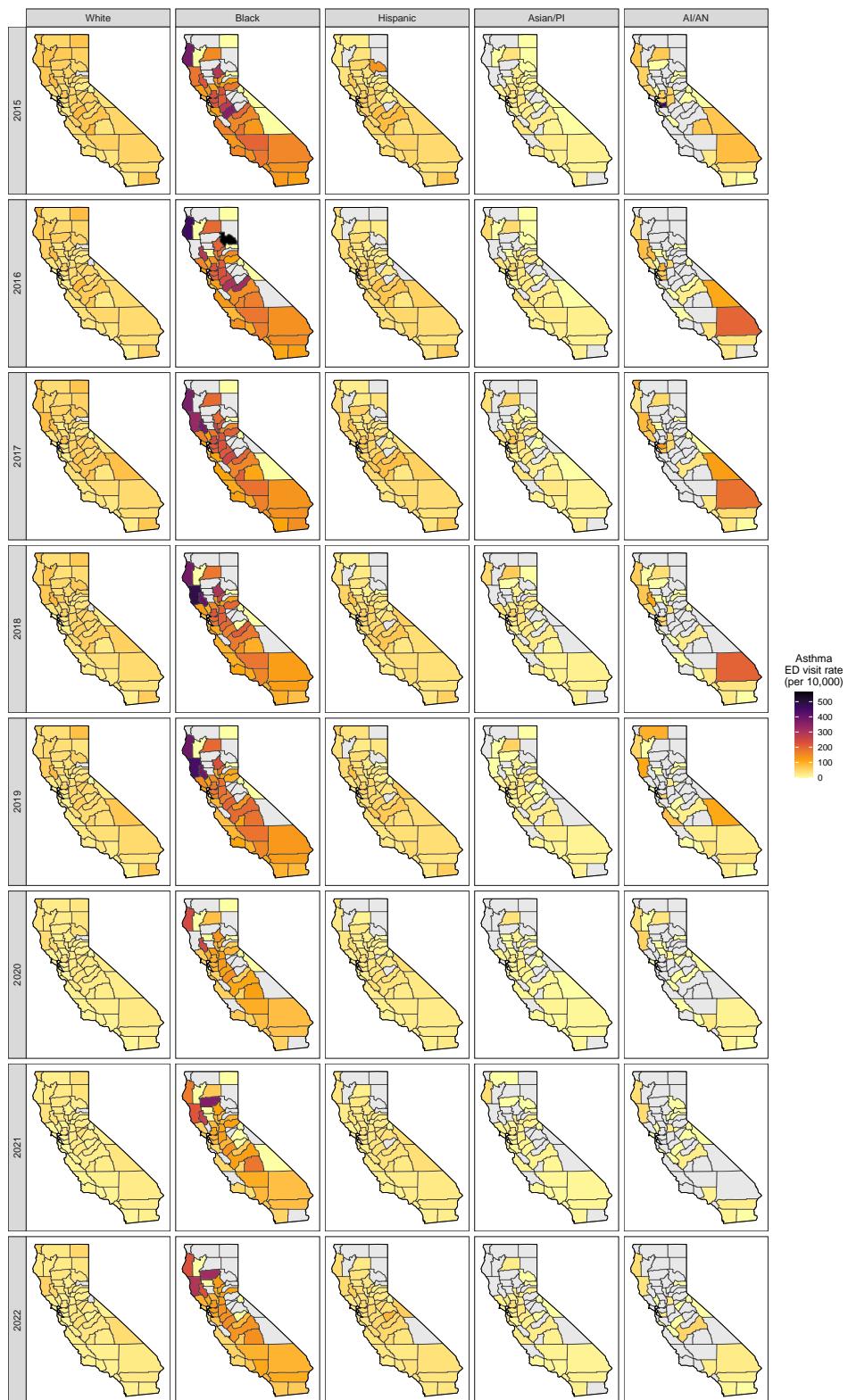


FIGURE A1. Asthma ED visit rates by race 2015-22

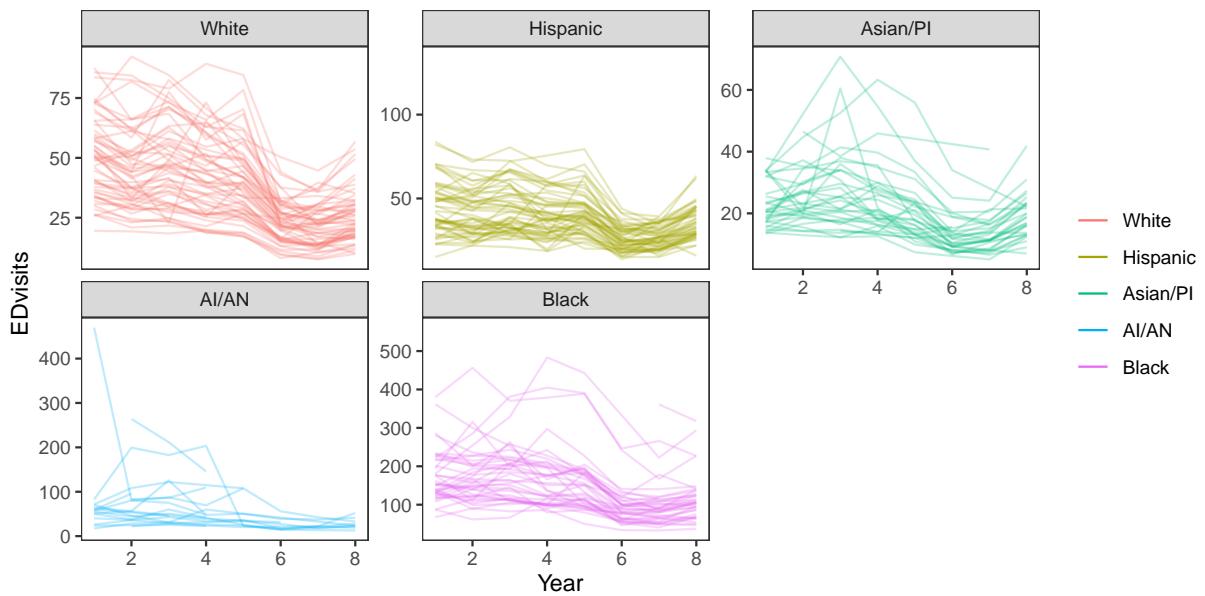


FIGURE A2. Annual trend in asthma-related ED visit rates per 10,000 