MODELING TEMPORAL GRADIENTS IN REGIONALLY AGGREGATED CALIFORNIA ASTHMA HOSPITALIZATION DATA

BY HARRISON QUICK, SUDIPTO BANERJEE AND BRADLEY P. CARLIN

Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota 55455

Advances in Geographical Information Systems (GIS) have led to enormous recent burgeoning of spatial-temporal databases and associated statistical modeling. Here we depart from the rather rich literature in space-time modeling by considering the setting where space is discrete (e.g. aggregated data over regions), but time is continuous. Our major objective in this application is to carry out inference on gradients of a temporal process in our dataset of monthly county level asthma hospitalization rates in the state of California, while at the same time accounting for spatial similarities of the temporal process across neighboring counties. Rather than use parametric forms to model time, we opt for a more flexible stochastic process embedded within a dynamic Markov random field framework. Through the cross-covariance function we can ensure that the temporal process realizations are mean square differentiable, and may thus carry out inference on temporal gradients in a posterior predictive fashion. We use this approach to evaluate temporal gradients where we are concerned with temporal changes in the residual and fitted rate curves after accounting for seasonality, spatiotemporal ozone levels, and several spatially-resolved important sociodemographic covariates.

1. Introduction. Technological advances in spatially-enabled sensor networks, and geospatial information storage, analysis, and distribution systems have led to a burgeoning of spatial-temporal databases. Accounting for associations across space and time constitute a routine component in analyzing geographically and temporally referenced datasets. The inference garnered through these analyses often supports decisions with important scientific implications, and it is therefore critical to accurately assess inferential uncertainty. The obstacle for researchers is increasingly not access to the right data, but rather implementing appropriate statistical methods and software.

There is a considerable literature in spatio-temporal modeling; see, for example, the recent book

Keywords and phrases: Gaussian process, Gradients, Markov chain Monte Carlo, Spatial process models, Spatially associated functional data
by Cressie and Wikle (2011) and references therein. Space-time modeling can broadly be classified as considering one of the following four settings: (a) space is viewed as continuous, but time is taken to be discrete, (b) space and time are both continuous, (c) space and time are both discrete, and (d) space is viewed as discrete, but time is taken to be continuous. Almost exclusively, the existing literature considers the first three settings. Perhaps the most pervasive case is the first. Here, the data are regarded as a time series of spatial process realizations. Early approaches include the STARMA (Pfeifer and Deutsch, 1980, a, b) and STARMAX (Stoffer, 1986) models, which add spatial covariance structure to standard time series models. Handcock and Wallis (1994) employ stationary Gaussian process models with an AR(1) model for the time series at each location to study global warming. Building upon previous work in the setting of dynamic models by West and Harrison (1997), several authors, including Stroud et al. (2001) and Gelfand et al. (2005), proposed dynamic frameworks to model residual spatial and temporal dependence.

When space and time are both viewed as continuous, the preferred approach is to construct stochastic processes using space-time covariance functions. Gneiting (2002) built upon earlier work by Cressie and Huang (1999) to propose general classes of nonseparable, stationary covariance functions that allow for space-time interaction terms for spatiotemporal random processes. Stein (2005) considered a variety of properties of space-time covariance functions and how these were related to process spatial-temporal interactions.

Finally, in settings where both space and time are discrete there has been much spatiotemporal modeling based on a Markov random field (MRF) structure in the form of conditionally autoregressive (CAR) specifications. See, for example, Waller et al. (1997), who developed such models in the service of disease mapping and Gelfand et al. (1998), whose interest was in single family home sales. Pace et al. (2000) work with simultaneous autoregressive (SAR) models extending them to allow temporal neighbors as well as spatial neighbors. Gelfand et al. (2004) attempt a survey in the context of real estate applications.

Our manuscript departs from this rich literature by considering the setting where space is discrete and time is continuous. This can be envisioned when, for instance, we have a collection of $N_t$ functions of time over $N_s$ regions, but the functions are posited to be spatially associated. That is, functions arising from neighboring regions are believed to resemble each other. The functional data analysis literature (Ramsay and Silverman, 1997, and references therein) deals almost exclusively with kernel smoothers and roughness-penalty type (spline) models. Baladandayuthapani et al. (2008) consider spatially correlated functional data modeling for point-referenced data by treating
space as continuous. A recent review by Delicado et al. (2010) reveals that spatially associated functional modeling of time has received little attention, especially for regionally aggregated data. This is unfortunate, especially given the datasets we encounter here (see Section 2 below).

As such, we propose a rich class of Bayesian space-time models based upon a dynamic MRF that evolve continuously over time. This accommodates spatial processes that are posited to be spatially indexed over a geographical map with a well-defined system of neighbors. This continuous temporal evolution sets our current article apart from the existing literature. Rather than modeling time using simple parametric forms, as is often done in longitudinal contexts, we employ a stochastic process, enhancing the model’s adaptability to the data.

The modeling also allows us to subsequently carry out inference on temporal gradients, that is, the rate of change of the underlying process over time. We show how such inference can be carried out in fully model-based fashion using exact posterior predictive distributions for the gradients at any arbitrary time point. The smoothness implications for the underlying process in this context are obvious. We deploy a mean square differentiable Gaussian process that provides a tractable gradient (or derivative) process to help us achieve these inferential goals.

The remainder of the manuscript is structured as follows. Section 2 describes the dataset that motivates our methodology and which we analyze in depth. Section 3 outlines a class of dynamic MRF indexed continuously over time. Section 4 provides details on the Bayesian hierarchical models that emerge from our rich space-time structures, while Section 5 derives the posterior predictive inferential procedure for temporal gradient process, verified via simulation in Section 6. Section 7 describes the detailed analysis of our dataset, while Section 8 summarizes and concludes.

