Spatiotemporal Gradient Modeling with Applications

HARRISON S. QUICK

Thesis Proposal

Submitted in partial fulfillment of the requirements for the preliminary oral examination of the Doctor of Philosophy degree in Biostatistics in the Graduate School of The University of Minnesota

February 27, 2012
Advised by
Sudipto Banerjee, Ph.D and Bradley P Carlin, Ph.D.

---

\[1\] Harrison S. Quick is a Graduate Assistant in the Division of Biostatistics, School of Public Health, 420 Delaware St. S.E., University of Minnesota, Minneapolis, MN, 55455.
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.

The overarching theme of my dissertation work will be the statistical estimation of temporal and spatiotemporal gradients and their application to real-world data. Representing instantaneous rates of change, an analysis of these gradients can allow researchers to learn how the residuals in a statistical model change over time. While many factors may lead to significant changes in the residual surface, an important yet unobserved covariate can be a likely culprit. For instance, a change in public health policy that waives co-pays for the uninsured could result in a sudden and possibly sustained increase in clinic visits across an entire region. Since ignoring important covariates can lead to biased parameter estimates and invalid inference, not only do I believe that an investigation of the residual surface is crucial, but also that analyzing its temporal and spatiotemporal gradients is a valuable tool for this task.

1.1 Statistical motivation

There is a considerable literature in spatio-temporal modeling; see, for example, the recent book 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. 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. Carroll et al. (1997) again use these processes, assuming a separable form for the space-time covariance function to study ground level ozone. 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. More recent examples include the dynamic CAR model proposed by Martínez-Beneito et al. (2008) and the latent structure models approach from Lawson et al. (2010).

1.1.1 Discrete space, continuous time setting

The first part of my dissertation work 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. In other words, 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.
Spatially associated functions have received little attention, especially for regionally aggregated data. This is unfortunate, especially given the datasets we face in our work (see Section 1.2.1).

As such, in Section 2 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 this work 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. Subsequently, we infer on temporal gradients, that is, the rate of change of the underlying process over time, from data that arise from the underlying stochastic process. The smoothness implications for the underlying process are apparent. We deploy a mean square differentiable Gaussian process that provides a tractable gradient process to help us achieve these inferential goals.

1.1.2 Extension to continuous space, continuous time setting

After developing methodology for the discrete space, continuous time setting, we look to extend this work to the case where space varies continuously. Specifically, we will model the data as coming from a stochastic process \( \{Y(s, t) : s \in D_s, t \in D_t\} \), where \( D_s \) and \( D_t \) are fixed subsets of 2- and 1-dimensional Euclidean space representing space and time, respectively. This scenario commonly arises in geostatistical settings, such as with monitoring stations at fixed locations that continuously collect data related to weather and air quality. Here, inferential interest is typically in both the estimation of model parameters as well as in spatiotemporal prediction (kriging).

One drawback of the continuous space setting is that many covariates are difficult to (or impossible) obtain at precise spatial locations, which is the idea of spatial misalignment. In the discrete space case, assigning attributes to a spatial region, such as population density, median income, etc., is not only easier to do (though can still pose challenges), but certainly easier to interpret. An analogous issue of temporal misalignment occurs for covariates that don’t properly align temporally, but this seems less troublesome, perhaps due to interpretability (e.g., assigning a region a population at a precise time point compared to assigning a precise spatial location a population for
any discrete time period). In general, misalignment, both spatial and temporal, can occur in either the continuous or discrete setting and its impact must be considered. For a detailed discussion of spatial misalignment (the concepts of which can be generalized to the temporal case), see Gelfand (2010).

In Section 3 of this proposal, I outline a methodology that I believe will accomplish these tasks while also permitting inference on not only temporal, but also spatial and spatiotemporal gradients. After validating its theory via simulation and implementing its use on air quality data from California (described in Section 1.2.2), I hope to further develop this work in order to fit larger, more complex datasets, while also making it accessible to researchers unfamiliar with the statistical theory from which it was designed. In this regard, Section 1.2.3 describes data collected during the clean-up of the Deepwater Horizon (BP) oil spill in the Gulf of Mexico, a dataset that will not only require specialized methods to account for the complex processes behind airborne particle transmission, but also algorithms which reduce the computational burden to a manageable level. Section 4 includes a brief discussion of these challenges, as well as preliminary ideas for their solutions. Finally, Section 5 concludes and offers a few ideas for future research.

1.2 Applications

1.2.1 California asthma hospitalization data

The first application of our methods will use a dataset consisting 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,
1986). For our analysis, the data are aggregated by month, for a total of 216 observations per county over the 18-year period. Figure 1, which maps the raw annual average rates, reveals that hospitalization for asthma demonstrates a statewide decreasing trend early in the study period, and appears to stabilize in later years.

