Title
Bayesian Modeling and Analysis for Gradients in Spatiotemporal Processes

Authors
Quick, Harrison
Banerjee, Sudipto
Carlin, Bradley P.

Publication Date
2015-01-27

Peer reviewed

Bayesian Modeling and Analysis for Gradients in Spatiotemporal Processes

Harrison Quick[1], Sudipto Banerjee[2*], and Bradley P. Carlin[3]

[1]Division of Heart Disease and Stroke Prevention, NCCDPHP/CDC, Atlanta, GA 30341-3717.

[2]Department of Biostatistics, University of California, Los Angeles, CA 90095-1772.

[3]Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota 55455.

* Corresponding author *email:* sudipto@ucla.edu

Summary. Stochastic process models are widely employed for analyzing spatiotemporal datasets in various scientific disciplines including, but not limited to, environmental monitoring, ecological systems, forestry, hydrology, meteorology and public health. After inferring on a spatiotemporal process for a given dataset, inferential interest may turn to estimating rates of change, or *gradients*, over space and time. This manuscript develops fully model-based inference on spatiotemporal gradients under continuous space, continuous time settings. Our contribution is to offer, within a flexible spatiotemporal process model setting, a framework to estimate arbitrary directional gradients over space at any given timepoint, temporal derivatives at any given spatial location and, finally, mixed spatiotemporal gradients that reflect rapid change in spatial gradients over time and vice-versa. We achieve such inference without compromising on rich and flexible spatiotemporal process models and use nonseparable covariance structures. We illustrate our methodology using a simulated data example and subsequently apply it to a dataset of daily PM2.5 concentrations in California, where the spatiotemporal gradient process reveals the effects of California's unique topography on pollution and detects the aftermath of a devastating series of wildfires.

Key words: Gaussian process, Gradients, Markov chain Monte Carlo

1 Introduction

Spatiotemporal modeling has enjoyed much attention over last two decades; see, for example, Cressie and Wikle (2011) and references therein. Such models assume a collection of real-valued random variables $\{Z(\mathbf{s},t) : (\mathbf{s},t) \in \Re^d \times \Re\}$, where $\mathbf{s}$ denotes spatial coordinates residing in the d-dimensional Euclidean space (usually $d = 2$ or 3 in spatial statistics), and t is time with domain $\Re$, the real line. Depending upon the application, $Z(\mathbf{s},t)$ can represent an observed response such as an environmental pollutant or, perhaps, an underlying physical process assumed to be generating the observeds response. The preferred modeling approach is to specify a space-time covariance function for $Z(\mathbf{s},t)$, which will ensure a valid joint distribution for the realizations of the process over a finite collection of locations and time points. Christakos (1992, 2000), Jones and Zhang (1997), Cressie and Huang (1999), Brown et al. (2000), de Iaco et al. (2002), Gneiting (2002), Ma (2003) and Hartfield and Gunst (2003) all offer classes of stationary space-time covariance functions.

Based upon observations obtained from a finite set of locations and timepoints, spatiotemporal models interpolate $Z(\mathbf{s},t)$ at arbitrary $\mathbf{s}$ and t. Interest, subsequently, can turn to estimating rates of change, or *gradients*, over space and time, where rapid changes occur over space and/or time. Examples include geographic features such as mountain ranges causing sharp spatial changes in precipitation levels, the degree of traffic congestion increasing abruptly at rush hour, and sudden changes in land values in urban areas, (e.g., Majumdar et al. 2006). In practice, it is difficult to distinguish between discontinuities in $Z(\mathbf{s},t)$ from rapid continuous change, especially when $Z(\mathbf{s},t)$ is a residual surface. Model-based inference on spatiotemporal gradients, as we propose below, seems much more tractable (analytically and computationally) than inferring about discontinuities.

The factors underlying sudden changes in observed spatiotemporal data are often unknown, and not easily accounted for. Here, estimating gradients exploiting the smoothness

of the process becomes relevant and is often referred to as *wombling*, named after a seminal paper by Womble (1951). Wombling has been explored in purely spatial contexts by Barbujani et al. (1989), Fortin (1994, 1997), Fortin and Drapeau (1995), Banerjee et al. (2003), Banerjee and Gelfand (2006), Liang et al. (2009) and Gabriel et al. (2011); see Banerjee (2010) for an overview. More recently, Quick et al. (2013) carried out inference on purely temporal (but not spatial) rates of change in asthma hospitalization in California.

Hitherto, inference on rates of change for spatiotemporal processes has gone unaddressed. Here, we offer inference for directional gradients over space at any given timepoint, temporal derivatives at any given spatial location, and *mixed* spatiotemporal gradients capturing rapid change in temporal gradients over space and vice-versa. Rather than infer on discretized spatiotemporal finite differences as approximations to gradients, which tend to become numerically unstable at sharper resolutions (Banerjee et al., 2003), we utilize limiting properties of Gaussian processes to develop the required distribution theory for spatiotemporal gradients. We deploy nonseparable covariance functions that ensure appropriately smooth specifications for $Z(\mathbf{s}, t)$. We construct valid *cross-covariance* functions that allow joint modeling for $Z(\mathbf{s}, t)$ and its derivatives across space and time. We offer full distributional details for Bayesian inference on the aforementioned gradients.

The remainder of this article evolves as follows. Section 2 outlines the notion of a spatiotemporal gradient process, followed by the distribution theory necessary for inference. Section 3 embeds this distribution theory within a Bayesian modeling framework and describes how to carry out posterior predictive inference on spatiotemporal gradients. Section 4 presents a simulated data example, which aims to verify our proposed methodology on gradients. In Section 5, we analyze air quality data (PM2.5, particulate matter less than 2.5 micrometers in diameter) from the Air Resources Board of the California Environmental Protection Agency, where we use the spatiotemporal gradient process to identify the impact

of mountain ranges on pollution surfaces and detect the aftermath of a devastating series of wildfires in Northern California. Finally, Section 6 concludes the chapter with a discussion.

2 Spatiotemporal gradients

2.1 Calculus of spatiotemporal gradients

Let $Z(\mathbf{s},t)$ be a (weakly) stationary real-valued spatiotemporal process on $\Re^d \times \Re$ with mean 0 and finite second moment and *covariance function* $\mathrm{Cov}\{Z(\mathbf{s},t), Z(\mathbf{s}',t')\} = K(\boldsymbol{\Delta}, \delta)$, where $\boldsymbol{\Delta} = \mathbf{s}' - \mathbf{s}$ and $\delta = t' - t$ for every pair of space-time coordinates $(\mathbf{s},t)$ and $(\mathbf{s}',t')$ in $\Re^d \times R$. This function must be *positive-definite* (e.g. Gneiting, 2002), which requires

$$\mathrm{Var}\left\{\sum_{i=1}^{k} a_i Z(\mathbf{s}_i,t_i)\right\} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j K(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) > 0 \tag{1}$$

for every finite collection of $(\mathbf{s}_i,t_i) \in \Re^d \times \Re$ and nonzero $a_i \in \Re$, $i = 1,2,\ldots,k$.