2. Data. Our dataset consists of asthma hospitalization rates in the state of California. According to the California Department of Health Services (2003), millions of residents of California suffer from asthma or asthma-like symptoms. As many studies have indicated (e.g. English et al. 1998), asthma rates are related to, among other things, pollution levels and socioeconomic status (SES)—two variables that likely induce a spatiotemporal distribution on such rates. Weather and climate also likely play a role, as cold air can trigger asthma symptoms.

The data we will analyze were collected daily from 1991 to 2008 at the county level, counting all discharges where asthma was the primary diagnosis. Due to confidentiality, data for days with between one and four hospitalizations are missing. To remedy this, county-specific values for these days are imputed using a method similar to Besag’s iterated conditional modes method (Besag,
We attempt to capture the effect of socioeconomic status by including population density in our model, using data from the 2000 U.S. Census and land area measurements from the National Association of Counties. To account for pollution, we use data from the California Environmental Protection Agency regarding the number of days in each month exceeding the 8 hour state standard for an acceptable level of ozone. Because our ozone data is compiled at the air basin level, county-specific values are calculated by taking the maximum value of all air basins that the county belonged to. Generally, ozone levels are highest during the summer months, with the highest values in southern California and the Central Valley region, and show little variation between years. As hospitalization rates are higher among youth and the black population, county-level covariates for percent under 18 and percent black are also included. These demographic covariates both have their highest values in southern California, though counties in the Central Valley region also have larger black populations. Rates per 1,000 residents are computed by dividing the monthly counts by the county’s 2000 Census population and multiplying by 1,000; the conversion from counts to rates for the purpose of fitting Gaussian spatiotemporal models is common in literature (see, for instance, Short et al., 2002). Here our goal is to detect temporal changes in the residuals that remain after the covariates are accounted for; significant changes may correspond to changes in the spatiotemporal covariates still missing from our model. We can also use this information to learn about changes in the fitted curve, which is also informed by our spatiotemporally-resolved ozone data.

3. Areally referenced temporal processes. As mentioned above, our methodological contribution is a modeling framework for areally referenced outcomes that, it can be reasonably assumed, arise from an underlying stochastic process continuous over time. To be specific, consider a map of a geographical region comprising \( N_s \) regions that are delineated by well-defined boundaries, and let \( Y_i(t) \) be the outcome arising from region \( i \) at time \( t \). For every region \( i \), we believe that \( Y_i(t) \) exists, at least conceptually, at every time point. However, the observations are collected not continuously but at discrete time points, say \( T = \{t_1,t_2,\ldots,t_{N_t}\} \). For the time being, we will assume that the data comes from the same set of time points in \( T \) for each region. This is not necessary for the ensuing development, but will facilitate the notation.
Fig 1. Raw asthma hospitalization rates, by year

A spatial random effect model for our data assumes

\[ Y_i(t) = \mu_i(t) + Z_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \sim \text{ind } N(0, \tau_i^2) \quad \text{for } i = 1, 2, \ldots, N_s, \]

where \( \mu_i(t) \) captures large scale variation or trends, for example using a regression model, and \( Z_i(t) \) is an underlying areally-referenced stochastic process over time that captures smaller-scale variations in the time scale while also accommodating spatial associations. Each region also has its own variance component, \( \tau_i^2 \), which captures residual variation not captured by the other components.

The process \( Z_i(t) \) specifies the probability distribution of correlated space-time random effects while treating space as discrete and time as continuous. We seek a specification that will allow temporal processes from neighboring regions to be more alike than from non-neighbors. As regards spatial associations, we will respect the discreteness inherent in the aggregated outcome. Rather than model an underlying response surface continuously over the region of interest, we want to treat the \( Z_i(t) \)’s as functions of time that are smoothed across neighbors.

The neighborhood structure arises from a discrete topology comprising a list of neighbors for each region. This is described using an \( N_s \times N_s \) adjacency matrix \( W = \{w_{ij}\} \), where \( w_{ij} = 0 \) if regions \( i \) and \( j \) are not neighbors and \( w_{ij} = c \neq 0 \) when regions \( i \) and \( j \) are neighbors, denoted by \( i \sim j \). By convention, the diagonal elements of \( W \) are all zero. To account for spatial association in the \( Z_i(t) \)’s, a temporally evolving MRF for the areal units at any arbitrary time point \( t \) specifies the full conditional distribution for \( Z_i(t) \) as depending only upon the neighbors of region \( i \),

\[ p(Z_i(t) \mid \{Z_{j \neq i}(t)\}) \sim N \left( \sum_{j \sim i} \alpha \frac{w_{ij}}{w_i} Z_j(t), \frac{\sigma^2}{w_i} \right), \]
where \( w_{i+} = \sum_{j=i} w_{ij} = \sigma^2 > 0 \), and \( \alpha \) is a propriety parameter described below. This means that the \( N_s \times 1 \) vector \( \mathbf{Z}(t) = (Z_1(t), Z_2(t), \ldots, Z_{N_s}(t))^T \) follows a multivariate normal distribution with zero mean and a precision matrix \( \frac{1}{\sigma^2} (D - \alpha W) \), where \( D \) is a diagonal matrix with \( w_{i+} \) as its \( i \)-th diagonal elements. The precision matrix is invertible as long as \( \alpha \in (1/\lambda(1), 1/\lambda(n)) \), where \( \lambda(1) \) (which can be shown to be negative) and \( \lambda(n) \) are the smallest (i.e., most negative) and largest eigenvalues of \( D^{-1/2} WD^{-1/2} \), and this yields a proper distribution for \( \mathbf{Z}(t) \) at each timepoint \( t \).