We attempt to capture the effect of socioeconomic status by including population density in our model, using population 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 asthma 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 will be to detect temporal changes in the residuals that remain after the covariates are accounted for; significant changes may indicate corresponding 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.

1.2.2 California air quality data

The second spatiotemporal dataset we plan to use pertains to air quality in California. We have access to a variety of measurements of air quality, such as ozone and carbon monoxide levels, recorded daily at a large number of stations over 18 years. Unlike the asthma hospitalization data,
these data are point-referenced; the exact spatial locations of the observations are known and no spatial aggregation has taken place. There is also some missingness in the data, but this appears to be a matter of whether or not the station was on-line or off-line during that period (e.g., some stations have complete data over a period of a number of years, but no observations before or after that period). For computational reasons, we will likely restrict our attention to a more recent time period (say, Jan. 2008 to Dec. 2009) and may likewise include only the stations which appear to be on-line during this entire period.

The data will be considered to vary continuously in both space and time, permitting inference on spatiotemporal gradients. We will assume that the data (or a transformation of the data, if necessary) are normally distributed. To model these data, we will look to collect covariate information related to pollution, such as population density, however here we will encounter both spatial and temporal misalignment issues discussed in Section 1.1.2. Other covariates that we can include in our model which won’t be affected by misalignment include latitude and longitude, seasonality, and day of the week (or perhaps a “weekday”/“weekend” indicator).

1.2.3 Deepwater Horizon oil spill data

In the aftermath of the Deepwater Horizon oil spill in the Gulf of Mexico in April, 2010, the National Institute of Environmental Health Sciences began conducting an epidemiological study to investi-
gate the possible adverse health effects of the workers engaged in the clean-up work. The exposure level of a number of chemical agents has been assessed, including total hydrocarbons, benzene, toluene, xylene, ethylbenzene, n-hexane, total polycyclic aromatic hydrocarbons, oil dispersants (2-butoxyethanol, propylene glycol), and PM2.5.

The workers have been classified into “exposure scenarios”, which were made on the basis of job or task, area, type of vessel, geographical location, and date worked. Exposure levels of workers assigned to a particular scenario are similar for the array of chemicals in the spilled oil as well as the dispersants and particulates to which the workers were exposed. Exposure varies by task, geographical location, and time since the spill occurred. The classification was determined on the basis of: (a) task/near field, derived from the type of task the workers were performing; (b) task/far field, derived from being in the proximity of someone else performing a task in a given geographical location; and (c) environmental/far field, derived from being in a particular location far from an emission source.

In total, the dataset consists of over 150,000 personal exposure measurements and hundreds of thousands of area concentration measurements collected by a number of organizations (BP, USEPA, NIOSH, NOAA) on the clean-up workers. These measurements will be assigned to the exposure scenarios, with the aim being to model, for each scenario, the exposure for each chemical agent. Preliminary assessments have indicated that a vast majority of the data is below the limits of detection, suggesting that extensions of our spatiotemporal modeling approach to handle censored data will be required.

2 Modeling temporal gradients in regionally aggregated data

2.1 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.

A spatial random effect model for our data assumes

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

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 easily 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 \( w_{ij} \) equals some non-zero value 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 \( i \),

\[
p(Z_i(t) \mid \{Z_{j \neq i}(t)\}) \sim N \left( \sum_{j \sim i} \frac{\alpha w_{ij}}{w_{i+}} Z_j(t), \frac{\sigma^2}{w_{i+}} \right), \tag{2}
\]

where \( w_{i+} = \sum_{j \sim i} w_{ij} \), \( \sigma^2 > 0 \), and \( \alpha \) is a propriety parameter described below. This means that
the $N_s \times 1$ vector $Z(t) = (Z_1(t), Z_2(t), \ldots, Z_{N_t}(t))^T$ follows a multivariate normal distribution with zero mean and a precision matrix $\frac{1}{\alpha^2}(D - \alpha W)$, 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 $Z(t)$ at each timepoint $t$.

