Published in final edited form as:

Spat Stat. 2012 December 1; 2: 15–32. doi:10.1016/j.spasta.2012.05.001.

Kernel Averaged Predictors for Spatio-Temporal Regression Models

Matthew J. Heaton¹ and Alan E. Gelfand²

¹Department of Statistical Science, Duke University, Box 90251, Durham, NC 27708-0251 (heaton@ucar.edu; phone: 303-497-2884)

²J.B. Duke Professor of Statistics and Decision Sciences, Duke University, Box 90251, Durham, NC 27708-0251 (alan@stat.duke.edu)

Abstract

In applications where covariates and responses are observed across space and time, a common goal is to quantify the effect of a change in the covariates on the response while adequately accounting for the spatio-temporal structure of the observations. The most common approach for building such a model is to confine the relationship between a covariate and response variable to a single spatio-temporal location. However, oftentimes the relationship between the response and predictors may extend across space and time. In other words, the response may be affected by levels of predictors in spatio-temporal proximity to the response location. Here, a flexible modeling framework is proposed to capture such spatial and temporal lagged effects between a predictor and a response. Specifically, kernel functions are used to weight a spatio-temporal covariate surface in a regression model for the response. The kernels are assumed to be parametric and non-stationary with the data informing the parameter values of the kernel. The methodology is illustrated on simulated data as well as a physical data set of ozone concentrations to be explained by temperature.

Keywords

Distributed lag; Stochastic integral; Gaussian process; Ozone

1. Introduction

Consider quantifying the effect of a covariate on a response where both the covariate and response are point and time-referenced over a spatio-temporal domain. More concretely, let $Y(s, t)$ denote a response variable at location s and time t and $X(s, t)$ denote a spatiotemporal covariate that is, potentially, associated with the response (unless explicitly stated, $Y(s, t)$ and $X(s, t)$ will be univariate). To simplify matters, throughout this article assume $(s, t) \in \mathbb{R}^d \times \mathbb{R}$ for some $d \in \mathbb{N}$ but note that if $(s, t) \in \mathcal{D} \times \mathcal{T}$ for bounded domains $\mathcal{D} \subset \mathbb{R}^d$ and $\mathcal{T} \subset \mathbb{R}$ then the methods below still apply with minimal alteration (see Section 2 for more details). Methods for capturing the spatio-temporal correlations within $Y(s, t)$ and $X(s, t)$ are abundant with reviews provided by Stein [1] and Gneiting et al. [2]. When relating $X(s, t)$ to $Y(s, t)$, the common method is to do so linearly through the mean according to,

Correspondence to: Matthew J. Heaton.

Present address: National Center for Atmospheric Research, PO Box 3000, Boulder, CO 80307.

$$\ell(\mathbb{E}(Y(s, t)|X(s, t), \beta_0, \beta_1)) = \beta_0 + \beta_1 X(s, t), \quad (1)$$

where $\ell(\cdot)$ is an appropriate link function (e.g. identity, log, etc.). Common extensions of (1) include spatially varying coefficient models [3] and dynamic spatial process models [4, 5, 6]. However, a fundamental assumption of (1), and its common extensions, is that only $X(s, t)$ affects $\mathbb{E}(Y(s, t))$ and neighboring covariate levels $X(s', t')$ for (s', t') close to (s, t) do not. In essence, by Equation (1) the relationship between Y and X is confined to a single spatial location and time period. However, if $Y(s, t)$ and/or $X(s, t)$ exhibit spatio-temporal correlation then the relationship between them may be more complex.

For example, ground-level ozone is the primary constituent of smog and has been linked to various negative health outcomes associated with the lungs such as chest pain, asthma, and bronchitis (www.epa.gov/ozone). For these reasons, the Environmental Protection Agency (EPA) monitors the levels of ozone near urban areas of the United States. Ozone formation is the result of a chemical reaction between volatile organic compounds (VOC) and nitric oxide (NO_x) in the presence of sunlight (i.e. solar radiation). In the absence of solar radiation data, temperature is often used as a surrogate predictor of the concentration of ozone [7, 8]. Specifying a statistical model for the relationship between ozone concentrations and temperature, however, may be more difficult than simply regressing ozone concentrations on temperature at a given location and time period. For example, if temperatures have been high for several days, ozone concentration may also be higher because such conditions, potentially, allow a greater number of reactions between VOC and NO_x to take place. Similarly, in the presence of wind, temperatures at one location in recent days may affect ozone concentrations at a different location on the present day. Finally, as temperature is merely a surrogate for solar radiation, the relationship between temperature and ozone concentration may be more spatially and temporally complex than had solar radiation been used directly. Each of these possibilities suggests that the effect of temperature on ozone concentration may be spatially and/or temporally lagged.