The process $Z(\mathbf{s},t)$ is L_2 (or mean-squared) continuous at $(\mathbf{s},t)$ if $E\{Z(\mathbf{s}+\boldsymbol{\Delta}_s, t+\Delta_t) - Z(\mathbf{s},t)\}^2 \to 0$ as $(\boldsymbol{\Delta}_s, \Delta_t) \to 0$. We say that $Z(\mathbf{s},t)$ is mean-square (totally) differentiable at $(\mathbf{s},t)$ if there exist a $d \times 1$ spatial gradient process, $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$, and a real valued temporal gradient process, $\nabla_t Z(\mathbf{s},t)$, such that, for any given vector $\mathbf{u} \in \Re^d$ and any real number v,

$$Z(\mathbf{s}+h\mathbf{u}, t+hv) = Z(\mathbf{s},t) + h\mathbf{u}^T\boldsymbol{\nabla}_s Z(\mathbf{s},t) + hv\nabla_t Z(\mathbf{s},t) + o(h) \;\text{ as }\; h \to 0 \tag{2}$$

in the L_2 sense for any scalar h. Equivalently, for any $\mathbf{u} \in \Re^d$ and any $v \in \Re$, we require

$$\lim_{h\to 0} E\left(\frac{Z(\mathbf{s}+h\mathbf{u}, t+hv) - Z(\mathbf{s},t)}{h} - \mathbf{u}^T\boldsymbol{\nabla}_s Z(\mathbf{s},t) - v\nabla_t Z(\mathbf{s},t)\right)^2 = 0\,. \tag{3}$$

We refer to $Z(\mathbf{s},t)$ as the *parent process* from which $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$ and $\nabla_t Z(\mathbf{s},t)$ are derived.

If we fix $v = 0$ in (2), then (3) supplies the spatial directional derivative process as

$$D_{\mathbf{u}}Z(\mathbf{s},t) = \lim_{h\to 0}\frac{Z(\mathbf{s}+h\mathbf{u},t)-Z(\mathbf{s},t)}{h} = \mathbf{u}^T\boldsymbol{\nabla}_s Z(\mathbf{s},t)\ . \tag{4}$$

If we collect a set of p directions into a $d \times p$ matrix $\mathbf{U} = [\mathbf{u}_1 : \mathbf{u}_2 : \ldots : \mathbf{u}_p]$, we then obtain a $p \times 1$ process $\mathbf{D}_{\mathbf{U}}Z(\mathbf{s},t) = (D_{\mathbf{u}_1}Z(\mathbf{s},t), D_{\mathbf{u}_2}Z(\mathbf{s},t), \ldots, D_{\mathbf{u}_p}Z(\mathbf{s},t))^T$. From (4), we see that $D_{\mathbf{U}}Z(\mathbf{s},t) = \mathbf{U}^T\boldsymbol{\nabla}_s Z(\mathbf{s},t)$. If, for instance, we take the standard (or canonical) Euclidean basis vectors in $\Re^d$ as our set of directions, then $p = d$, $\mathbf{U} = \mathbf{I}_d$ and

$$\boldsymbol{\nabla}_s Z(\mathbf{s},t) = D_{\mathbf{I}_d}Z(\mathbf{s},t) = \left(\frac{\partial}{\partial s_1}Z(\mathbf{s},t), \frac{\partial}{\partial s_2}Z(\mathbf{s},t), \ldots, \frac{\partial}{\partial s_d}Z(\mathbf{s},t)\right)^T\ , \tag{5}$$

where $\mathbf{s} = \sum_{i=1}^{d} s_i\mathbf{e}_i$, so the s_i's are the coordinates of $\mathbf{s}$ with respect to the canonical basis, and $D_{\mathbf{e}_i}Z(\mathbf{s},t) = (\partial/\partial s_i)Z(\mathbf{s},t)$. Henceforth, we refer to $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$ in (5) as the *spatial gradient process*. Since $D_{\mathbf{U}}Z(\mathbf{s},t) = \mathbf{U}^T\boldsymbol{\nabla}_s Z(\mathbf{s},t)$, the directional derivative process along a set of arbitrary directions is a linear transformation of the spatial gradient process, and inference for $D_{\mathbf{U}}Z(\mathbf{s},t)$ proceeds immediately from inference for $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$. Also, when inferring about $D_{\mathbf{u}}Z(\mathbf{s},t)$ it suffices to consider $\mathbf{u}$ to be a unit vector (i.e., $\|\mathbf{u}\| = 1$) since $D_{\mathbf{w}}Z(\mathbf{s},t) = \|\mathbf{w}\|D_{\mathbf{u}}Z(\mathbf{s},t)$ for any vector $\mathbf{w}$ that is parallel to $\mathbf{u}$.

If $\mathbf{u} = \mathbf{0}$ and $v = 1$ in (2), then (3) yields the temporal gradient process

$$\nabla_t Z(\mathbf{s},t) = \frac{\partial}{\partial t}Z(\mathbf{s},t) = \lim_{h\to 0}\frac{Z(\mathbf{s},t+h)-Z(\mathbf{s},t)}{h}\ . \tag{6}$$

We must assume $Z(\mathbf{s},t)$ is mean-square totally differentiable for the directional gradient processes to exist in every direction because the partial derivative processes in (5) and (6) do not ensure differentiability of $Z(\mathbf{s},t)$ along every direction.

Since the processes $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$ and $\nabla_t Z(\mathbf{s},t)$ are well-defined and have explicit canonical

representations with respect to partial derivatives, it is natural to consider gradient processes arising from mixed derivatives. For example, we can construct the $d \times 1$ process $\nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s}, t)$ whose coordinates are given by mixed derivative in the L_2 limit

$$\lim_{h \to 0} \frac{D_{\mathbf{e}_i} Z(\mathbf{s}, t+h) - D_{\mathbf{e}_i} Z(\mathbf{s}, t)}{h} = \frac{\partial^2}{\partial t \partial s_i} Z(\mathbf{s}, t) \quad \text{for} \quad i = 1, 2, \ldots, d \, . \tag{7}$$

This is the temporal derivative of the spatial gradient, reflecting the continuous rate of change in spatial gradients over time. Alternatively, we could first take the temporal derivative followed by the spatial gradient, which yields the $d \times 1$ process $\boldsymbol{\nabla}_s \nabla_t Z(\mathbf{s}, t)$ with coordinates

$$\lim_{h \to 0} \frac{\nabla_t Z(\mathbf{s} + h\mathbf{e}_i, t) - \nabla_t Z(\mathbf{s}, t)}{h} = \frac{\partial^2}{\partial s_i \partial t} Z(\mathbf{s}, t) \quad \text{for} \quad i = 1, 2, \ldots, d \, . \tag{8}$$

Assuming that all mixed second order partial derivatives are mean-square continuous at every space-time coordinate $(\mathbf{s}, t)$, we can legitimately change the order of the derivatives so that $\nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s}, t) = \boldsymbol{\nabla}_s \nabla_t Z(\mathbf{s}, t)$ almost surely for every space-time coordinate. Hence, we can unambiguously define the $d \times 1$ mixed derivative process as

$$\boldsymbol{\nabla}_{s,t} Z(\mathbf{s}, t) = \boldsymbol{\nabla}_s \nabla_t Z(\mathbf{s}, t) = \nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s}, t) \, . \tag{9}$$

While equivalent, the practical interpretations of (7) and (8) may differ; e.g., it may be more natural to think of $\nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s}, t)$ when modeling the movement of a storm over time.

2.2 Distribution theory for spatiotemporal gradients

Let $Z(\mathbf{s}, t)$ be a univariate stationary zero-centered Gaussian random field $GP(0, K(\cdot, \cdot; \boldsymbol{\theta}))$, where $\boldsymbol{\theta}$ is a collection of process parameters. For notational convenience, we suppress the dependence on $\boldsymbol{\theta}$ and simply write $K(\cdot, \cdot)$ for the stationary covariance function defined on

$\Re^d \times \Re$. Based upon observations from a finite set of locations, say $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{N_s}\}$, and timepoints $\mathcal{T} = \{t_1, t_2, \ldots, t_{N_t}\}$, we wish to predict the different spatiotemporal gradients outlined in Section 2.1 at an arbitrary space-time coordinate $(\mathbf{s}, t)$. This suggests developing a multivariate process comprising the parent $Z(\mathbf{s}, t)$ and its derivative processes.