The MRF in (2) does not allow temporal dependence; the \( \mathbf{Z}(t) \)'s are independently and identically distributed as \( N(0, \sigma^2 (D - \alpha W)^{-1}) \). We could allow time-varying parameters \( \sigma_t^2 \) and \( \alpha_t \) so that \( \mathbf{Z}(t) \sim N(0, \sigma_t^2 (D - \alpha_t W)^{-1}) \) for every \( t \). If time were treated discretely, then we could envision dynamic autoregressive priors for these time-varying parameters, or some transformations thereof. However, there are two reasons why we do not pursue this further. First, we do not consider time as discrete because that would preclude inference on temporal gradients, which, as we have mentioned, is a major objective here. Second, time-varying hyperparameters, especially the \( \alpha_t \)'s, in MRF models are usually weakly identified by the data; they permit very little prior-to-posterior learning and often lead to over-parametrized models that impair predictive performance over time.

Here we prefer to jointly build spatial-temporal associations into the model using a multivariate process specification for \( \mathbf{Z}(t) \). A highly flexible and computationally tractable option is to assume that \( \mathbf{Z}(t) \) is a zero-centered multivariate Gaussian process, \( GP(0, K_Z(\cdot, \cdot)) \), where the cross-covariance matrix function (e.g., Cressie, 1993) \( K_Z(t, u) = \text{cov}\{ \mathbf{Z}(t), \mathbf{Z}(u) \} \) is defined to be the \( N_s \times N_s \) matrix with \((i, j)\)-th entry \( \text{cov}\{ Z_i(t), Z_j(u) \} \) for any \((t, u) \in \mathbb{R}^+ \times \mathbb{R}^+ \). Thus, for any two positive real numbers \( t \) and \( u \), \( K_Z(t, u) \) is an \( N_s \times N_s \) matrix with \((i, j)\)-th element given by the covariance between \( Z_i(t) \) and \( Z_j(u) \).

The cross-covariance matrix plays a central role in ensuring a valid stochastic process as it completely determines the joint dispersion structure implied by the spatial process. It need not itself be symmetric or positive definite but must satisfy the following two conditions: (i) \( K_Z(t, u) = K_Z(u, t)^T \) for any \((t, u) \in \mathbb{R}^+ \times \mathbb{R}^+ \), and (ii) \( \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \mathbf{u}_i^T K_Z(t_i, t_j) \mathbf{u}_j > 0 \) for all \( \mathbf{u}_i, \mathbf{u}_j \in \mathbb{R}^{N_s} \setminus \{ \mathbf{0} \} \). The first condition follows immediately from the symmetry of covariances. For any \( N_t \) and any arbitrary collection of timepoints \( \mathcal{T} = \{t_1, t_2, \ldots, t_{N_t}\} \), the \( N_s N_t \times 1 \) vector of realizations \( \mathbf{Z} = (\mathbf{Z}(t_1))^T, \ldots, (\mathbf{Z}(t_{N_t}))^T \) will have the variance-covariance matrix given by \( \Sigma_Z \), an \( N_s N_t \times N_s N_t \) block matrix whose \((i, j)\)-th block is the cross-covariance matrix \( K_Z(t_i, t_j) \). Conditions (i) and (ii) on the cross-covariance matrix ensure that \( \Sigma_Z \) is symmetric and positive-definite. In the limiting sense, as \( u \to t \), \( K_Z(t, u) \) must approach the symmetric positive definite matrix \( K_Z(t, t) \), which represents
the spatial association of the $Z_i(t)$’s at a given time point $t$. These multivariate processes are **stationary** when the cross-covariances are functions of the separation between the time-points, in which case we write $K_Z(t, u) = K_Z(\Delta)$, and **isotropic** when $K_Z(t, u) = K_Z(|\Delta|)$, where $\Delta = t - u$.

Characterizing cross-covariance matrix functions $K_Z(t, u)$ that ensure positive-definiteness of $\Sigma_Z$ is not immediate, requiring that for an arbitrary number and choice of locations the resulting $\Sigma_Z$ be positive definite. A fundamental characterization theorem for cross-covariance matrix functions (Yaglom, 1987) says that real-valued functions, say $K_{ij}(t)$, will form the elements of a valid cross-covariance matrix $K_Z(t) = \{K_{ij}(t)\}_{i,j=1}^{N_s}$ if and only if each $K_{ij}(t)$ has the **cross-spectral representation** $K_{ij}(t) = \int \exp(2\pi iut)d(K_{ij}(u))$, where $i = \sqrt{-1}$, with respect to a positive definite measure $K(\cdot)$, i.e. the cross-spectral matrix $M(B) = \{K_{ij}(B)\}_{i,j=1}^{N_s}$ is positive definite for any Borel subset $B \subseteq \mathbb{R}^d$. The cross-spectral representation provides a very general representation for cross-covariance functions. Matters simplify when $K_{ij}(t)$ is assumed to be square-integrable, ensuring that a spectral density function $K_{ij}(t)$ exists such that $d(K_{ij}(u)) = k_{ij}(u)du$. Now, one simply needs to ensure that $\{k_{ij}(u)\}_{i,j=1}^{N_s}$ are positive definite for all $t \in \mathbb{R}^d$. Corollaries of the above representation lead to the approaches proposed by Gaspari and Cohn (1999) for constructing valid cross-covariance functions as convolutions of covariance functions of stationary random fields.