residents for different racial groups.

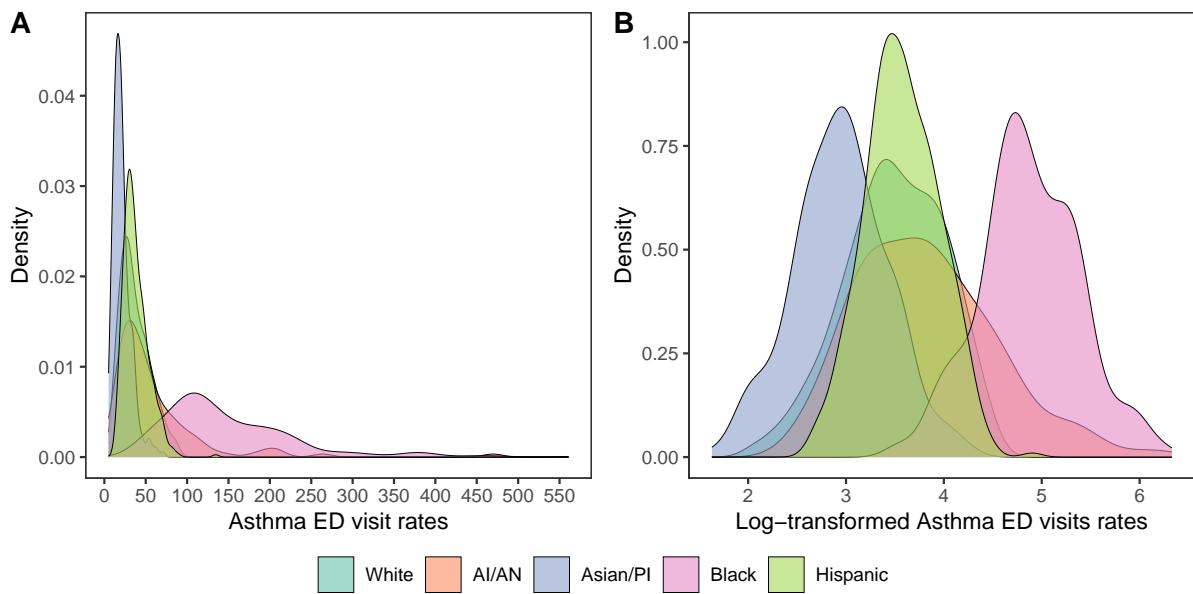


FIGURE A3. Race-specific estimated density of county-level age-adjusted rates (per 10,000) of asthma ED visits during 2015-2022.