To be precise, let $\mathbf{W}(\mathbf{s}, t) = (\mathbf{W}_1(\mathbf{s}, t)^T, (\nabla_t \mathbf{W}_1(\mathbf{s}, t))^T)^T$ be the $2(d+1) \times 1$ process, where $\mathbf{W}_1(\mathbf{s}, t) = (Z(\mathbf{s}, t), (\boldsymbol{\nabla}_s Z(\mathbf{s}, t))^T)^T$ and $\nabla_t \mathbf{W}_1(\mathbf{s}, t) = (\nabla_t Z(\mathbf{s}, t), (\nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s}, t))^T)^T$ are each $(d+1) \times 1$. Its cross-covariance matrix $\mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta}, \delta)$ is $2(d+1) \times 2(d+1)$ partitioned as

$$\begin{bmatrix} \text{Cov}\{\mathbf{W}_1(\mathbf{s}, t), \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\} & \text{Cov}\{\mathbf{W}_1(\mathbf{s}, t), \nabla_t \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\} \\ \text{Cov}\{\nabla_t \mathbf{W}_1(\mathbf{s}, t), \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\} & \text{Cov}\{\nabla_t \mathbf{W}_1(\mathbf{s}, t), \nabla_t \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\}\} \end{bmatrix}, \quad (10)$$

where each of the above blocks is $(d+1) \times (d+1)$. The cross-covariance matrix in (10) need not be symmetric or positive definite (unlike a covariance matrix), but must satisfy:

$$\text{(i) } \mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta}, \delta) = \mathbf{C}_{\mathbf{W}}(-\boldsymbol{\Delta}, -\delta)^T \text{ and (ii) } \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{a}_i^T \mathbf{C}_{\mathbf{W}}(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) \mathbf{a}_j > \mathbf{0}\,, \quad (11)$$

for every $(\boldsymbol{\Delta}, \delta) \in \Re^d \times \Re$, $\mathbf{a}_i \in \Re^{2(d+1)} \setminus \{\mathbf{0}\}$, and finite set of space-time coordinates $\{(\mathbf{s}_1, t_1), (\mathbf{s}_2, t_2), \ldots, (\mathbf{s}_n, t_n)\}$. We write (10) in terms of the covariance function $K(\boldsymbol{\Delta}, \delta)$ and its derivatives by first constructing the corresponding finite-difference process and then passing to limits. Since the finite difference processes arise as linear transformations of the original process, the associated cross-covariance matrices are valid (i.e., they satisfy (11)) by construction. This ensures that $\mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta}, \delta)$ in (10) is also valid because it arises as limits of the valid finite-difference cross-covariances. The (i, j)-th block in (10) can be expressed as

$$\begin{bmatrix} \nabla_t^{i-1} \nabla_t^{j-1} K(\boldsymbol{\Delta}, \delta) & (\nabla_t^{i-1} \nabla_t^{j-1} \boldsymbol{\nabla}_s K(\boldsymbol{\Delta}, \delta))^T \\ -\nabla_t^{j-1} \nabla_t^{i-1} \boldsymbol{\nabla}_s K(\boldsymbol{\Delta}, \delta) & -\nabla_t^{i-1} \nabla_t^{j-1} \mathbf{H}_K(\boldsymbol{\Delta}, \delta) \end{bmatrix}, \; i, j = 1, 2\,,$$

where ∇_t acts element-wise on vectors and matrices. See Web Appendix A for details.

We now turn to choosing $K(\boldsymbol{\Delta}, \delta)$, which determines the smoothness of the spatiotemporal process realizations, thereby ensuring the existence of the spatiotemporal gradient processes. When constructing the parent process, one could conceivably use separable models, which factor the spatiotemporal covariance function in terms of a purely spatial and a purely temporal covariance function, e.g., $K(||\boldsymbol{\Delta}||, |\delta|) = K_s(||\boldsymbol{\Delta}||)K_t(|\delta|)$. However, the resulting correlations are sensitive to small perturbations in locations, since they are not smoother away from the origin than they are at the origin (Stein, 2005).

We adapt a class of nonseparable covariance functions in Gneiting (2002), and use

$$K(\boldsymbol{\Delta}, \delta) = \frac{\sigma^2}{(\phi_t^2|\delta|^2 + 1)} \left(1 + \frac{\phi_s||\boldsymbol{\Delta}||}{(\phi_t^2|\delta|^2 + 1)^{1/2}}\right) \exp\left[-\frac{\phi_s||\boldsymbol{\Delta}||}{(\phi_t^2|\delta|^2 + 1)^{1/2}}\right], \tag{12}$$

which yields a mean-square differentiable temporal process. This completes our specifications for $\mathbf{C_W}(\boldsymbol{\Delta}, \delta)$, which, by construction, is a valid cross-covariance because the required derivatives of $K(\boldsymbol{\Delta}, \delta)$ exist. Therefore, the collection of random variables $\{\mathbf{W}(\mathbf{s}_i, t_j) : (\mathbf{s}_i, t_j) \in \mathcal{S} \times \mathcal{T}\}$ has a well-defined multivariate normal distribution with a positive-definite $2N_sN_t(d+1) \times 2N_sN_t(d+1)$ variance-covariance matrix, which is an $N_sN_t \times N_sN_t$ block matrix and the block associated with the pair of space-time coordinates $(\mathbf{s}_i, t_k)$ and $(\mathbf{s}_j, t_l)$ is given by the $2(d+1) \times 2(d+1)$ matrix $\mathbf{C_W}(\mathbf{s}_j - \mathbf{s}_i, t_l - t_k)$.

For subsequent inferential development, we will be particularly interested in the joint distribution of the set $\{Z(\mathbf{s}_i, t_j) : (\mathbf{s}_i, t_j) \in \mathcal{S} \times \mathcal{T}\}$ and the spatiotemporal gradient process at an arbitrary space-time coordinate $(\mathbf{s}_0, t_0)$. We collect the $Z(\mathbf{s}_i, t_j)$'s into an $N_sN_t \times 1$ vector by first stacking them over time to form an $N_t \times 1$ column $\mathbf{Z}(\mathbf{s}_i)$ for each $\mathbf{s}_i \in \mathcal{S}$, and then stacking the $\mathbf{Z}(\mathbf{s}_i)$'s to form $\mathbf{Z} = (\mathbf{Z}(\mathbf{s}_1)^T, \mathbf{Z}(\mathbf{s}_2)^T, \ldots, \mathbf{Z}(\mathbf{s}_{N_s})^T)^T$. Note that $\mathbf{Z} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_Z)$, where $\boldsymbol{\Sigma}_Z$ is the $N_sN_t \times N_sN_t$ variance-covariance matrix whose elements correspond to

each pair of space-time coordinates in $\mathcal{S} \times \mathcal{T}$. The element associated with $(\mathbf{s}_i, t_k)$ and $(\mathbf{s}_j, t_l)$ is $K(\boldsymbol{\Delta}_{ij}, \delta_{kl})$, where $\boldsymbol{\Delta}_{ij} = \mathbf{s}_j - \mathbf{s}_i$ and $\delta_{kl} = t_l - t_k$. From (12), it is clear that $\boldsymbol{\Sigma}_Z = \sigma^2 \mathbf{R}_Z(\phi_s, \phi_t)$, where $\mathbf{R}_Z(\phi_s, \phi_t)$ is the corresponding correlation matrix.