To ensure valid joint distributions for process realizations, we use a constructive approach based on latent processes. We assume that $Z(t)$ arises as a (possibly temporally-varying) linear transformation $Z(t) = A(t)v(t)$ of a simpler process $v(t) = (v_1(t), v_2(t), \ldots, v_{N_s}(t))^T$ where the $v_i(t)$’s are univariate temporal processes, independent of each other, and with unit variances. The cross-covariance matrix for $v(t)$, say $K_v(t, u)$, thus has a simple diagonal form and $K_Z(t, u) = A(t)K_v(t, u)A(u)^T$. The dispersion matrix for $Z$ is $\Sigma_Z = A\Sigma_vA^T$ with $A$ being a block-diagonal matrix with $A(t_j)$’s as blocks, and $\Sigma_v$ is the dispersion matrix constructed from $K_v(t, u)$. Constructing simple valid cross-covariance functions for $v(t)$ automatically ensures valid probability models for $Z(t)$. Also note that for $t = u$, $K_v(t, t)$ is the identity matrix so that $K_Z(t, t) = A(t)A(t)^T$ and $A(t)$ is a square-root (e.g. obtained from the triangular Cholesky factorization) of the cross-covariance matrix at time $t$.

The above framework subsumes several simpler and more intuitive specifications. One particular specification that we pursue here assumes that each $v_i(t)$ follows a stationary Gaussian Process $GP(0, \rho(\cdot; \cdot; \phi))$, where $\rho(\cdot; \cdot; \phi)$ is a positive definite correlation function parametrized by $\phi$ (e.g. Stein, 1999), so that $\text{cov}(v_i(t), v_i(u)) = \rho(t, u; \phi)$ for every $i = 1, 2, \ldots, N_s$ for all non-negative real numbers $t$ and $u$. Since the $v_i(t)$ are independent across $i$, $\text{cov}\{v_i(t), v_j(u)\} = 0$ for $i \neq j$.

The cross-covariance matrix for $Z(t)$ becomes $K_Z(t, u) = \rho(t, u; \phi)A(t)A(u)^T$. If we further
assume that \( A(t) = A \) is constant over time, then the process \( Z(t) \) is stationary if and only if the \( v(t) \) is stationary. Further, we obtain a separable specification, so that \( K_Z(t, u) = \rho(t, u; \phi)AA^T \).

Letting \( A \) be some square-root (e.g. Cholesky) of the \( N_s \times N_s \) dispersion matrix \( \sigma^2(D - \alpha W)^{-1} \) and \( R(\phi) \) be the \( N_t \times N_t \) temporal correlation matrix having \((i, j)\)-th element \( \rho(t_i, t_j; \phi) \) yields

\[
\text{Model 1: } K_Z(t, u) = \sigma^2 \rho(t, u; \phi)(D - \alpha W)^{-1} \quad \text{and } \Sigma_Z = R(\phi) \otimes \sigma^2(D - \alpha W)^{-1}.
\]

It is straightforward to show that the marginal distribution from this constructive approach for each \( Z(t_i) \) is \( N(0, \sigma^2(D - \alpha W)^{-1}) \), the same marginal distribution as the temporally independent MRF specification in (2). Therefore, our constructive approach ensures a valid space-time process, where associations in space are modeled discretely using a MRF, and those in time through a continuous Gaussian process.

This separable specification is easily interpretable as it factorizes the dispersion into a spatial association component (areal) and a temporal component. Another significant practical advantage is its computational feasibility. Estimating more general space-time models usually entails matrix factorizations with \( O(N_s^3 N_t^3) \) computational complexity. The separable specification allows us to reduce this complexity substantially by avoiding factorizations of \( N_s N_t \times N_s N_t \) matrices. One could design algorithms to work with matrices whose dimension is the smaller of \( N_s \) and \( N_t \), thereby accruing massive computational gains.

We would, however, be remiss not to recognize the limitations of the separable model and how, at least in theory, our constructive approach can obviate such limitations. The separable model imposes the same temporal correlation structure for every areal unit because each of the \( v_i(t) \)'s are assumed to follow independent Gaussian processes with the same correlation function for every region \( i \). Letting each \( v_i(t) \) have its own set of process parameters \( \phi_i \) yields:

\[
\text{Model 2: } K_Z(t, u) = A\tilde{R}(t, u; \phi_1, \ldots, \phi_{N_s})A^T \quad \text{and } \Sigma_Z = \Sigma_v(I \otimes A^2),
\]

where \( \tilde{R}(t, u; \phi_1, \ldots, \phi_{N_s}) \) is an \( N_s \times N_s \) diagonal matrix with \( \rho(t, u; \phi_j) \) as the \( i \)-th diagonal element, \( A \) is a square-root matrix such that \( AA^T = \sigma^2(D - \alpha W)^{-1} \), and \( \Sigma_v \) is the variance-covariance matrix for the \( v_i(t_j) \)'s, that is an \( N_s N_t \times N_s N_t \) block matrix with \( \tilde{R}(t_i, t_j; \phi_1, \ldots, \phi_{N_s}) \) as the \((i, j)\)-th block. Thus, we still assume that the elements of \( v(t) \) are independent Gaussian processes but are no longer identical. Now each region will be able to account for its own temporal correlation characteristics, including the smoothness of its temporal curve and the rate of decay of temporal correlation.