Other examples where spatio-temporal lagged effects occur include the effect of pollution on public health [9, 10, 11], economic indicators on consumption [12], and disease incidence on disease propagation [13]. In all of these examples, the relationship between the response and covariate is not confined to a single spatio-temporal location and lagged effects need to be incorporated into the statistical model.

Models with temporally lagged effects are not new with the most common being the distributed lag model of Almon [14] and its variations [see 12, and the references therein]. Distributed lag models extend (1) to

$$\ell(\mathbb{E}(Y(s, t)|X(s, t), \dots, X(s, t-L), \beta_0, \alpha_0, \dots, \alpha_L)) = \beta_0 + \sum_{l=0}^L \alpha_l X(s, t-l), \quad (2)$$

for some known maximum lag L . Alternatively, L could be infinite yielding the Koyck distributed lag model [15, 16]. Distributed lag models, however, suffer from several limitations. First, if $X(s, t)$ exhibits strong temporal correlation then the set of covariates $\{X(s, t-l) : l = 0, \dots, L\}$ are highly collinear resulting in unstable estimates of the coefficients α_l , $l = 1, \dots, L$. To stabilize the estimates, various constraints are imposed on $\{\alpha_l\}$. For example, $\{\alpha_l\}$ may be assumed to follow some function such as a polynomial [9] or spline [17]. Welty et al. [11] build constraints into a prior distribution and estimate the coefficients from a Bayesian perspective. Second, (2) is typically constrained to time periods where $X(s, t)$ is observed. That is, only observed X 's are used. For problems where covariates are measured at different locations or time scales than $Y(s, t)$, or there is much

missing data, using only observed $X(s, t)$ may result in undesirable model restrictions. Third, (2) only accounts for temporal lags while ignoring spatially “lagged” effects.

The problem of incorporating spatially “lagged” effects (i.e. effects spread over space) has recently been taken up by Heaton and Gelfand [18]. They propose the use of kernels to properly weight a spatial covariate surface in a regression model for the response. That is, they propose

$$\mathbb{E}(Y(s)) = \beta_0 + \beta_1 \int_{\mathbb{R}^d} K(s, r) X(r) dr \quad (3)$$

where K is a kernel (density) function properly normalized to integrate to one over $\mathbb{R}^d$. They find that, depending on the kernel specification, (3) can capture a variety of spatially lagged effect behaviors. However, their approach is limited in that their kernels do not incorporate temporally lagged effects. Furthermore, calculation of the likelihood for their models is computationally challenging as they rely on expensive Monte Carlo integration techniques.

The goal of this article is to extend the kernel averaged predictor models of Heaton and Gelfand [18] to the spatio-temporal setting. Specific extensions are as follows. First, the purely spatial kernels of Heaton and Gelfand [18] are extended to spatio-temporal kernels to capture spatial *and* temporal lagged effects between the predictors and the response. Hence, the models herein are referred to as spatio-temporal kernel averaged predictor (STKAP) models. Second, non-stationary kernels (i.e. kernels which vary over space and time) are considered to create the spatially and temporally averaged covariate (say, $\tilde{X}(s, t)$) which is then used as a predictor for $Y(s, t)$. The non-stationary kernels proposed here are assumed to be parametric with parameters which vary over the domain thus allowing the space-time interactions between X and Y to also vary over the spatio-temporal region. And, third, this article defines the kernel averaged predictors $\tilde{X}(s, t)$ using a discrete measure (as opposed to Lebesgue measure) such that the associated computational burden of fitting the STKAP model is lightened in that covariance calculation is done via matrix operations rather than Monte Carlo integration.