Let $\boldsymbol{\nabla} Z(\mathbf{s}, t) = \left((\boldsymbol{\nabla}_s Z(\mathbf{s}, t))^T, \nabla_t Z(\mathbf{s}, t), (\boldsymbol{\nabla}_{s,t} Z(\mathbf{s}, t))^T\right)^T$ be the $(2d+1) \times 1$ spatiotemporal gradient process comprising the spatial, temporal and mixed derivatives of $Z(\mathbf{s}, t)$. For any arbitrary space-time coordinate $(\mathbf{s}_0, t_0)$, we can write $\text{Cov}\{\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0), \mathbf{Z}\}$ as the $(2d+1) \times N_s N_t$ matrix partitioned as $\boldsymbol{\nabla}\mathbf{K}_0 = [\boldsymbol{\nabla}\mathbf{K}_{0,1} : \boldsymbol{\nabla}\mathbf{K}_{0,2} : \ldots : \boldsymbol{\nabla}\mathbf{K}_{0,N_s}]$, where each $\boldsymbol{\nabla}\mathbf{K}_{0,i}$ is $(2d+1) \times N_t$ with j-th column $\boldsymbol{\nabla} K(\boldsymbol{\Delta}_{i0}, \delta_{j0})$ and with $\boldsymbol{\nabla} K(\boldsymbol{\Delta}, \delta)$ being defined analogously to $\boldsymbol{\nabla} Z(\mathbf{s}, t)$. Hence, the joint distribution of $\mathbf{Z}$ and $\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0)$ is

$$\begin{pmatrix} \mathbf{Z} \\ \boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_Z & (\boldsymbol{\nabla}\mathbf{K}_0)^T \\ \boldsymbol{\nabla}\mathbf{K}_0 & \mathbf{C}_{\nabla Z}(\mathbf{0}, 0) \end{bmatrix}\right) , \tag{13}$$

where $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0)$ is the cross-covariance matrix of $\boldsymbol{\nabla} Z(\mathbf{s}, t)$ evaluated at $(\mathbf{0}, 0)$. Since $Z(\mathbf{s}, t)$ and $\boldsymbol{\nabla} Z(\mathbf{s}, t)$ are subsets of a valid process $\mathbf{W}(\mathbf{s}, t)$, the positive-definiteness of the above variance-covariance matrix is implicit. The conditional distribution of the spatiotemporal gradient process $\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0)$ given $\mathbf{Z}$ is, therefore, a well-defined normal distribution. Subsequently, we show how to carry out posterior predictive inference on the gradient process.

3 Hierarchical modeling and inference

In the following we work with $d = 2$ and consider the spatiotemporal process model

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + Z(\mathbf{s}, t) + \epsilon(\mathbf{s}, t) \, , \tag{14}$$

where $\mu(\mathbf{s},t)$ captures large scale variation or trends (e.g., a regression model), $Z(\mathbf{s},t)$ is an underlying spatiotemporal process, and $\epsilon(\mathbf{s},t)$ is a zero centered white noise process to capture micro-scale variability and other unstructured random disturbances in the data. For (14), inferring on gradients associated with $Z(\mathbf{s},t)$ is more general than on gradients associated with $Y(\mathbf{s},t)$. The white-noise process in (14) introduces a discontinuity in $Y(\mathbf{s},t)$, so $\boldsymbol{\nabla} Y(\mathbf{s},t)$ does not exist. However, if $\mu(\mathbf{s},t)$ is smooth enough so that $\boldsymbol{\nabla}\mu(\mathbf{s},t)$ exists, then the formulation in Sections 2.1 and 2.2 allow legitimate inference for $\boldsymbol{\nabla}\mathrm{E}[Y(\mathbf{s},t)\,|\,\mu(\mathbf{s},t), Z(\mathbf{s},t)]$, which is simply the sum of the gradients for $\mu(\mathbf{s},t)$ and $Z(\mathbf{s},t)$. Without the white noise process in (14) (e.g., in models with no "nugget"), we can legitimately infer on $\boldsymbol{\nabla} Y(\mathbf{s},t)$. We adopt the Bayesian paradigm, which conveniently accommodates each of the above settings including, in particular, full inference for $\boldsymbol{\nabla} Z(\mathbf{s},t)$ at arbitrary $\mathbf{s}$ and t.

The model in (14) is applicable to settings where $Y(\mathbf{s},t)$ exists for every space-time coordinate in $\Re^d \times \Re$. While the data will be collected only over a finite subset $\mathcal{S}\times\mathcal{T} \subset \Re^d \times \Re$, inference may be sought at any location and time point, e.g., interpolation and prediction. For notational convenience, we will assume that every location in $\mathcal{S}$ produces observations over the same set of time points in $\mathcal{T}$, but this assumption is not necessary and, in fact, does not apply to our data analysis in the subsequent sections.

We now extend (14) to a Bayesian hierarchical model. Following customary assumptions, we model the trend using regressors indexed by both space and time, i.e., $\mu(\mathbf{s},t) = \mathbf{x}(\mathbf{s},t)^T\boldsymbol{\beta}$, and model $Z(\mathbf{s},t)$ as a spatiotemporal process specified by the covariance kernel in (12). Customary prior specifications produce the posterior distribution $p(\boldsymbol{\theta}, \mathbf{Z}\,|\,\mathbf{Y})$ proportional to

$$\begin{aligned} &U(\phi_s\,|\,a_{\phi_s}, b_{\phi_s}) \times U(\phi_t\,|\,a_{\phi_t}, b_{\phi_t}) \times IG(\sigma^2\,|\,a_\sigma, b_\sigma) \times IG(\tau^2\,|\,a_\tau, b_\tau) \times N(\boldsymbol{\beta}\,|\,\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \\ &\quad \times N\left(\mathbf{Z}\,|\,\mathbf{0}, \sigma^2\mathbf{R}_Z(\phi_s,\phi_t)\right) \times \prod_{i=1}^{N_s}\prod_{j=1}^{N_t} N(Y(\mathbf{s}_i,t_j)\,|\,\mathbf{x}(\mathbf{s}_i,t_j)^T\boldsymbol{\beta} + Z(\mathbf{s}_i,t_j), \tau^2)\ , \end{aligned} \tag{15}$$

where $\mathbf{Y}$ is the set of all observed $Y(\mathbf{s}_i, t_j)$'s and $\boldsymbol{\theta} = \{\phi_s, \phi_t, \sigma^2, \tau^2\}$ is the set of process parameters in the spatiotemporal and the white-noise processes. The parametrizations for the standard densities are as in Carlin and Louis (2009, Appendix A). We assume all the other hyperparameters in (15) are known.

We use Markov chain Monte Carlo (MCMC) to draw samples from the posterior distribution in (15) using Gibbs steps for all parameters except ϕ_s and ϕ_t, which will be updated using Metropolis steps. Sampling-based Bayesian inference is advantageous here as it seamlessly delivers inference on the residual space-time effects. That is, for an arbitrary space-time coordinate $(\mathbf{s}_0, t_0)$, we can sample from the posterior predictive distribution $p(Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Y}) = \int p(Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}, \boldsymbol{\theta}) p(\boldsymbol{\theta}, \mathbf{Z} \,|\, \mathbf{Y}) d\boldsymbol{\theta} d\mathbf{Z}$ using *composition sampling*. To be precise, for each posterior sample for the parameters in (15), say $\{\boldsymbol{\beta}^{(j)}, \boldsymbol{\theta}^{(j)}, \mathbf{Z}^{(j)}\}$, we draw $Z(\mathbf{s}_0, t_0)^{(j)}$ from $p(Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}^{(j)}, \boldsymbol{\theta}^{(j)})$ for $j = 1, 2, \ldots, M$, where M is the number of (post-burn-in) posterior samples. Predicting the outcome at $(\mathbf{s}_0, t_0)$ is also straightforward: drawing $Y(\mathbf{s}_0, t_0)^{(j)}$ from its full conditional distribution, $N(\mathbf{x}(\mathbf{s}_0, t_0)^T \boldsymbol{\beta}^{(j)} + Z(\mathbf{s}_0, t_0)^{(j)}, \tau^{2(j)})$, where we assume that $\mathbf{x}(\mathbf{s}_0, t_0)$ is available, yields samples from the posterior predictive distribution $p(Y(\mathbf{s}_0, t_0) \,|\, \mathbf{Y})$. Thus, the sampler adapts easily to situations where $Y(\mathbf{s}_0, t_0)$ is missing or unobserved, as these values can be estimated directly from their posterior predictive distributions. In practice, we recommend selecting $\mathbf{s}_0$ by laying a grid over (a subset of) the spatial domain and sample $Z(\mathbf{s}_0, t_0)$ at each location in the grid for a given time, t_0.