Model 2, while temporally non-stationary, still assumes the same spatial association structure across time. This, perhaps, is a less stringent limitation. It is not clear when one would envision a
temporally evolving spatial association between the areal units, particularly for data with temporally short duration. However, if this were the case, then temporally evolving spatial associations can be modeled using $A(t)$ as a function of $t$. For instance, we could extend Model 2 to:

**Model 3:** $K_Z(t,u) = A(t)\bar{R}(t,u;\phi_1,\phi_2,\ldots,\phi_N)A(u)^T$ and $\Sigma_Z = \tilde{A}\Sigma_v\tilde{A}^T$,

where $A(t)$ is a square root of $\sigma^2(t)(D - \alpha(t)W)^{-1}$, $\sigma^2(t)$ and $\alpha(t)$ are continuous functions of time, and $\tilde{A}$ is an $N_sN_t \times N_sN_t$ block diagonal matrix with $A(t_i)$’s as its diagonal blocks. This raises the question of what type of function $\alpha(t)$ should be. Recall that $\alpha(t) \in (0,1)$ for all $t$. One could perhaps think of a polynomial-spline, say $\psi(t)$, on the logit-link of the smoothness parameter: $\log(\alpha(t)/(1-\alpha(t))) = \psi(t)$. Indeed $\psi(t)$ will likely have parameters so one will need to ascertain identifiability issues in such general settings. Even more generally, if one does not want to assume a CAR structure, we can keep the $A(t)$ as unknown functions and estimate them. For example, we could treat its diagonal elements as a log-Gaussian, and the lower-triangular elements are independent Gaussian processes.

While further modeling of these functions using stochastic processes is conceivable in Bayesian contexts, such models will be excessively complex, both for estimation and interpretation, and without much foreseeable inferential gains in most practical settings. This is especially true in our current context, where we are more interested in estimating gradients from models. Gradient analysis can be performed seamlessly once the posterior distribution for any of the above models has been estimated. Since the more complex Models 2 and 3 do not offer anything new in terms of temporal gradients, we do not pursue them in the remainder of this paper.

4. **Hierarchical modeling.** In this section, we build a hierarchical modeling framework to analyze the data in Section 2 using the likelihood from our spatial random effects model in (1) and the distributions emerging from the temporal Gaussian process discussed in Section 3. The mean $\mu_i(t)$ in (1) is often indexed by a parameter vector $\beta$, for example a linear regression with regressors indexed by space and time so that $\mu_i(t;\beta) = x_i(t)^T\beta$.

The posterior distributions we seek can be expressed as

$$p(\theta,Z|Y) \propto p(\phi) \times IG(\sigma^2|a_\sigma,b_\sigma) \times \left(\prod_{i=1}^M IG(\tau_i^2|a_\tau,b_\tau)\right) \times N(\beta|\mu_\beta,\Sigma_\beta)$$

$$\times N(Z|0,R(\phi) \otimes \sigma^2(D - \alpha W)^{-1}) \times \prod_{j=1}^{N_t} \prod_{i=1}^{N_s} N(Y_i(t_j)|x_i(t_j)^T\beta + Z_i(t_j),\tau_i^2),$$

(4)
where \( \theta = \{ \phi, \sigma^2, \beta, \tau_1^2, \tau_2^2, \ldots, \tau_{K_i}^2 \} \) and \( Y \) is the vector of observed outcomes defined analogous to \( Z \). The parametrizations for the standard densities are as in Carlin and Louis (2009). We assume all the other hyperparameters in (4) are known, and we fix \( \alpha = 0.9 \).

Recall the separable cross-covariance function in (3). The correlation function \( \rho(\cdot; \phi) \) determines process smoothness and we choose it to be an isotropic Matérn correlation function given by

\[
\rho(t, u; \phi) = \rho(\Delta; \phi) = \frac{1}{\Gamma(\nu)2^{\phi_2-1}} \left( 2\sqrt{\phi_2|\Delta|\phi_1} \right)^{\phi_2} K_{\phi_2} \left( 2\sqrt{\phi_2|\Delta|\phi_1} \right),
\]
where \( \phi = \{ \phi_1, \phi_2 \} \), \( \Delta = t - u \), \( \Gamma(\cdot) \) is the Gamma function, \( K_{\phi_2}(\cdot) \) is the modified Bessel function of the second kind, and \( \phi_1 \) and \( \phi_2 \) are non-negative parameters representing rate of decay in temporal association and smoothness of the underlying process, respectively.

We use Markov chain Monte Carlo (MCMC) to evaluate the joint posterior in (4), using Metropolis steps for updating \( \phi \) and Gibbs steps for all other parameters; details of which are shown in the supplemental article (Quick et al. 2011). Sampling-based Bayesian inference seamlessly delivers inference on the residual spatial effects. Specifically, if \( t_0 \) is an arbitrary unobserved time-point, then, for any region \( i \), we sample from the posterior predictive distribution \( p(Z_i(t_0) | Y) = \int p(Z_i(t_0) | Z, \theta) p(\theta, Z | Y) \, d\theta \, dZ \). This is achieved using composition sampling: for each sampled value of \( \{ \theta, Z \} \), we draw \( Z_i(t_0) \), one for one, from \( p(Z_i(t_0) | Z, \theta) \), which is Gaussian. Also, our sampler easily adapts to situations where \( Y_i(t) \) is missing (or not monitored) for some of the time points in region \( i \). We simply treat such variables as missing values and update them, from their associated full conditional distributions, which of course are \( N(x_i(t)T \beta + Z_i(t), \tau_i^2) \). We assume that all predictors in \( x_i(t) \) will be available in the space-time data matrix, so this temporal interpolation step for missing outcomes is straightforward and inexpensive.