In practice, the entire covariate surface $\{X(s, t) : s \in \mathbb{R}^d, t \in \mathbb{R}\}$ will not be available. Rather, applications will observe the covariate surface at only a finite number of spatio-temporal locations. Therefore, in order to define a model which is not dependent on the sampling locations and sampling frequency, a Gaussian process model is used for $X(s, t)$. In this way, the entire covariate surface is used as a predictor for $Y(s, t)$ as opposed to only observed $X(s, t)$ as in (2); thus, the models here are not constrained to observed time periods or spatial locations. Additionally, by using a Gaussian process model for $X(s, t)$, the models here can be fit in situations where the responses and covariates are measured at different spatial locations (spatial misalignment) and when the response or covariate is missing for a particular spatio-temporal location.

In concluding this introduction, the motivation for this work is to capture spatio-temporal explanatory relationships between responses and predictors. In other words, STKAP models are primarily concerned with inference on the parameters in an explanatory model for the response surface given the covariate surface. They are not intended for prediction (kriging) and would not be expected to improve upon Bayesian kriging. This focus on inference does not discount the usefulness of the models developed herein. In many spatio-temporal applications (e.g. pollution and morbidity, heat and mortality, disease mapping, etc.), the goal of data analysis is often inference on the effect of covariates rather than prediction to unobserved spatial locations. This point is discussed further below.

This article proceeds as follows. Section 2 details the STKAP model providing associated distribution theory and model properties. Section 3 discusses implementation issues regarding STKAP models. Section 4 provides guidelines and precautions for specifying appropriate kernels and proposes a spatially non-stationary kernel. Section 5 evaluates the properties of the STKAP model in practice via simulation studies and then applies the model to a physical data set of ozone concentrations to be explained by average daily temperature. Section 6 provides a summary and outlines directions for future work.

2. Spatio-Temporal Kernel Averaged Predictor Models

2.1. Model Specification

Assume $Y(s, t)$ and $X(s, t)$ take values in $\mathbb{R}^1$ and that a bivariate Gaussian process model is appropriate for X and Y (perhaps following a transformation). In particular, specify the bivariate process as follows. Let $X(s, t)$ follow a Gaussian process (GP) such that

$$X(s, t) = \mu_X(s, t) + \sigma_X w_X(s, t) \quad (4)$$

where $\mu_X(s, t)$ is the spatio-temporal mean surface, σ_X is the spatio-temporal standard deviation, $w_X(s, t)$ follows a mean 0, unit variance spatio-temporal Gaussian process with valid (i.e. positive definite) space-time correlation function denoted by,

$$\text{Corr}(w_X(s, t), w_X(s', t') | \phi_X) \equiv \rho_X(s, s'; t, t' | \phi_X) \quad (5)$$

and ϕ_X is a vector of parameters for ρ_X . For simplicity in notation, the vector ϕ_X will be suppressed and (5) will be written as $\rho_X(s, s'; t, t')$. The space-time covariance function ρ_X is not restricted here; the interested reader is directed to Gneiting [19], Stein [1], and the review by Gneiting et al. [2] for discussions on valid space-time covariance functions and their properties. For applications where covariates are measured with error, the observed covariate would be

$$H(s, t) = X(s, t) + \tau_H \varepsilon_H(s, t)$$

where $\varepsilon_H(s, t)$ follows a mean 0, unit variance white noise and τ_H is the standard deviation of the measurement error. Even if measurement error is present, the models below are still defined in terms of $X(s, t)$ because the goal is to capture the effect of the true covariate levels on the response. For simplicity, for the remainder of this article assume $\tau_H = 0$.

To account for the effect of spatial and temporal covariate neighbors on $Y(s, t)$, define,

$$Y(s, t) | \tilde{X}(s, t) = \beta_0 + \beta_1 \tilde{X}(s, t) + \sigma_Y w_Y(s, t) + \tau_Y \varepsilon_Y(s, t) \quad (6)$$

where $\tilde{X}(s, t)$ is a spatially and temporally weighted (smoothed) covariate, $w_Y(s, t)$ and $\varepsilon_Y(s, t)$ are defined similarly to $w_X(s, t)$ and $\varepsilon_H(s, t)$ above with

$$\text{Corr}(w_Y(s, t), w_Y(s', t') | \phi_Y) \equiv \rho_Y(s, s'; t, t' | \phi_Y),$$