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

$$\begin{aligned} p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Y}) &= \int p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}) p(\mathbf{Z} \,|\, \mathbf{Y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \mathbf{Y}) d\boldsymbol{\theta} d\mathbf{Z} \\ &= \int p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}, \boldsymbol{\theta}) p(\mathbf{Z} \,|\, \mathbf{Y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \mathbf{Y}) d\boldsymbol{\theta} d\mathbf{Z} \end{aligned} \tag{16}$$

where the second equality follows from the fact that the gradient process is derived entirely

from its parent process, so $p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})$ does not depend upon $\mathbf{Y}$. From (13), it follows that $p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}, \boldsymbol{\theta})$ is multivariate normal with mean $(\boldsymbol{\nabla}\mathbf{K}_0)\,\boldsymbol{\Sigma}_Z^{-1}\mathbf{Z}$ and variance-covariance matrix $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0) - (\boldsymbol{\nabla}\mathbf{K}_0)\,\boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{\nabla}\mathbf{K}_0)^T$. Sampling from (16) proceeds, again, using composition: for each $\boldsymbol{\theta}^{(j)}$ and $\mathbf{Z}^{(j)}$, we draw $\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0)^{(j)} \sim p(\nabla Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}^{(j)}, \boldsymbol{\theta}^{(j)})$, which results in draws from (16). Therefore, inference on spatiotemporal gradients proceeds in posterior predictive fashion, and requires only the post-convergence MCMC samples for the spatiotemporal model parameters. See Web Appendix B for derivations of $\boldsymbol{\nabla}\mathbf{K}_0$.

4 Simulated Data Example

To demonstrate the effectiveness of our methods, we present an experiment using data generated from a true underlying gradient process, enabling us to validate our Bayesian estimation of spatiotemporal gradients. Our data is generated from $N_s = 100$ locations on a grid on the unit square with coordinates (s_{i1}, s_{i2}), for $i = 1, \ldots, N_t$. Each site is observed $N_t = 9$ times at evenly spaced increments, $t \in \{1, 2, \ldots, 9\}$, and all locations are observed at each time point, giving us a total of $N = 100 \times 9 = 900$ observations. To analyze these data, we use an intercept-only regression model, allowing the Gaussian process to capture all spatiotemporal variability. Our results are based on 10,000 iterations of the MCMC sampler, discarding the first 5,000 as burn-in.

We generate our data from a model where the true gradient is available in closed form:

$$Y(\mathbf{s}_i, t_j) \sim N\left(5\left[\sin(s_{i1}3\pi) + \cos(s_{i2}3\pi)\cos(t_j\pi/7)\right], \tau^2\right). \tag{17}$$

The reason we chose the expression in (17) is two-fold: not only does it allow us to assess the quality of our gradient estimates, but it also features an interesting interaction between space and time. Specifically, in the s_{i2} direction, our spatial gradients evolve over time,

allowing us to determine if our mixed gradients (and thus the spatial and the temporal gradients) are able to accurately estimate the true values, and if our model is capable of detecting *significant* mixed gradients. By contrast, the s_{i1} direction does *not* evolve over time, providing us with temporal replicates of their gradients to help identify locations in which the gradient estimation performed well and where it did not. We can also check how our mixed gradients perform when there are truly no interesting trends to be found.

Our Bayes-MCMC procedure accurately estimated the error variance parameter as $\tau^2 = 0.93$ with a 95% CI of (0.84, 1.04), covering its true value of 1. We also fared well regarding the spatiotemporal random effects, $\mathbf{Z}$, with 96% of the 900 CI's for $Z(\mathbf{s}_i, t_j)$ covering their true values. Gradients are computed on a grid comprised of points $\mathbf{s}_{i0} = (s_{i01}, s_{i02})$ where $s_{i0j} \in \{0.05, 0.15, \ldots, 0.95\}$ for $j = 1, 2$ at times $t_0 \in \{1, 2, \ldots, 15\}$. Overall, our gradient estimation methods performed quite well: 99% of our estimated gradients' CI's covered their true values. In Figure 1, we display the true temporal and the s_{i2} mixed gradients at $t_0 = 3$ compared to their estimated values, as well as a map displaying significant gradients. Our estimated gradients appear to accurately capture the true underlying gradient surface, and we generally see significant gradients where we would expect. For instance, the temporal gradient at location (0.5, 0.5) in Figure 1(a) is roughly equal to 0 and this location appears to be in the middle of a downward slope in the vertical axis. Looking at this location in Figure 1(b), we find that this location indeed has a significantly negative gradient in the s_{i2} direction. Also, though not shown, we find zero significant mixed gradients in the s_{i1} direction, as we had anticipated.

Figure 1(b) also highlights an important issue regarding statistical significance in the mixed gradients. Based on our work, detecting significant mixed gradients near the boundaries of the spatial domain can be difficult. For instance, when computing mixed gradients in the s_{i2} direction (i.e., the "north/south" gradient), the estimated mixed gradients in the

middle panel of Figure 1(b) accurately reflect the true underlying gradient process. However, the horizontal bands across the top and bottom of the spatial domain fail to consistently achieve significance. We believe this is due in part to the fact that we're conducting inference two levels of derivatives removed from the estimated spatio-temporal process, and thus we encounter a considerable amount of variability in these estimates. Furthermore, this issue is particularly exacerbated across these bands, as we can essentially only learn about the gradient process from a single direction. As such, we are less concerned with detecting significance in the mixed gradients; instead we focus our attention on the *patterns* observed.

This example illustrates how an investigation of the spatiotemporal gradient process could be used to identify important missing covariates. Based on the lack of interesting patterns in the mixed gradients found in the s_{i1} direction, this suggests that the missing factors do not substantially vary in this direction. On the other hand, the patterns observed in Figure 1(b) indicate the potential for an important factor that moves in the s_{i2} over time. Recalling the structure in (17), we see that both of these conclusions are accurate.

5 Data Analysis

Our data consist of concentrations of PM2.5 (in parts per million; ppm) measured at monitoring sites throughout the state of California and were collected by and obtained from the Air Resources Board of the California Environmental Protection Agency. We observe 777 observations collected during June 2008 from 79 monitoring sites across the state. Of these, 10 sites record observations nearly every day (26+ days out of 30), while the rest are observed only observed once every three to six days. For instance, we observe data from 78 of our monitoring sites on June 11, yet only nine sites are observed on the 9th and ten sites on the 10th, 12th and 13th.

In Figure 2(a), we display the locations from which our data were observed, with black circles denoting sites observed once every three to six days and red triangles denoting sites observed on an almost daily basis. Note that while the monitoring sites are relatively well dispersed throughout the state, only two monitoring sites in Northern California (and only one as far north as Sacramento) are observed daily. This lack of information may pose challenges such as a decrease in precision, which may lead to a difficulty to detect significance in the gradient process developed in Section 3. This sparsity of observations may also be seen in Figure 2(b), which displays the distribution of the observed PM2.5 concentrations.