Model checking is facilitated by simulating independent replicates for each observed outcome: for each region \( i \) and observed timepoint \( t_j \), we sample from \( p(Y_\text{rep,}i(t_j) | Y) = \int N(Y_\text{rep,}i(t_j) | x_i(t_j)^T \beta + Z_i(t_j), \tau_i^2) \, p(\beta, Z_i(t_j), \tau_i^2 | Y) \, d\beta \, dZ_i(t_j) \, d\tau_i^2 \), where \( p(\beta, Z_i(t_j), \tau_i^2 | Y) \) is the marginal posterior distribution of the unknowns in the likelihood. Sampling from the posterior predictive distribution is straightforward, again, using composition sampling.

5. Gradient Analysis. Our primary goal is to carry out statistical inference on temporal gradients with data arising from a temporal process indexed discretely over space. We will do so using the notions of smoothness of a Gaussian process and its derivative. Adler (1981), Mardia et al. (1996) and Banerjee and Gelfand (2003) discuss derivatives (more generally, linear functionals) of Gaussian processes, while Banerjee, Gelfand and Sirmans (2003) lay out an inferential framework
for directional gradients on a spatial surface. Most of the existing work on derivatives of stochastic processes deal either with purely temporal or purely spatial processes (see, e.g., Banerjee, 2010). Here, we consider gradients for a temporal process indexed discretely over space.

Assume that \( \{ Z_i(t) : t \in \mathbb{R}^1 \} \) is a stationary random process for each region \( i \). The process is \( L^2 \) (or mean square) continuous at \( t_0 \) if \( \lim_{t \to t_0} E(|Z_i(t) - Z_i(t_0)|) = 0 \). The notion of a mean square differentiable process can be formalized using the analogous definition of total differentiability of a function in a non-stochastic setting (see, e.g., Banerjee and Gelfand, 2003): \( Z_i(t) \) is mean square differentiable at \( t_0 \) if it admits a first order linear expansion for any scalar \( h \),

\[
Z_i(t_0 + h) = Z_i(t_0) + hZ'_i(t) + o(h)
\]

in the \( L^2 \) sense as \( h \to 0 \), where we say that \( \frac{d}{dt}Z_i(t) = Z'_i(t_0) \) is the gradient or derivative process derived from the parent process \( Z_i(t) \). In other words, we require

\[
\lim_{h \to 0} E \left( \frac{Z_i(t_0 + h) - Z_i(t_0)}{h} - Z'_i(t_0) \right)^2 = 0.
\]

Equations (6) and (6') ensure that mean square differentiable processes are mean square continuous.

For a univariate stationary process, smoothness in the mean square sense is determined by its covariance or correlation function. A stationary multivariate process \( Z(t) \) with cross-covariance function \( K_Z(\Delta) \) will admit a well-defined gradient process \( Z'(t) = (Z'_1(t), \ldots, Z'_N(t))^T \) if and only if \( K''_Z(0) \) exists, where \( K''_Z(0) \) is the element-wise second-derivative of \( K_Z(\Delta) \) evaluated at \( \Delta = 0 \).

A Gaussian process with a Matérn correlation function has sample paths that are \( \lceil \phi_2 - 1 \rceil \) times differentiable. As \( \phi_2 \to \infty \), the Matérn correlation function converges to the squared exponential (or the so-called Gaussian) correlation function, which is infinitely differentiable and leads to acute oversmoothing. When \( \phi_2 = 0.5 \), the Matérn correlation function is identical to the exponential correlation function (see, e.g., Stein, 1999). To ensure that the underlying process is differentiable so that the gradient process exists, we need to restrict \( \phi_2 > 1 \). However, letting \( \phi_2 > 2 \) usually leads to oversmoothing as the data can rarely distinguish among values of the smoothness parameter greater than 2. Hence, we restrict \( \phi_2 \in (1, 2] \). We could either assign a prior on this support or simply fix the \( \phi_2 \) somewhere in this interval. Since it is difficult to elicit informative priors for the smoothness parameter, we would most likely end up with a uniform prior. In our experience, this delivers only modest posterior learning, and the substantive inference is not very different from what is obtained by fixing \( \phi_2 \).
As such, in our subsequent analysis we fix \( \phi_2 = 3/2 \), which has the side benefit of yielding the closed form expression \( \rho(\Delta; \phi_1) = (1 + \phi_1|\Delta|) \times \exp(-\phi_1|\Delta|) \). The first and second order derivatives for the cross-covariance function in (3) can now be obtained explicitly as

\[
    K_Z'(\Delta) = -\sigma^2 \phi_1^2 \Delta \exp(-\phi_1|\Delta|)(D - \alpha W)^{-1} \quad \text{and} \quad -K_Z''(0) = \sigma^2 \phi_1^2 (D - \alpha W)^{-1}.
\]

Turning to inference for gradients, we seek the joint posterior predictive distribution,

\[
    p(Z'(t_0) | Y) = \int p(Z'(t_0) | Y, Z, \theta) p(Z | \theta, Y) p(\theta | Y) d\theta dZ
\]

\[
= \int p(Z'(t_0) | Z, \theta) p(Z | \theta, Y) p(\theta | Y) d\theta dZ,
\]

where the second equality follows from the fact that the gradient process is derived entirely from the parent process and so \( p(Z'(t_0) | Y, Z, \theta) \) does not depend on \( Y \).