σ_Y and τ_Y are the spatio-temporal and nugget standard deviations, respectively, and ϕ_Y is a vector of parameters associated with ρ_Y . Spatially and temporally lagged effects enter into Equation (6) via $\tilde{X}(s, t)$ which is defined as,

$$\tilde{X}(s, t) = \int_{\mathbb{R}^d} \int_{-\infty}^t K_{st}(\mathbf{r} - \mathbf{s}; u - t) X(\mathbf{r}, u) \nu(d\mathbf{r} du), \quad (7)$$

$K_{st}(\mathbf{r} - \mathbf{s}; u - t) = 0$ for all $(s, t, \mathbf{r} - \mathbf{s}, u - t)$ is a spatio-temporal smoothing kernel normalized to integrate to one (i.e. $\int_{\mathbb{R}^d} \int_{-\infty}^0 K_{st}(\mathbf{r}, u) \nu(d\mathbf{r} du) = 1$ and $\nu(\cdot)$ is a measure (typically Lebesgue or discrete). For problems where the spatio-temporal domain of interest is compact (e.g. $\mathcal{D} \times \mathcal{T}$ where $\mathcal{D} \subset \mathbb{R}^d$ and $\mathcal{T} \subset \mathbb{R}$) then the limits of integration in (7) are changed to $\mathcal{D} \times \mathcal{U}$ where $\mathcal{U} = \{u - t : u \in \mathcal{T}\}$. Similarly, if time is assumed to be discrete then the second integral in (7) is replaced by a summation. In any case, $K_{st}(\mathbf{r} - \mathbf{s}; u - t)$ should always be normalized to integrate to one over the appropriate domain.

By (7), the predictor $\tilde{X}(s, t)$ is a point-wise kernel average of a spatio-temporal covariate surface and, hence, the models proposed here are referred to as spatio-temporal kernel averaged predictor (STKAP) models. Combining (6) and (7), notice that the kernel K_{st} determines the extent of the spatio-temporal lagged effects of X on Y at the point (s, t) . To illustrate, suppose $K_{st}(\mathbf{r} - \mathbf{s}; u - t)$ is maximized at $(\mathbf{r}, u) = (s + \boldsymbol{\theta}, t - \delta)$ then $X(s + \boldsymbol{\theta}, t - \delta)$ is given the largest weight on the right-hand-side of (6), i.e., $X(s + \boldsymbol{\theta}, t - \delta)$ has the largest effect on $\mathbb{E} Y(s, t)$. However, if $K_{st}(\mathbf{r} - \mathbf{s}; u - t) > 0$ for $(\mathbf{r}, u) = (s, t)$, $X(\mathbf{r}, u)$ still affects $Y(s, t)$ (albeit with a smaller magnitude than $X(s + \boldsymbol{\theta}, t - \delta)$) but the effect of $X(\mathbf{r}, u)$ on $Y(s, t)$ is always 0 if $u > t$ since future covariate levels should not affect current responses.

By assuming a normalized form for K_{st} , the parameter β_1 is well-identified and quantifies the global cumulative effect (in space and time) of the covariate surface on the response. The kernel K_{st} defines how the effect (β_1) is allocated over the region $\mathbb{R}^d \times (-\infty, 0)$. Additionally, the “coefficient” of any $X(s', t')$ on $Y(s, t)$ is given by $\beta_1 K_{st}(s' - s; t' - t)$ if $t' \leq t$ and 0 otherwise, enabling space-time varying coefficients due to the dependence of K on (s, t) . Extending β_1 to $\beta_1(s, t)$ using spatially varying coefficients modeling as described in Gelfand et al. [3] could be envisioned to create an arbitrarily rich specification. However, such models are expected to be poorly identified since they introduce the product form $\beta(s, t) \tilde{X}(s, t)$ where each variable comes from a process model but neither is observed.

The STKAP model in (6) defines a general class of spatio-temporal regression models that incorporates various other models proposed in the literature. For example, the point predictor model given by (1) fits into the proposed framework where $K_{st}(\mathbf{r} - \mathbf{s}; u - t) = \delta_0(\|\mathbf{r} - \mathbf{s}\| + |u - t|)$ where $\delta_0(\cdot)$ is the Dirac function (point mass) at 0. In this case, $\tilde{X}(s, t) = X(s, t)$. Additionally, the purely spatial kernel averaged predictors of Heaton and Gelfand [18] can be defined by the associated spatio-temporal kernel,

$$K_{st}(\mathbf{r} - \mathbf{s}; u - t) = K(\mathbf{r} - \mathbf{s}) \delta_0(|u - t|) \quad (8)$$

where $K(\mathbf{r} - \mathbf{s})$ is a purely spatial kernel whose shape does not vary over spatial domain of interest.