The MRF in (2) does not allow temporal dependence; the $Z(t)$’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 $Z(t) \overset{ind}{\sim} N(0, \sigma_t^2(D - \alpha_t W)^{-1})$ 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 $Z(t)$. A highly flexible and computationally tractable option is to assume that $Z(t)$ is a zero-centered multivariate Gaussian process, $GP(0, K_Z(\cdot, \cdot))$, where 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) $\sum_{i=1}^{N_t} \sum_{j=1}^{N_t} u_i^T K_Z(t_i, t_j) u_j > 0$ for all $u_i, u_j \in \mathbb{R}^{N_t} \setminus \{0\}$. The first condition follows immediately from the symmetry of covariances. For any $N_t$ and any arbitrary collection of timepoints $T = \{t_1, t_2, \ldots, t_{N_t}\}$, the $N_s N_t \times 1$ vector of realizations $Z = (Z(t_1)^T, \ldots, Z(t_{N_t})^T)^T$ 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 \(K_{ij}(t) = \int \exp(2\pi i ut)d(K_{ij}(u))\), 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 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 \( u \). Note that the \( v_i(t) \)'s are independent across \( i \), so that \( \text{cov}\{v_i(t), v_j(u)\} = 0 \) whenever \( i \neq j \).

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 (nonstationary) whenever the \( v(t) \) is stationary (nonstationary). 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 \( \sigma^2(D - \alpha W)^{-1} \) yields

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

where \( R(\phi) \) is the \( N_t \times N_t \) temporal correlation matrix having \( (i, j) \)-th element \( \rho(t_i, t_j; \phi) \). 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.
2.2 Hierarchical modeling

In this section, we build a hierarchical modeling framework to analyze the data in Section 1.2.1 using the likelihood from our spatial random effects model in (1) and the distributions emerging from the temporal Gaussian process discussed in Section 2.1. 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

\[
p(\theta, Z | Y) \propto p(\phi) \times IG(\sigma^2 | a_\sigma, b_\sigma) \times \left( \prod_{i=1}^{M} IG(\tau_i^2 | a_\tau, b_\tau) \right) \times N(\beta | \mu_\beta, \Sigma_\beta) \times N(Z | 0, R(\phi) \otimes \sigma^2(D - \alpha W)^{-1}) \times \prod_{j=1}^{N_t} \prod_{i=1}^{N_s} N(Y_i(t_j) | x_i(t_j)^T \beta + Z_i(t_j), \tau_i^2),
\]

where \( \theta = \{ \phi, \sigma^2, \beta, \tau_1^2, \tau_2^2, \ldots, \tau_{N_s}^2 \} \) 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

\[
\rho(t, u; \phi) = \rho(\Delta; \phi) = \frac{1}{\Gamma(\nu)2^{\phi_2-1}} \left( 2\sqrt{\phi_2} | \Delta | \phi_1 \right)^{\phi_2} K_{\phi_2} \left( 2\sqrt{\phi_2} | \Delta | \phi_1 \right), \tag{5}
\]

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. Sampling-based Bayesian inference seamlessly delivers inference on the residual spatial effects. Specifically, if \( t_0 \) is an arbitrary unobserved timepoint, 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 \( N(x_i(t)^T \beta + Z_i(t), \tau^2) \). 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 \textit{independent} replicates for each observed outcome: for each region \( i \) and observed timepoint \( t_j \), we sample from
\[
p(Y_{\text{rep},i}(t_j) | Y) = \int N(Y_{\text{rep},i}(t_j) | x_i(t_j)^T \beta + Z_i(t_j), \tau^2) \ p(\beta, Z_i(t_j), \tau^2 | Y) d\beta dZ_i(t_j) d\tau^2,
\]
where \( p(\beta, Z_i(t_j), \tau^2 | Y) \) is the marginal posterior distribution of the unknowns in the likelihood. Sampling from the posterior predictive distribution is straightforward, again, using composition sampling.

### 2.3 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.

Let us assume that \( Z_i(t) \) is a stationary random process for each region \( i \).\(^2\) The process \( \{Z_i(t) : t \in \mathbb{R}^{1+}\} \) is \( L_2 \) (or mean square) continuous at \( t_0 \) if \( \lim_{t \to t_0} \ E(|Z_i(t) - Z_i(t_0)|^2) = 0 \). 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). In particular, \( Z_i(t) \) is mean square differentiable at \( t_0 \) if it admits a first order linear

\(^2\)Stationarity is not required. We only use it to ensure smoothness of realizations and to simplify forms for the induced cross-covariance function.
expansion for any scalar $h$,

$$Z_i(t_0 + h) = Z_i(t_0) + hZ_i'(t) + o(h)$$  \hspace{1cm} (6)

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

$$\lim_{h \to 0} E\left(\frac{Z_i(t_0 + h) - Z_i(t_0)}{h} - Z_i'(t_0)\right)^2 = 0.$$  \hspace{1cm} (6')

This first order linearity ensures 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 $Z'(t) = (Z_1'(t), Z_2'(t), \ldots, Z_N'(t))^T$ 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 $\nu \to \infty$, 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 $\phi = 0.5$, 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 $\phi > 1$. However, letting $\phi > 2$ usually leads to oversmoothing as the data can rarely distinguish among values of the smoothness parameter greater than 2. Hence, we restrict $\phi \in (1, 2]$. We could either assign a prior on this support or simply fix the $\phi$ 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 $\phi$.