We evaluate (8) using composition sampling. Here, we first obtain \( \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(M)} \sim p(\theta | Y) \) and \( Z^{(j)} \sim p(Z | \theta^{(j)}, Y), j = 1, 2, \ldots, M \), where \( M \) is the number of (post-burn-in) posterior samples. Next, for each \( j \) we draw \( Z^{(j)} \sim p(Z | \theta^{(j)}, Y) \), and finally \( Z'(t_0)^{(j)} \sim p(Z'(t_0) | Z^{(j)}, \theta^{(j)}) \). The conditional distribution for the gradient can be seen to be multivariate normal with mean and variance-covariance matrix given by

\[
    \mu_{Z'|Z, \theta} = \text{cov}(Z'(t_0), Z) \text{var}(Z)^{-1} Z = -(K_Z')^T \Sigma^{-1}_Z Z
\]

and

\[
    \Sigma_{Z'|Z, \theta} = -K_Z''(0) - (K_Z')^T \Sigma^{-1}_Z (K_Z'),
\]

where \( \Sigma^{-1}_Z = \frac{1}{\sigma^2} R(\phi)^{-1} \otimes (D - \alpha W) \) and \( (K_Z')^T \) is a block matrix whose \( j \)-th block is given by the \( N_s \times N_s \) matrix \( K_Z'(\Delta_{0j}) \), with \( \Delta_{0j} = t_j - t_0 \). Note that \( \Sigma_{Z'|Z, \theta} \) is an \( N_s N_t \times N_s N_t \) matrix, but we can use the properties of the MRF to only invert \( N_t \times N_t \) matrices.

6. Simulation. To validate our methodology, we conducted a simulation using the 58 counties of California as our spatial grid, \( N_t = 50 \), and \( t_j = j = 1, 2, \ldots, N_t \). To generate an interesting temporal pattern with spatial clustering, we assumed

\[
    Y_i(t_j) \overset{\text{iid}}{\sim} N \left( 5 + x_{i1} \cdot \sin \left( \frac{t_j}{2} \right) + x_{i2} \cdot \cos \left( \frac{t_j}{2} \right), \tau^2 \right),
\]

where \( x_{i1} \) is the \( i \)-th county’s percent black, \( x_{i2} \) is the \( i \)-th county’s ozone level from April 1991, as described in Section 2. For illustration purposes, we fix \( \tau = 0.1 \). We then modeled the data using only an intercept, leaving the spatiotemporal random effects to capture the sinusoidal curve, and conducted the gradient analysis at the midpoints of each time interval. Our Gaussian Process
model provided a good fit for the random effects, and our temporal gradient estimates accurately reproduced the theoretical gradient curves derived using elementary calculus, as can be seen in the left and right panels of Figure 2 for a particular spatial region. While these data had little random noise, this example demonstrates the validity of the gradient theory derived in Section 5.

7. Data Analysis. As first mentioned in Section 2, our dataset is comprised of monthly asthma hospitalization rates in the counties of California over an 18-year period. As such, \( N_t = 12 \times 18 = 216 \), and we will again use \( t_j = j = 1, 2, \ldots, N_t \). The covariates in this model include population density, ozone level, the percent of the county under 18, and percent black. Population-based covariates are calculated for each county using the 2000 U.S. Census, thus they do not vary temporally. However, the covariate for ozone level is aggregated at the air basin level and varies monthly, though show little variation annually. In order to accommodate seasonality in the data, monthly fixed effects are included, using January as a baseline. Thus, \( x_i(t) \) is a \( 16 \times 1 \) vector.

To justify the use of the model we’ve described, we compare it to three alternative models using the DIC criterion (Spiegelhalter et al. 2002). These models are all still of the form

\[
Y_i(t) = x_i(t)' \beta + Z_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{\text{iid}}{\sim} N(0, \tau^2_i) \quad \text{for} \quad i = 1, 2, \ldots, N_s,
\]

but with different \( Z_i(t) \). Our first model is a simple linear regression model which ignores both the spatial and the temporal autocorrelation, i.e., \( Z_i(t) = 0 \ \forall \ i, t \). The second model allows for a random intercept and random temporal slope, but ignores the spatial nature of the data, i.e., here \( Z_i(t) = \alpha_{0i} + \alpha_{1i}t \), where \( \alpha_{ki} \overset{\text{iid}}{\sim} N(0, \sigma^2_k) \), for \( k = 0, 1 \). In this model, to preserve model
identifiability, we must remove the global intercept from our design matrix, \( x_i(t) \). Our third model builds upon the second, but introduces spatial autocorrelation by letting \( \alpha_k = (\alpha_{k1}, \ldots, \alpha_{kN_s})' \sim CAR(\sigma_k^2), k = 0, 1 \). The results of the model comparison can be seen in Table 1, which indicates that our Gaussian process model has the lowest DIC value, and is thus the preferred model and the only one we consider henceforth. The surprisingly large \( p_D \) for the areally referenced Gaussian process model arises due to the very large size of the dataset (58 counties \( \times \) 216 timepoints).

Table 2 shows our fixed effect parameter estimates. The coefficients for the monthly covariates indicate decreased hospitalization rates in the summer months, a trend which is consistent with previous findings. The coefficients for population density, percent under 18, and percent black are all significantly positive, also as expected. The coefficient for ozone level is significantly negative, however, which is surprising but consistent with the patterns in the monthly trends for both hospitalization rates and ozone levels. The strong spatial story seen in the maps is emphasized by the estimates of \( \sigma^2 \) and the \( \tau_i^2 \)'s \( [\sigma^2/(\tau_i^2 + \sigma^2)] > 0.8 \) for most \( i \), and there is a relatively strong temporal correlation, with \( \phi = 1.03 \) corresponding to \( \rho(t_i, t_j; \phi) \geq 0.4 \) for \( |t_j - t_i| < 2 \) months.

Maps of the yearly (averaged across month) spatiotemporal random effects can be seen in Figure 3. Since here we’re dealing with the residual curve after accounting for a number of mostly non-time-varying covariates, it comes as no surprise that the spatiotemporal random effects capture most of the variability in the model, including the striking decrease in yearly hospitalization rates...
over the study period. It also appears that our model is providing a better fit to the data in the years surrounding 2000, perhaps indicating that we could improve our fit by allowing our demographic covariates to vary temporally. Our model also appears to be performing well in the central counties, where asthma hospitalization rates remained relatively stable for much of the study period.