In addition to exact spatial coordinates (in latitude and longitude), each monitoring site is attributed with an elevation covariate. For convergence purposes, we have centered and scaled elevation to avoid unnecessary correlation between samples of our slope and intercept coefficients. We use the sinusoidal projection $s_{i1} = E\lambda\cos\theta$ and $s_{i2} = E\theta$, where $E =$ 6371km is the radius of the earth, and θ and λ denote latitude and longitude, respectively. Distances were computed by letting one spatial unit equal 100km, while temporal distances were computed by letting one temporal unit equal one week. In order to make informed decisions regarding air quality warnings, public health officials rely on statistical models to predict each day's pollution levels. In the analysis of pollution data, weather-related covariates such as wind speed and wind direction are typically included, however the only covariate information available to us here is the elevation of the monitoring site.

A summary of our full model's parameter estimates can be found in Table 1. As expected, the regression coefficient for elevation is significantly negative, suggesting that locations at higher elevation tend to have lower levels of pollution. The samples of our variance parameters, τ^2 and σ^2, can be used to compute the posterior distribution for $\sigma^2/(\sigma^2+\tau^2)$, which has a median (95% CI) of 0.986 (0.98, 0.99), indicating that 98.6% of the residual variability in our data is being explained by the spatiotemporal process. This high a value

is not surprising given the limited number of covariates in our model. Finally, our spatial and temporal range parameters, ϕ_s and ϕ_t, can be interpreted as controlling how quickly the correlation between two observations drops off. For instance, the spatial correlation between two observations 100 km apart but recorded at the same time has a 95% CI of (0.44, 0.59), and the temporal correlation between two observations recorded at the same location 2 days apart has a 95% CI of (0.21, 0.44). Similarly, the spatiotemporal correlation between two observations 100 km and 2 days apart has a 95% CI of (0.17, 0.32).

Figure 3 displays the maps of the fitted values (i.e., $\widehat{Y}(\mathbf{s}_i, t_j) = \mathbf{x}(\mathbf{s}_i, t_j)^T\widehat{\boldsymbol{\beta}} + \widehat{Z}(\mathbf{s}_i, t_j)$) for each of the days in the study period. Given our limited number of covariates, the majority of the variation in this figure is due to the variability in $\mathbf{Z}$; as such, our focus will be on the underlying spatiotemporal process. Here, we do not see much day-to-day variability during the first 3 weeks of the month, save for a slight increase on June 12. Things change dramatically during the week of June 22, however, when a rash of wildfires in Northern California occurred, causing substantial levels of particulate matter to be released.

In Figure 4, we display maps of the spatial gradients for the transition between June 12 and June 13; for comparison purposes, we also display a topographical map of California in Figure 4(c). These maps represent a typical day during the first three weeks of June, when the topography of California appears to play a major role in the patterns we see. For instance, we can clearly see that the Coast Ranges (along the coast from northern California to the Los Angeles area) and Transverse Ranges (stretching from west to east, creating the northern border of Los Angeles) appear to strongly influence the gradient estimates. During this three-week time period, we fail to detect any significant gradient activity.

The gradient story changes dramatically at the start of our fourth week, however, when a series of dry thunderstorms crossed the northern half of the state on June 20 and June 21, bringing with it over 5,000 lightning strikes and causing over 2,000 fires (California

Department of Forestry and Fire Protection, 2008). As these fires spread over the course of the weekend, levels of PM2.5 increased substantially, as evidenced by the map of temporal gradients from June 21 in Figure 5. Over the course of the following week, we find a large number of significant temporal gradients, corresponding first to the increased levels of smoke and ash being released into the air. As the military and fire departments from across the nation battled the fire, we see levels begin to drop at the end of the month. As in our simulated example, identifying significance in the mixed gradient surfaces proves challenging. While this may be due in part to the sparsity of monitoring sites in Northern California, this may also indicate that the wildfires remained contained to one (albeit large) area. Maps of all of the spatiotemporal gradients can be found in the Web Appendix C.

6 Discussion

We have outlined a fully process-based inferential framework for infinitesimal rates of change in space and time within a Bayesian setting, expanding upon earlier work on rates of change in purely spatial or in discrete-space continuous-time settings. Full posterior distribution of spatiotemporal gradients can be obtained in a posterior predictive fashion at arbitrary locations and timepoints. This is attractive from an execution standpoint because we do not need to recompute the posterior distribution of the model parameters when inference is sought on new locations and timepoints; e.g., we can store the post burn-in posterior samples of the model parameters and estimate the spatiotemporal gradients using composition sampling, which is fast and exact. Our sampling-based Bayesian framework outlined in Section 3 also easily adapts to finding gradients from *misaligned* or *unbalanced* spatiotemporal data, where the outcome may not have been observed in the same set of locations across time.

Turning to our analysis of the California air pollution data, supplying the posterior distribution for the spatiotemporal gradients can help investigators determine the nature of the important factor(s) not accounted for in our mean model. During the first 3 weeks of June 2008, the spatial gradients consistently display a trend corresponding to the topography of California. This could also be inferred by the lack of interesting patterns in the mixed gradients during this period, suggesting that the observed spatial gradients were due to "fixed" spatial features. Then, once the lightning storm passed and left a trail of wildfires in its wake, we suddenly observe new patterns in our gradients. Based on our results, it appears the pollution levels increase primarily over time rather than over space; i.e., once the storm hit and the fires started, pollution levels in the affected areas rose dramatically over the entire region. As such, while we see some activity in both the spatial and mixed gradients, the majority of the action here occurs in the temporal gradient process. If our data were collected at a finer spatiotemporal resolution, we may have been able to track the trajectory of the storm as it set the state's forests ablaze using our entire gradient process.

Some alternate methods present themselves immediately. One could, for example, construct finite-difference spatiotemporal processes, e.g., $\frac{Z(\mathbf{s}+h\mathbf{u},t+h)-Z(\mathbf{s},t)}{h}$. Inference for such processes is straightforward and requires no new methodological developments. Furthermore, these processes may be able to approximate the gradient processes effectively. However, we may encounter numerical instabilities with finite difference processes when inference is sought at finer resolutions, i.e., smaller values of h (e.g., Banerjee et al., 2003). While the distribution theory for gradients becomes more complicated as we pass to L_2 limits, we obtain tractable joint distributions that do not encounter any numerical problems.

Also, we have focused only upon fully process-based models, where space and time are both treated as continuous. There are a variety of alternate models, such as dynamic linear models, that model spatiotemporal data. These alternate models, almost invariably, treat

only one of space and time as continuous. For example, dynamic linear models usually treat space as continuous but time as discrete. We can still infer about spatial gradients over time, but not about temporal derivatives. Extensions to non-Gaussian data are achieved by replacing the Gaussian likelihood in (14) with an exponential family.

Finally, we provide some pointers toward future work. Our developments naturally extend to machine learning and functional regression contexts, where $Z(\mathbf{x}, t)$ is now a dynamic Gaussian process on the predictor space and inference is sought on the rates of change of this process, both with respect to the predictors $\mathbf{x}$ and time t. We also intend to extend this work to massive datasets by extending the theory for gradients to dimension-reducing spatiotemporal (or dynamic functional regression) processes. Our current work can also be extended to gradient analysis for spatiotemporal Dirichlet processes, building upon earlier work in purely spatial settings by Guindani and Gelfand (2006). When the number of spatial locations and/or temporal units become large, implementing spatiotemporal processes become cumbersome. Inference can still proceed from dimension-reducing and computationally scalable predictive process models that can be adjusted with tapered covariance functions (e.g., Sang and Huang, 2011). While gradients can be estimated from the predictive process (see, e.g., Liang, Banerjee and Carlin, 2009), choosing an appropriately smooth tapered covariance function will likely perform better. Lastly, we intend to explore gradient theory in cases where second-order stationarity is not a valid assumption. This extension is nontrivial as nonstationary covariance functions do not submit themselves as easily to inference for estimation on gradients.