As such, in our subsequent analysis we fix $\phi = 3/2$, 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

$$
K'_Z(\Delta) = -\sigma^2 \phi_1^2 \Delta \exp(-\phi_1|\Delta|)(D - \alpha W)^{-1} \text{ and } -K''_Z(0) = \sigma^2 \phi_1^2(D - \alpha W)^{-1}.
$$

(7)

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

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

(8)

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 $\Sigma^{-1}_Z = \frac{1}{\sigma^2} R(\phi)^{-1} \otimes (D - \alpha W)$, $(K'_Z)^T$ is an $N_s \times N_sN_t$ block matrix whose $j$-th block is given by the $N_s \times N_s$ matrix $K'_Z(\Delta_{0j})$, with $\Delta_{0j} = t_j - t_0$, and $-K''_Z(0)$ as defined in (7). Note that $\Sigma_{Z'} | Z, \theta$ is an $N_sN_t \times N_sN_t$ matrix, but we can use the properties of the MRF to only invert $N_t \times N_t$ matrices.

### 2.4 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 induce spatiotemporal clustering, we
used covariates from the California data to generate our data. To generate an interesting temporal pattern, we assumed

\[ Y_i(t_j) \sim N \left( 5 + x_{i1} \sin \left( \frac{t_j}{2} \right) + x_{i2} \cos \left( \frac{t_j}{2} \right), \tau^2 \right), \quad (9) \]

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 1.2.1. 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 2.3.

### 2.5 Data analysis

As first mentioned in Section 1.2.1, 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
Figure 3: Maps of the standardized covariates used in asthma hospitalization study. Color cut-offs are the same for each map, range from light to dark, and are as follows: -1.0, -0.4, 0.2, 0.8, 1.4, 2.0. Ozone level maps are averaged over the length of the study for each month. Population density, % Under 18, and % Black remain fixed for all timepoints. Note: San Francisco County has a standardized population density of 7.2, nearly six times that of any other county. Due to its size (47 sq. miles), it’s difficult to see on the map.

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, maps of which can be seen in Figure 3. 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.

Fixed effect parameter estimates for this analysis can be found in Table 1. 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
Table 1: Parameter estimates for asthma hospitalization data, where estimates for $\tau^2$ represent the median (95% CI) for all of the $\tau^2_i$ maps is emphasized by the estimates of $\sigma^2$ and the $\tau^2_i$s $[\sigma^2/(\tau^2_i + \sigma^2)] > 0.8$ for most $i$, and there is a relatively strong temporal correlation, with the posterior mean of $\phi = 1.03$ corresponding to $\rho(t_i, t_j; \phi) \geq 0.4$ for $|t_j - t_i|$ less than 2 months.

Maps of the yearly (averaged across month) spatiotemporal random effects can be seen in Figure 4. 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.

Because our data are aggregated monthly, we felt it was most important to investigate the gradients on a month-to-month basis over the course of the study. For instance, Figure 5 reveals the gradients between August and September decrease substantially statewide over the course of the study. Coupling this with the information in Table 1, 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 asthma 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 quite 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.

3 Spatiotemporal gradients in spatially continuous data

3.1 Point-Level spatiotemporal processes

In the continuous space, continuous time setting described in Section 1.1.2, a typical spatiotemporal random effect model assumes the form

\[ Y(s, t) = \mu(s, t) + w(s, t) + e(s, t), \quad \forall (s, t) \in D_s \times D_t, \]  

(10)

where \( \mu(s, t) = x(s, t)'\beta \), \( w(s, t) \sim GP(0, K_\phi(\cdot, \cdot)) \),

\[ K_\phi = \text{cov}\{w(s_1, t_1), w(s_2, t_2)\}, \]  

(11)

and \( \phi \) represents parameter (or vector of parameters) which control the spatiotemporal covariance. A property which we desire from (11) is stationarity over space-time, that is

\[ K_\phi((s_1, t_1), (s_2, t_2)) = K_\phi(s_2 - s_1, t_2 - t_1). \]  

(12)

A common method for achieving (12) is via a separable model, that is

\[ K_\phi((s_1, t_1), (s_2, t_2)) = K_{\phi_s}(s_1, s_2) \cdot K_{\phi_t}(t_1, t_2), \]

where we assume \( K_{\phi_s}(s_1, s_2) = K_{\phi_s}(s_2 - s_1) \) and \( K_{\phi_t}(t_1, t_2) = K_{\phi_t}(t_2 - t_1) \).