2.2. Model Properties

Combine (4) and (7) so that $\tilde{X}(s, t) = \mu_X(s, t) + \sigma_X w_X(s, t)$ where,

$$\mu_X(s, t) = \int_{\mathbb{R}^d} \int_{-\infty}^t K_{st}(\mathbf{r} - \mathbf{s}, u - t) \mu_X(\mathbf{r}, u) \nu(d\mathbf{r} du), \quad (9)$$

and

$$w_{\tilde{X}}(\mathbf{s}, t) = \int_{\mathbb{R}^d} \int_{-\infty}^t K_{st}(\mathbf{r} - \mathbf{s}, u - t) w_X(\mathbf{r}, u) v(dr du). \quad (10)$$

By (10), $\mathbb{E} w_{\tilde{X}}(\mathbf{s}, t) = 0$ for all $\mathbf{s}$ and t , and $\text{Cov}(w_{\tilde{X}}(\mathbf{s}, t), w_{\tilde{X}}(\mathbf{s}', t'))$ given by

$$\mathbb{C}_{\tilde{X}}(\mathbf{s}, \mathbf{s}'; t, t') = \int_{\mathbb{R}^d} \int_{-\infty}^t \int_{\mathbb{R}^d} \int_{-\infty}^{t'} K_{st}(\mathbf{r} - \mathbf{s}; u - t) K_{s't'}(\mathbf{v} - \mathbf{s}'; m - t') \times \rho_X(\mathbf{r}, \mathbf{v}; u, m) v(dr du) v(dv dm) \quad (11)$$

and cross-covariance function $\text{Cov}(w_{\tilde{X}}(\mathbf{s}, t), w_{\tilde{X}}(\mathbf{s}', t'))$ denoted by

$$\mathbb{C}_{X\tilde{X}}(\mathbf{s}, \mathbf{s}'; t, t') = \int_{\mathbb{R}^d} \int_{-\infty}^{t'} K_{s't'}(\mathbf{r} - \mathbf{s}'; u - t') \rho_X(\mathbf{s}, \mathbf{r}; t, u) v(dr du). \quad (12)$$

Although not explicit in the notation, $\mathbb{C}_{\tilde{X}}$ and $\mathbb{C}_{X\tilde{X}}$ are dependent on ϕ_X .

Notice from (11) and (12) that properties such as stationarity, full-symmetry, and separability of the smoothed covariate process $\tilde{X}(\mathbf{s}, t)$ are inherited from the specification of ρ_X and K_{st} . For example, $\tilde{X}(\mathbf{s}, t)$ will be a stationary process if ρ_X is stationary and K_{st} only depends on the separation vectors $\mathbf{r} - \mathbf{s}$ and $u - t$ (i.e. $K_{st}(\mathbf{r} - \mathbf{s}, u - t) = K(\mathbf{r} - \mathbf{s}, u - t)$). In contrast, if $K_{st}(\mathbf{r}, u)$ is not stationary (e.g. the shape of K_{st} varies over the spatio-temporal domain) then $\tilde{X}(\mathbf{s}, t)$ will *not* follow a stationary process even if ρ_X is a stationary covariance function. For applications that cover a bounded spatio-temporal domain, $\tilde{X}(\mathbf{s}, t)$ will *never* follow a stationary process because the normalizing constants for K_{st} will depend upon $\mathbf{s}$ and t . In similar fashion, $\tilde{X}(\mathbf{s}, t)$ will be a separable process if both ρ_X and K_{st} are separable.