Supplementary Materials

Web Appendices referenced in Sections 3, 4, and 5 are available with this paper at the Biometrics website on Wiley Online Library.

References

Banerjee, S. (2010). Spatial gradients and wombling. In *Handbook of Spatial Statistics*, eds. A.E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. Boca Raton, FL: CRC Press, pp. 559–575.

Banerjee, S. and Gelfand, A.E. (2006). Bayesian wombling: Curvilinear gradient assessment under spatial process models. *Journal of the American Statistical Association*, **101**, 1787–1501.

Banerjee, S., Gelfand, A.E. and Sirmans, C.F. (2003). Directional rates of change under spatial process models. *Journal of the American Statistical Association*, **98**, 946–954.

Barbujani, G., Oden, N.L., and Sokal, R.R. (1989). Detecting areas of abrupt change in maps of biological variables. *Systematic Zoology*, **38**, 376–389.

Bochner, S. (1955). *Harmonic analysis and the theory of probability*, University of California Press.

Brown, P.E., Kåresen, K.F., Roberts, G.O., and Tonellato, S. (2000). Blur-generated non-separable space-time models. *Journal of the Royal Statistical Society, Ser. B*, **62**, 847–860.

California Department of Forestry and Fire Protection (2008). June 2008 Fire Siege. http://www.fire.ca.gov/fire_protection/downloads/siege/2008/2008FireSiege_full-book_r6.pdf

Carlin, B.P. and Louis, T.A. (2009). *Bayesian Methods for Data Analysis*, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Christakos, G. (1992). *Random Field Models in Earth Sciences*, San Diego, CA: Academic Press.

Christakos, G. (2000). *Modern Spatiotemporal Geostatistics*, New York: Oxford University Press.

Cressie, N. and Huang, H.C. (1999). Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330–1340.

Cressie, N. and Wikle, C.K. (2011). *Statistics for Spatio-Temporal Data*, Hoboken, NJ: John Wiley & Sons.

De Iaco, S., Myers, D.E., and Posa, D. (2002). Space-time variograms and a functional form for total air pollution measurements. *Computational Statistics and Data Analysis*, **41**, 311–328.

Environmental Protection Agency (2012). Ground-level Ozone: Basic Facts. http://www.epa.gov/air/ozonepollution/basic.html

Fortin, M.J. (1994). Edge detection algorithms for two-dimensional ecological data. *Ecology*, **75**, 956–965.

Fortin, M.J. (1997). Effects of data types on vegetation boundary delineation. *Canadian Journal of Forest Research*, **27**, 1851–1858.

Fortin, M.J. and Drapeau, P. (1995). Delineation of ecological boundaries: Comparisons of approaches and significance tests. *Oikos*, **72**, 323–332.

Gabriel, E., Allard, D., and Barco, J.-N. (2011). Estimating and testing zones of abrupt change for spatial data. *Statistics and Computing*, **21**, 107–120.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, **97**, 590–600.

Guindani, M. and Gelfand, A.E. (2006). Smoothness properties and gradient analysis under spatial dirichlet process models. *Methodology and Computing in Applied Statistics*, **8**, 159–189.

Hartfield, M. and Gunst, R. (2003). Identification of model components for a class of continuous spatiotemporal models. *Journal of Agricultural, Biological, and Environmental Statistics*, **8**, 105–121.

Jones, R.H. and Zhang, Y. (1997). Models for continuous stationary space-time processes. In *Modelling Longitudinal and Spatially Correlated Data*, eds. T.G. Gregoire, D.R. Brillinger, P.J. Diggle, E. Russek-Cohen, W.G. Warren, and R.D. Wolfinger. New York, NY: Springer-Verlag, pp. 289–298.

Liang, S., Banerjee, S. and Carlin, B.P. (2009). Bayesian wombling for spatial point processes. *Biometrics*, **65**, 1243–1253.

Ma, C. (2003). Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference*, **116**, 489–501.

Majumdar, A., Munneke, H.J., Gelfand, A.E., Banerjee, S., and Sirmans, C.F. (2006). Gradients in spatial response surfaces with application to urban land values. *Journal of Business and Economic Statistics*, **24**, 77–90.

Manolopoulou, I., Matheu, M.P., Cahalan, M.D., West, M. and Kepler, T.B. (2012). Bayesian spatio-dynamic modeling in cell motility studies: Learning nonlinear toxic

fields guiding the immune response. *Journal of the American Statistical Association*, **107**, 855–865.

Morris, M.D., Mitchell, T.J., and Ylvisaker, D. (1993). Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics*, **35**, 243–255.

O'Hagan, A. (1992). Some Bayesian numerical analysis (with discussion). In *Bayesian Statistics 4*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. Oxford: Oxford University Press, pp. 345–363.

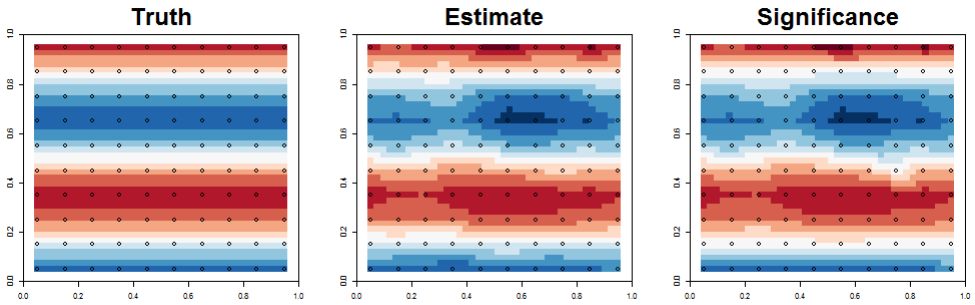
Quick, H., Banerjee, S., and Carlin, B.P. (2013). Modeling temporal gradients in regionally aggregated California asthma hospitalization data. *Annals of Applied Statistics*. **7**, 154–176.

Sang, H. and Huang, J.Z. (2012). A full-scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society, Series B*. **74**, 111–132.

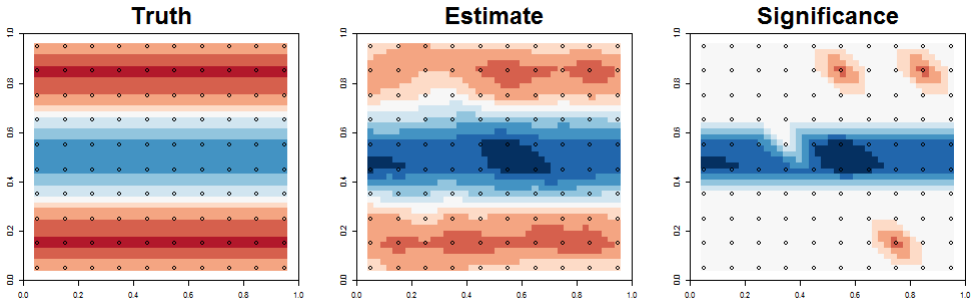
Stein, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.

Stein, M.L. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, **140**, 310–321.

Womble, W.H. (1951). Differential systematics. *Science*, **114**, 315–322.