While computationally convenient, it has been shown (Kyriadkidis and Journel, 1999; Cressie and Huang, 1999) that separable models do not allow for space-time interaction, which leads to processes that are too simple and frequently fail to adequately model the underlying physical process. Fortunately, a very rich literature (a discussion of which can be found in Gneiting and Guttorp, 2010) exists for space-time covariance functions, including those which are non-stationary and non-separable.

In practice, we have data collected over \( S = \{s_1, s_2, \ldots, s_{N_s}\} \) spatial locations and \( T = \)
\{t_1, t_2, \ldots, t_{N_t}\} time points. The likelihood for such data can be expressed as

\[ Y(s_i, t_j) = x(s_i, t_j)' \beta + w(s_i, t_j) + e(s_i, t_j), \ i = 1, \ldots, N_s, \ j = 1, \ldots, N_t, \]  

(13)

and may be collected either over space or over time. For instance, if collected over time

\[ Y(s_i) = X(s_i)' \beta + w(s_i) + e(s_i), \ i = 1, \ldots, N_s, \]

(14)

where \( \text{cov}\{w(s_i), w(s_j)\} = K_w(s_i, s_j) = \{K_\phi((s_i, t_k), (s_j, t_\ell))\}_{k, \ell=1}^{N_t} \) and

\[ \text{cov}\{e(s_i), e(s_j)\} = \begin{cases} 0, & s_i \neq s_j \\ \Sigma_e, & s_i = s_j \end{cases} \]

for \( \Sigma_e = \text{diag}\{\tau_1^2, \ldots, \tau_{N_t}^2\} \) or simply \( \tau^2 I_{N_t} \).

### 3.1.1 Spatiotemporal predictive processes

One challenge associated with the use of Gaussian likelihoods in this setting is that they require evaluating the \( N_t N_s \times N_t N_s \) precision matrix \( K_\theta^{-1} \). As matrix inversion requires computations on the order of \( O(N_t^3 N_s^3) \), one solution is to use a low-rank method, such as predictive process modeling, as proposed by Banerjee et al. (2008). Recent work in this area by Liang et al. (2010) includes the application of spatial gradients in predictive process models, where its use was compared to methods for areal data.

In predictive process models, we begin by considering a set of knots, \( S^* = \{s_1^*, s_2^*, \ldots, s_{N_s^*}\} \), \( T^* = \{t_1^*, t_2^*, \ldots, t_{N_t^*}\} \), where \( N_s^* < N_s \) and \( N_t^* < N_t \). The Gaussian process in model (13) yields \( w^* \sim N(0, \Sigma_{w^*}) \), where \( w^* = (w^*(s_1^*)', w^*(s_2^*)', \ldots, w^*(s_{N_s^*}')')' \), \( w^*(s_i^*) = (w(s_i^*, t_1^*), w(s_i^*, t_2^*), \ldots, w(s_i^*, t_{N_t^*}'))' \), \( \Sigma_{w^*} = \{K^*(s_i^*, s_j^*)\}_{i,j=1}^{N_s^*} \), and \( K^*(s_i^*, s_j^*) = \{K_\phi((s_i^*, t_k^*), (s_j^*, t_\ell^*))\}_{k, \ell=1}^{N_t^*} \). The spatiotemporal interpolant at site \((s, t)\) is given by

\[ \bar{w}(s, t) = E[w(s, t) \mid w^*] = k^*_\phi(s, t)'(\Sigma_{w^*})^{-1}w^*, \]

(15)
where \( k_\phi^*(s, t)' = \text{cov}\{w(s, t), w^*\} \). Moreover, \( \tilde{w}(s, t) \sim GP(0, \tilde{K}_\phi(\cdot)) \) with covariance function

\[
\tilde{K}_\phi[(s, t), (s', t')] = k_\phi^*(s, t)'(\Sigma_{w^*})^{-1}k_\phi^*(s', t').
\]

Thus we can replace \( w(s_i, t_j) \) in model (13) with \( \tilde{w}(s_i, t_j) \), to obtain the predictive process model

\[
Y(s_i, t_j) = \mathbf{x}(s_i, t_j)'\beta + \tilde{w}(s_i, t_j) + \epsilon(s_i, t_j).
\]

By (15), it is clear that \( \tilde{w}(s_i, t_j) \) is a spatiotemporally varying linear transformation of \( w^* \), and hence we have achieved a dimension reduction. In fitting model (16), the \( N_sN_t \) random effects \( \{w(s_i, t_j), i = 1, \ldots, N_s, j = 1, \ldots, N_t\} \) have been replaced by just the \( N_s^*N_t^* \) random effects in \( w^* \).

More information regarding predictive processes, including various properties and drawbacks and a discussion of the selection of knots, is given in Banerjee et al. (2008).