Note that, $(X(\mathbf{s}, t), \tilde{X}(\mathbf{s}, t))^T$ follows a bivariate GP. Furthermore, from (6), $\mathbf{Z}(\mathbf{s}, t) = (X(\mathbf{s}, t), \tilde{X}(\mathbf{s}, t), Y(\mathbf{s}, t))^T$ follows a trivariate Gaussian process with mean,

$$\mathbb{E}(\mathbf{Z}(\mathbf{s}, t)) = (\mu_X(\mathbf{s}, t), \mu_{\tilde{X}}(\mathbf{s}, t), \beta_0 + \beta_1 \mu_{\tilde{X}}(\mathbf{s}, t))^T \quad (13)$$

and cross covariance matrix $\text{Cov}(\mathbf{Z}(\mathbf{s}, t), \mathbf{Z}(\mathbf{s}', t'))$ for $(\mathbf{s}, t) \neq (\mathbf{s}', t')$ given by,

$$\begin{pmatrix} \sigma_X^2 \rho_X(\mathbf{s}, \mathbf{s}'; t, t') & \sigma_X^2 \mathbb{C}_{X\tilde{X}}(\mathbf{s}, \mathbf{s}'; t, t') & \sigma_X^2 \beta_1 \mathbb{C}_{\tilde{X}}(\mathbf{s}, \mathbf{s}'; t, t') \\ \sigma_X^2 \mathbb{C}_{\tilde{X}}(\mathbf{s}', \mathbf{s}; t', t) & \sigma_X^2 \mathbb{C}_{\tilde{X}}(\mathbf{s}, \mathbf{s}'; t, t') & \sigma_X^2 \beta_1 \mathbb{C}_{\tilde{X}}(\mathbf{s}, \mathbf{s}'; t, t') \\ \sigma_X^2 \beta_1 \mathbb{C}_{\tilde{X}}(\mathbf{s}', \mathbf{s}; t', t) & \sigma_X^2 \beta_1 \mathbb{C}_{\tilde{X}}(\mathbf{s}, \mathbf{s}'; t, t') & \sigma_X^2 \beta_1^2 \mathbb{C}_{\tilde{X}}(\mathbf{s}, \mathbf{s}'; t, t') + \sigma_Y^2 \rho_Y(\mathbf{s}, \mathbf{s}'; t, t') \end{pmatrix} \quad (14)$$

(for derivation of these results see the Appendix). By (13) and (14), note that marginal the covariance properties of $Y(\mathbf{s}, t)$ are determined by $\mathbb{C}_{\tilde{X}}$ and ρ_Y .

As emphasized in the Introduction, the STKAP model given by (6) above is motivated by the objective of quantifying complex, non-stationary spatio-temporal lagged effects between a covariate and response. Prediction (co-kriging) under the STKAP model is not expected to differ much from that under the point predictor model (1). To see why this is the case, notice that the induced bivariate model for $(X(\mathbf{s}, t), Y(\mathbf{s}, t))$ is a GP with a covariance function which differs from that ordinarily used for spatial regression problems (i.e. $\mathbb{C}_{\tilde{X}}$ appears in the covariance function for $Y(\mathbf{s}, t)$). Thus, co-kriging under the two models will only differ in the specification of the cross-covariance function. As discussed in Zhang [20], kriging under two different covariance functions produces similar predictions (see the Appendix for more details). To emphasize, however, this focus on inference rather than prediction does not discount the applicability of the STKAP model because the goal of many spatio-temporal analyses is to capture the effect of covariates on the response (e.g. pollution and morbidity studies focus on the inferring the effect of pollutants on public health).

Using the GP model for $\mathbf{Z}(s, t)$ given in (13) and (14), various questions related to model misspecification can be directly addressed. For example, if (6) is the “true” model, what are the consequences of using the point predictor $X(s, t)$ in the conditional specification for $Y(s, t)$ rather than $\tilde{X}(s, t)$? The following two results, whose proofs are provided in the Appendix, provide the answers and are the spatio-temporal extensions of the results provided in Heaton and Gelfand [18].

Result 1. *The implied coefficient of $X(s, t)$ on $Y(s, t)$ is $\beta_1 \mathbb{C}_{X\tilde{X}}(s, s; t, t)$; i.e. it has a multiplicative bias of $\mathbb{C}_{X\tilde{X}}(s, s; t, t) \in (0, 1]$ where $\mathbb{C}_{X\tilde{X}}(s, s; t, t) = 1$ if $K_s(\mathbf{r} - s; u - t) = \delta_0(\|\mathbf{r} - s\| + |u - t|)$.*

Thus, if the point level covariate $X(s, t)$ is used rather than the kernel averaged covariate $\tilde{X}(s, t)$, the estimated effect of the covariate on the response will be shrunk towards zero. This result can have serious implications in epidemiological and other studies where $X(s, t)$ is often used rather than $\tilde{X}(s, t)$. For example, if pollution were used to explain health outcomes, the estimated pollution effect could be markedly masked.