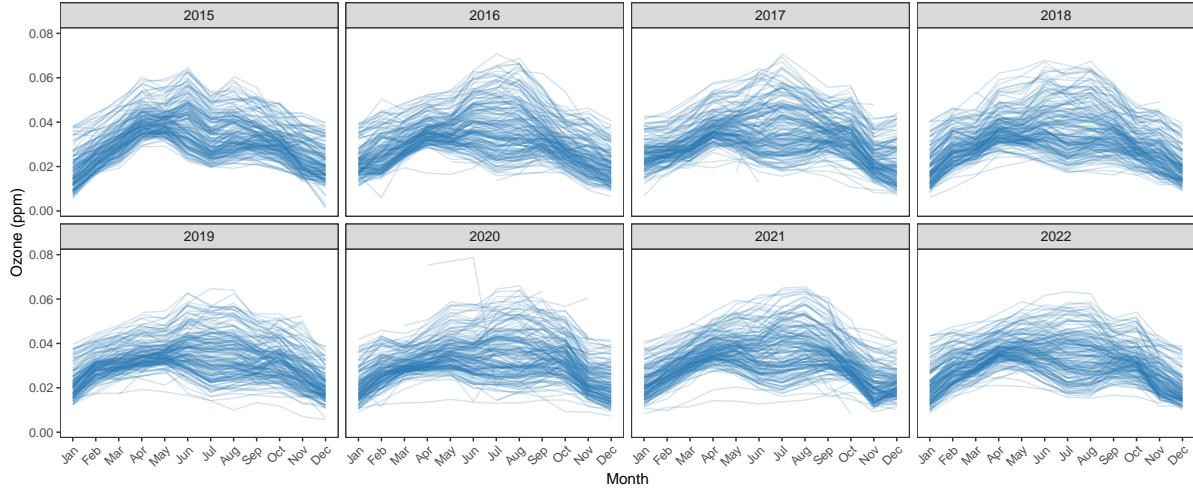


FIGURE A4. Monthly average ozone concentrations (in parts per million, ppm) recorded at various ozone monitoring sites across California from 2015 to 2022. A clear seasonal pattern is evident, with ozone levels peaking as well as exhibiting higher variability during the warmer months (May to October) and reaches troughs during the colder months (November to April).

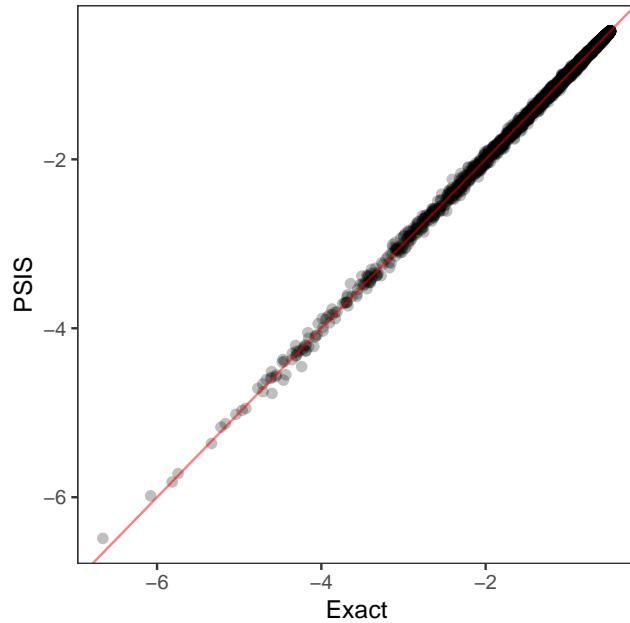


FIGURE A5. Comparison of leave-one-out predictive densities computed using exact closed form expression given by (17) and Pareto smoothed importance sampling (PSIS) for a spatial regression model on a simulated dataset of sample size 5000.