### 3.1.2 Spatiotemporal tapered predictive processes

One drawback of the predictive process model outlined in Section 3.1.1 is that it will sacrifice small scale spatial dependence in the name of computational efficiency. Because this small scale dependence is arguably the most influential, this is not ideal. Another method that can be used to reduce the computational burden in spatiotemporal analyses is the use of covariance tapering, as proposed by Furrer et al. (2006). In essence, their idea is to coerce the covariance matrix of the spatiotemporal random effects into a sparse matrix, permitting the use of sparse matrix libraries and their associated computational benefits. This is done by tapering the covariance of “distant” observations to zero through a positive definite but compactly supported function. More specifically, let \( \Sigma \) be a generic covariance matrix and let \( K_\gamma \) be a covariance matrix that is identically zero outside a particular range controlled by \( \gamma \). Then a tapered covariance matrix, \( \Sigma_{\text{top}} \), can be constructed using the Schur product of \( \Sigma \) and \( K_\gamma \), that is, for any two points \( \mathbf{x}, \mathbf{x}' \):

\[
\Sigma_{\text{top}}(\mathbf{x}, \mathbf{x}') = \Sigma(\mathbf{x}, \mathbf{x}')K_\gamma(\mathbf{x}, \mathbf{x}').
\]
If we suppose our data varies spatiotemporally, that is, \( x = (s, t) \), we can let \( \gamma = (\gamma_s, \gamma_t) \) to allow for separate tapering spatially and temporally.

Clearly, the use of tapering will put an emphasis on the small scale spatial dependence that the predictive process lacks. Thus, we will also look to the approach used by Sang and Huang (2012) and then combine the two methods, producing a covariance structure that simultaneously has the benefits of both. To achieve this, we replace (15) with

\[
\bar{w}(s, t) = E[w(s, t) | w^*] = k^*_\phi(s, t)'\Sigma^{-1}w^* + \bar{\epsilon}(s, t),
\]

where \( \bar{\epsilon}(s, t) \sim GP(0, K_{\bar{\epsilon}}(\cdot, \cdot)) \) and

\[
K_{\bar{\epsilon}}((s_1, t_1), (s_2, t_2)) = [K((s_1, t_1), (s_2, t_2)) - k^*_\phi(s_1, t_1)'\Sigma^{-1}k^*_\phi(s_2, t_2)]
\cdot K_{\gamma_s, \gamma_t}((s_1, t_1), (s_2, t_2)).
\]

Here, \( K_{\gamma_s, \gamma_t}((s_1, t_1), (s_2, t_2)) \) represents our tapered covariance and takes the value zero whenever \( ||s_1 - s_2|| > \gamma_s \) or \( ||t_1 - t_2|| > \gamma_t \). From this, we define

\[
\bar{K}((s_1, t_1), (s_2, t_2)) = \text{cov}\{\bar{w}(s_1, t_1), \bar{w}(s_2, t_2)\}
= k^*_\phi(s_1, t_1)'\Sigma^{-1}k^*_\phi(s_2, t_2) + K_{\bar{\epsilon}}((s_1, t_1), (s_2, t_2)).
\]

Note that if either \( ||s_1 - s_2|| > \gamma_s \) or \( ||t_1 - t_2|| > \gamma_t \) holds, our covariance is exactly that from the predictive process model, otherwise it’s a weighted average of the predictive process covariance and the covariance assumed under our model, \( K((s_1, t_1), (s_2, t_2)) \). We refer to the model of the form (16) where \( \bar{w}(s, t) \) is from (17) as our spatiotemporal tapered predictive process (STTPP). Properties of covariance matrices of the form (18), such as results on its positive (semi-)definiteness, can be found in Sang and Huang (2012).
3.2 Hierarchical modeling

Similar to Section 2.2, we will build a hierarchical modeling framework to analyze the California air quality data described in Section 1.2.2 using our STTPP. The posterior distributions we seek can be expressed as

\[
p(\theta, \tilde{w}, w^* | Y) \propto p(\phi, \gamma) \times \prod_{j=1}^{N_t} IG(\tau_j^2 | a_r, b_r) \times N(\beta | \mu_\beta, \Sigma_\beta) \times N(w^* | 0, \Sigma_{w^*}) \\
\times N(\tilde{w} | B_\phi w^*, \tilde{D}_\overline{\omega}) \times \prod_{i=1}^{N_s} N(y(s_i) | X(s_i)\beta + \tilde{w}(s_i), \Sigma_e),
\]