An analysis of the average annual temporal gradients in Figure 4 shows that while rates generally decreased, there are a number of years in which average asthma hospitalization rates increased statewide. Figure 5, which compares the monthly temporal profiles of the random effects and their gradients for Los Angeles and San Francisco Counties, explores these results. For Los Angeles County, the spatiotemporal random effects (top-left panel) decrease at a consistent, moderate rate throughout the length of the study. Several large spikes prior to 2000, are apparent here and also in the temporal gradients (lower-left panel). In contrast, San Francisco County’s random effects (upper-right) have fewer and less dramatic spikes, which is supported by its temporal gradients (lower-right). In addition, San Francisco County appears to have had a changepoint in its spatiotemporal random effects around 2000, where they transition from a fairly steady decline to a period of lower variability and very little mean change. Further investigation may reveal a corresponding change in social, environmental, or health care reimbursement policy.

As our data are aggregated monthly, we felt it was also important to investigate the gradients on a month-to-month basis over the course of the study. For instance, Figure 6 reveals the gradients between August and September decrease substantially statewide over the course of the study. Coupling this with the information in Table 2, which indicates that hospitalization rates in September are $\beta_{12} - \beta_{11} = 1.61$ per thousand higher than those in August, suggests that the difference in asthma hospitalization rates between August and September has nearly disappeared, going from
roughly 2.91 at the beginning of the period to just 0.62 by the end. An investigation of the raw hospitalization rates shows a similar trend, but this is to be expected since most of the spatiotemporal variability in the model is accounted for by the random effects. A similar, though not as striking phenomenon occurs between March and April, where the gradients are increasing. As these two pairs of months lie on the transition between the warmer months and the cooler months, this result would seem to suggest that the effect of seasonality has moderated over the length of the study.

One limitation of this analysis is that the data records asthma hospitalizations, not overall prevalence. This is an important distinction, as factors that trigger symptoms of asthma may not be the same as or have the same impact on asthma hospitalizations. For instance, residents of
regions with high risk environments may be better educated about and/or prepared for managing their symptoms, which could lead to a relative decrease in asthma hospitalization rates.

8. Summary and Conclusions. In this paper, we have provided an overview of parent and gradient processes, building on previous work in spatiotemporal Gaussian process modeling. We then described our modeling framework and methods that allow for inference on temporal gradients. An implementation for one of these methods was outlined in Section 4, and its theory was verified via simulation. Its use was then illustrated on a real dataset in Section 7; where our results showed real insight can be gained from an assessment of temporal gradients in the residual Gaussian process, indicating overall trends as well as motivating a search for temporally interesting covariates still missing from our model (say, one that changes abruptly in San Francisco County around 2000).

An important point of discussion is the importance of significance with respect to the temporal gradients. We believe it depends on the problem being modeled. In our case, a significant gradient in our data indicates a significant difference between two adjacent months after adjusting for important fixed effects, namely weather patterns and pollution levels. While we have accounted for monthly differences in our design matrix, the $Z_i(t)$ here may simply be capturing the remaining cyclical trend, and this is why we felt it was more beneficial to focus on the trends of the twelve month-to-month comparisons rather than solely on whether a specific gradient for a particular county was significant. In situations where it’s reasonable to assume two time-points are comparable—say, in data measured annually—investigating significant temporal gradients can indicate periods of important changes in the data. We also point out that the methodology for gradients outlined here can be applied to more general spatial functional data analysis contexts and will be especially
useful for estimating gradients from high-resolution samples of the function.

We certainly have not exhausted our modeling options. Some of the richer association structures we mentioned in Section 3 may be appropriate in alternate inferential contexts. While we demonstrated the advantages of the process-based specifications over some simpler parametric options for $Z_i(t)$ in our data analysis, one could envision alternative specifications depending upon the inferential question at hand. For example, if interested in separating the variability between time and space using two variance parameters, additive specifications such as $Z_i(t) = u_i + w(t)$, where $u_i$'s follow a Markov random field and $w(t)$ is a temporal Gaussian process, could be explored. Now the $u_i$'s and $w(t)$'s could have their own variance components. This, however, would not allow the temporal functions to borrow strength across the neighbors as effectively as we do here.

Apart from exploring such alternate specifications, our future work includes expanding our focus to include spatiotemporal gradients for point-referenced (geostatistical) data, where our response arises from a spatiotemporal process $Y(s; t)$ with $s \in \mathbb{R}^d$. Typically, we have a finite collection of sites $S = \{s_1, \ldots, s_n\}$ and time points $t \in T = \{t_1, \ldots, t_N\}$ (as before) where the responses $Y(s_i; t_j)$ have been observed. Spatiotemporal gradient analysis in this setting offers richer possibilities, and of course avoids the problems associated with the CAR model’s failure to offer a true spatial process (Banerjee et al, 2004, p. 82-83). Here one can conceptualize spatial (directional) gradients, temporal gradients or even “mixed” gradients.

SUPPLEMENTARY MATERIAL

Imputation of missing daily hospitalization counts and MCMC details (www.biostat.umn.edu/~brad/QBCsupplement.pdf). As data for days with between one and four asthma hospitalizations are missing, we impute county-specific values for these days using a method similar to Besag’s iterated conditional modes method (Besag, 1986) but with means. We also lay out the details for the MCMC implementation.

REFERENCES


2nd edition.