(a) Temporal Gradient

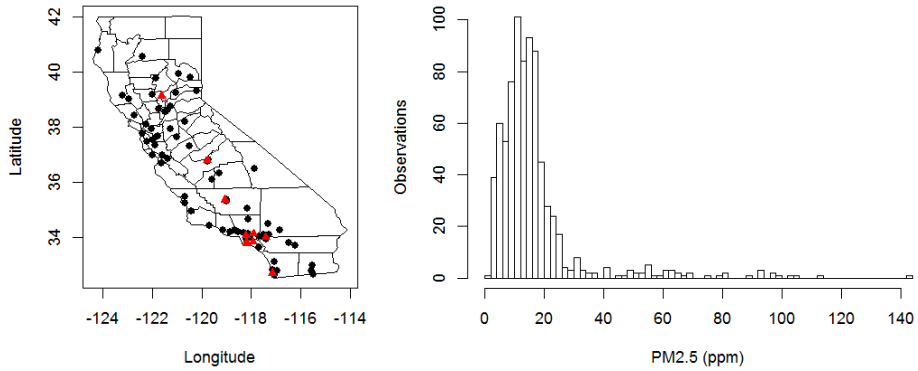


(b) s_{i2} Mixed Gradient

Figure 1: Comparison of the true temporal and the s_{i2} mixed gradients for a particular time point and their posterior median values based on our gradient theory. The color scheme here goes from highly negative (blue) to highly positive (red), centered around 0 (light gray), with locations of observed locations plotted as open circles. The third panel highlights significant gradients; i.e., gradients whose 95% CI's contain 0 are assigned value 0, while gradients whose 95% CI's do not contain 0 are assigned their posterior medians.

Parameter	Median (95% CI)	Parameter	Median (95% CI)
β_0 (Intercept)	16.561 (16.414, 16.707)	σ^2 (S-T Var)	306.457 (247.138, 404.785)
β_1 (Elevation)	-1.654 (-2.589, -0.606)	ϕ_s (Spatial)	1.647 (1.407, 1.863)
τ^2 (Error Var)	4.415 (3.315, 5.855)	ϕ_t (Time)	4.908 (3.944, 6.876)

Table 1: Parameter estimates from the analysis of the PM2.5 data. ϕ_s is on a scale of 1 unit = 100 km, and ϕ_t is on a scale of 1 unit = 1 week.



(a) Monitoring Sites (b) Histogram of PM2.5

Figure 2: Left panel displays the locations of the monitoring sites observed in the PM2.5 data. Black circles denote sites observed once every three to six days, while red triangles denote sites observed nearly daily. Right panel displays the distribution of observed PM2.5 concentrations (in ppm).

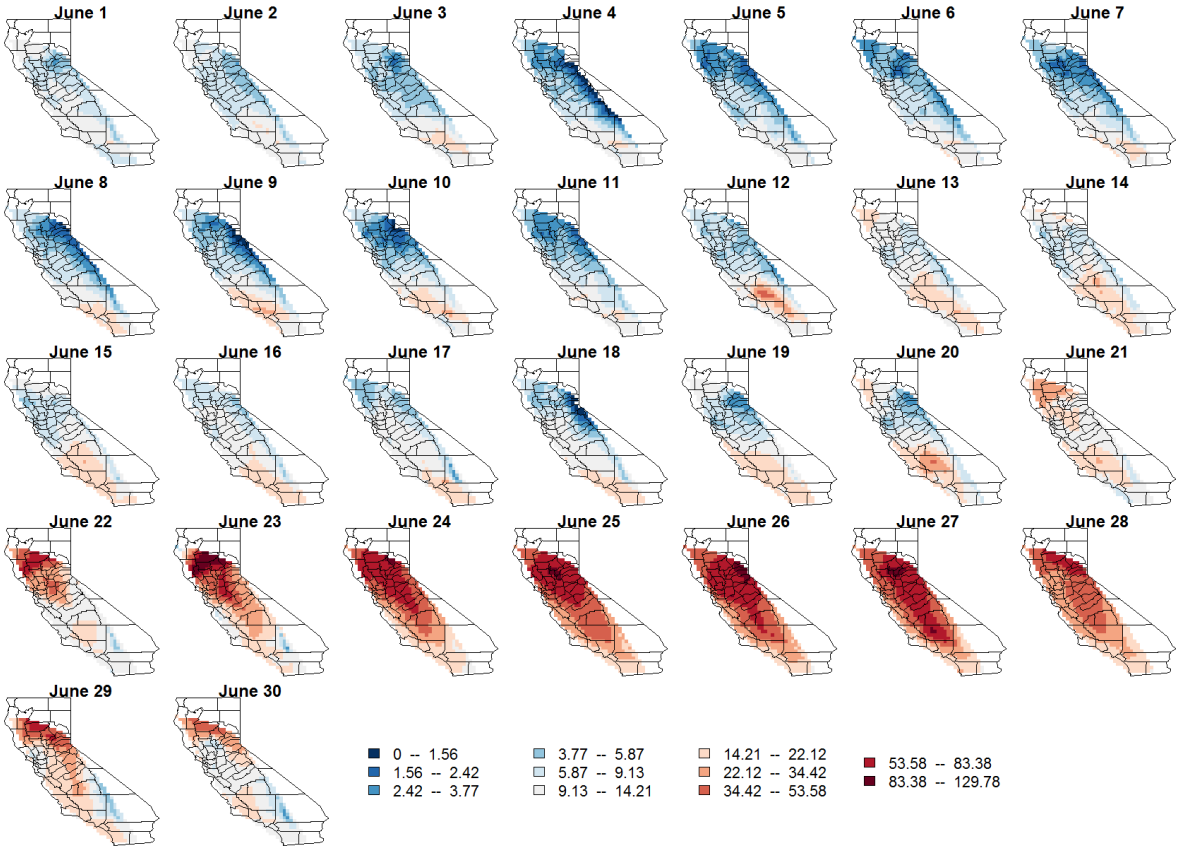
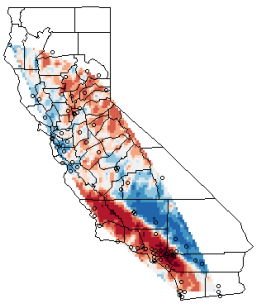
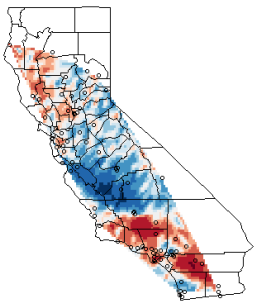
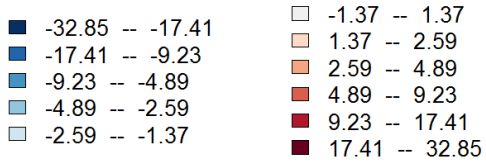


Figure 3: Predicted levels of PM2.5 for the month of June 2008. The maps are situated so that the first column of days are on Sunday, the second are Mondays, etc.



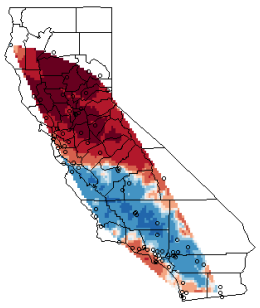


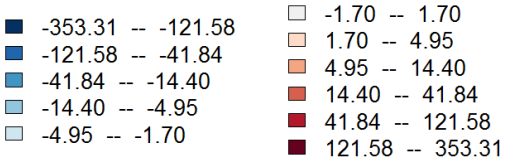
(a) Eastern Gradient (b) Northern Gradient



(c) Topographical Map (d) Spatial Gradient Key (per 100km change)

Figure 4: Estimated spatial gradient plots for the transition period between June 12 and June 13, as well as a topographical map of California for comparison purposes. Spatial gradients in the southern portion of the state appear to correspond to the Transverse Ranges, which wrap around north and east of Los Angeles. We are also able to identify the valley that runs vertically through the state.





(a) Temporal Gradient (b) Temporal Gradient Key (per 1 week change)

Figure 5: Estimated temporal gradient plots for the transition period between June 21 and June 22. Positive gradients in the northern California likely correspond to wildfires resulting from a series of lightning storms on the afternoon of June 21.