where \( \theta = \{\phi, \beta, \tau_1^2, \tau_2^2, \ldots, \tau_{N_s}^2\} \), \( y(s_i) = (y(s_i, t_1), \ldots, y(s_i, t_{N_t}))' \), and \( X(s_i) = (x(s_i, t_1), \ldots, x(s_i, t_{N_t}))' \) is the \( N_t \times p \) design matrix for the \( i \)th spatial region, where the \( j \)th row corresponds to the transpose of the \( p \times 1 \) vector, \( x(s_i, t_j) \). The mean vector for \( \tilde{w} \) is expressed as a linear transformation of \( w^* \), where it is premultiplied by

\[
B_\phi = \begin{bmatrix}
K^*_\phi(s_1)' \\
\vdots \\
K^*_\phi(s_{N_s})'
\end{bmatrix} \Sigma_{w^*}^{-1},
\]

with \( K^*_\phi(s_i)' = \{K^*_\phi(s_i, t_j)\}'_{j=1}^{N_t} \) being an \( N_t \times N_s^* \) matrix. Its covariance matrix, \( \tilde{D}_\overline{\omega} = \{K_\tilde{\omega}(s_i, s_j)\}_{i,j=1}^{N_s} \), is an \( N_s N_t \times N_s N_t \) matrix made up of \( N_t \times N_t \) blocks,

\[
K_\tilde{\omega}(s_i, s_j) = \{K(s_i, s_j) - K^*_\phi(s_i)\Sigma_{w^*}^{-1}K^*_\phi(s_j)\} \odot K_\gamma(s_i, s_j),
\]

where \( K(s_i, s_j) = \{K((s_i, t_k), (s_j, t_\ell))\}_{k,\ell=1}^{N_t} \), \( K_\gamma(s_i, s_j) = \{K_\gamma((s_i, t_k), (s_j, t_\ell))\}_{k,\ell=1}^{N_t} \), and \( \odot \) denotes the Schur product for component-wise multiplication.

The joint posterior in (19) will be evaluated via MCMC; \( \phi \) will be estimated using Metropolis steps while all other model parameters will be estimated via the Gibbs sampler. Note also that the distribution on \( \phi \) will depend upon the choice of covariance function, which will be described in the following subsection.
3.3 Gradient analysis

As mentioned in the introduction, our primary goal in this section is to make inference on spatiotemporal gradients from data collected in a setting where both space and time are considered to vary continuously. Much of the theory regarding the existence of a gradient process follows from the work in Section 2.3 and the paper by Banerjee et al. (2003).

Before we can outline the derivation of our gradient process, we must first discuss the covariance and tapering covariance functions ($K_\phi$ and $K_\gamma$, respectively), as their differentiability is necessary in order for the gradient process to exist. While the exact form of these functions has yet to be decided, some properties are desirable. For instance, for two given spatial locations, a fully symmetric function is unable to distinguish between the effect of time moving forward from time moving backward. For the dataset discussed in Section 1.2.2, this is undesirable, as factors such as wind tendencies may cause the air quality on the coast to effect the air quality inland in the future, but not vice versa. Thus, a nonseparable (for reasons discussed in Section 3.1), non fully symmetric correlation function is desired for $K_\phi$, as well as one that is differentiable, a property of the Matérn($\nu = 3/2$). The choice of the tapering covariance, $K_\gamma$, appears to be more simple, as Furrer et al. (2006) construct their tapers using convenient forms of the Matérn correlation structure. More specifically, they suggest that the choice of taper depends on the number of derivatives you wish to exist. For instance, if you wish to have a once differentiable covariance structure, the Matérn($\nu = 3/2$) is a common choice, and there’s a taper derived from and corresponding to this choice of $\nu$.

After choosing our covariance functions, gradient calculations are straightforward. We will seek

$$\frac{d}{dt} \nabla w(s, t) = \nabla \left( \frac{d}{dt} w(s, t) \right),$$

where $\nabla w(s, t)$ is the spatial gradient analogue of $Z_i'(t)$ in (6), and thus

$$\frac{d}{dt} D_\mathbf{u} w(s, t) = \mathbf{u}' \left( \frac{d}{dt} \nabla w(s, t) \right),$$

where $D_\mathbf{u} w(s, t)$ denotes the gradient of $w(s, t)$ in the direction of $\mathbf{u}$, for any unit vector, $\mathbf{u}$. Also,
note that in our STTPP model, (17) can be written as

\[
\tilde{w}(s, t) = k^*_\phi(s, t)'\Sigma_w^{-1}w^* + \eta(s, t)\gamma(s, t),
\]

and so

\[
\frac{d}{dt} \nabla w(s, t) = \left(\frac{d}{dt} \nabla k^*_\phi(s, t)'\right) \Sigma_w^{-1}w^* + \frac{d}{dt} \nabla (\eta(s, t)\gamma(s, t))
\]

\[
= (\eta(s, t)\gamma(s, t)k^*_\phi(s, t)') \Sigma_w^{-1}w^* + \eta(s, t) \left(\frac{d}{dt} \nabla \gamma(s, t)\right) + \left(\frac{d}{dt} \nabla \eta(s, t)\right) \gamma(s, t). \tag{20}
\]

3.4 Analysis of California air quality data

To analyze the California air quality data, we will model the data using (16) where \(\tilde{w}(s_i, t_j)\) uses the low-rank and tapered covariance structure from (17). Covariates for this model will likely include latitude and longitude and covariates to seasonal trends, though a literature review will be conducted to determine other covariates typically used to model this type of data.

Much as with the asthma hospitalization data, we will first fit the model using MCMC, and then calculate gradients using composition sampling. In this case, however, it will be impossible to compute every temporal and every spatial gradient, as there are an infinite number of potential spatial and temporal sites. One feasible option could be to select points either at random or using a grid, and then compute temporal and simple directional gradients for each selected point. A more thorough gradient analysis could then occur either near the more “interesting” points from that initial analysis, or simply by selecting sites via a visual inspection of the residual maps, though an automated selection method may be preferred, since visual inspection can be too subjective.
4 Spatiotemporal gradients in interval-censored airborne exposures data from the clean-up of the *Deepwater Horizon* oil spill

The final part of my dissertation work will undertake a thorough analysis of the airborne exposure data collected during the clean-up of the *Deepwater Horizon* (BP) oil spill in the Gulf of Mexico in 2010. While this will mostly be applied work from a statistical perspective—building off the STTPP methodology used for the California air quality—this data presents a number of interesting challenges that will need to be addressed.

First and foremost, because of the number of observations that are below the limits of detection, our methods must be equipped for interval censoring. Statistically speaking, I don’t expect much complication from this, as the censored data can be estimated directly from the Gibbs sampler, but this will add another layer of computation to what will likely be an already computationally burdensome project.

Before getting more in-depth on the issue of computing, another aspect of this data which needs attention from both a statistical as well as an environmental science perspective is the correlation structure of the spatiotemporal process. If we are to expect that our methods will be used in practice, we need to take into account knowledge of how airborne particles transfer from point to point, and verify that our methods (e.g., choice of correlation function, method for tapering, etc.) are sensibly selected.

The final challenge I expect to encounter is the overall computational burden this data poses. Bayesian methods are inherently computationally intensive, as they can require repeating estimating a large number of parameters over tens of thousands of iterations. In our work with the asthma hospitalization data, the separable model allowed us to avoid the daunting (and almost crippling) task of inverting a $12,000 \times 12,000$ matrix through a clever restructuring of the data, resulting in a speed-up of a factor of 10. For the air quality, we hope the use of statistical methods such as predictive processes and tapering and the transition in programming language from R to C++ will result in more feasible problem. For this project, however, the size of the dataset itself may present challenges, since a standard computer may not have sufficient memory to store, let alone model,
the data. As such, we plan to work with computer scientists (including Prof. Shashi Shekhar) to investigate ways to handle this computational problem. One possible solution would be to devise ways to break up the data into smaller pieces, in such a way that storing it without sacrificing valuable information is no longer an issue, but also avoids double-counting observations.

5 Discussion

The Section 2 results from the asthma hospitalization data suggest that temporal and spatiotemporal gradients can become valuable tools. Building off of previous work in spatiotemporal Gaussian process modeling, we have already developed a modeling framework and methods that allow for inference on temporal gradients that can be verified via simulation. In the asthma data, 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 lurking covariates, still missing from our model.

This concept of using gradients to find “lurking covariates” is potentially an area of future research. For instance, we could simulate data in a scenario where we know the data’s structure, fit the model with an important covariate left out, and then conduct a full gradient analysis. This process could be repeated for a number of covariate types, such as continuous and discrete covariates, both spatially- and temporally-varying, with interesting features noted. Given that gradients estimate rates of change, one could imagine significantly high gradients to be representative of lurking covariates that differ in their values across adjacent time points (for temporal gradients) or across nearby locations (for spatial gradients). As such, it may be possible to identify (and possibly categorize) missing covariates in this fashion, which could give investigators a better idea of what is missing from their model.

An important point of discussion is the importance of statistical significance of the temporal gradients. We believe it depends on the problem being modeled. In the case of our asthma hospitalization data, a significant gradient 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 or when the spatiotemporal distance between two points is small enough that residual trends are negligible—investigating significant temporal and/or spatiotemporal gradients can indicate places (whether in space, in time, or a mixture of space and time) of important changes in the data.

References