Second, the percent of variation in $Y(s, t)$ explained by $\tilde{X}(s, t)$ is greater than or equal to the percent of variation in $Y(s, t)$ explained by $X(s, t)$ as detailed in the following result.

Result 2. *Let $R_X^2(s, t)$ and $R_{\tilde{X}}^2(s, t)$ be the proportion of variation in $Y(s, t)$ explained by $\tilde{X}(s, t)$ and $X(s, t)$, respectively. Then, under the assumptions of (6), $R_X^2(s, t) \geq R_{\tilde{X}}^2(s, t)$ for all (s, t) with equality if $K_s(\mathbf{r} - s; u - t) = \delta_0(\|\mathbf{r} - s\| + |u - t|)$.*

If $X(s, t)$ is incorrectly used as the predictor of $Y(s, t)$ then not only is the effect of the covariate on the response being underestimated, but less variance is being explained. Thus, for situations where misspecification between $\tilde{X}(s, t)$ and $X(s, t)$ may be a concern, a parametric kernel that can tend toward the “Dirac” case should be used.

3. Implementation Details

In the most general setting, responses and predictors do not have to be measured at the same spatio-temporal locations and observed spatial locations can vary in number and location across time periods. Let, $Y(s_{Y,i}, t_{Y,i})$ be the observed response at spatial location $s_{Y,i}$ and time period $t_{Y,i}$ for $i = 1, \dots, N_Y$ and $\mathbf{Y} = (Y(s_{Y,1}, t_{Y,1}), \dots, Y(s_{Y,N_Y}, t_{Y,N_Y}))^T$ denote the vector of observed responses of length N_Y . In similar notation, let $\mathbf{X} = (X(s_{X,1}, t_{X,1}), \dots, X(s_{X,N_X}, t_{X,N_X}))^T$ and $\tilde{\mathbf{X}} = (\tilde{X}(s_{Y,1}, t_{Y,1}), \dots, \tilde{X}(s_{Y,N_Y}, t_{Y,N_Y}))^T$ be the observed covariates and the (latent) kernel averaged covariates. First, notice that if the data are misaligned (i.e. the responses and covariates are measured at distinct spatio-temporal locations), the methods here still hold because a valid joint distribution was defined for $\mathbf{Z}(s, t) = (X(s, t), \tilde{X}(s, t), Y(s, t))^T$. Second, notice the latent kernel averaged predictors $\tilde{\mathbf{X}}$ are aligned with the observed response variables. Finally, because the joint distribution of $(\mathbf{X}, \tilde{\mathbf{X}}, \mathbf{Y})^T$ is Gaussian, $\tilde{\mathbf{X}}$ can be easily integrated out and parameter estimation can be done using the joint distribution $[\mathbf{X}, \mathbf{Y}]$ where $[\cdot]$ denotes a density function (in this case, the Gaussian density function). While this results in a non-conjugate update for β_1 , the mixing of the Markov chain used to obtain draws from the posterior distribution is significantly improved.

Heaton and Gelfand [18] assume $\nu(\cdot)$ to be Lebesgue measure. However, when $\nu(\cdot)$ is Lebesgue measure, the elements of the covariance matrix for $(\mathbf{X}, \mathbf{Y})^T$ involve integrals which are not available in closed form. By expressing (11) and (12) as expectations, Monte Carlo (MC) integration can be used to calculate the integrals. However, MC integration presents several challenges. First, MC integration can give poor results, especially if the spatio-temporal range of $X(s, t)$ is small. Second, MC integration is computationally expensive

because $N_Y(N_Y + 1)/2 + N_Y N_X$ such integrations will need to be performed at each iteration. For problems involving a large sample size, this can be a computationally daunting task. And, third, the Monte Carlo integration would need to be carefully constructed so as to guarantee that the estimated covariance matrix will be positive definite.

As opposed to Heaton and Gelfand [18], by assuming $\nu(\cdot)$ to be discrete the computational burden associated with fitting STKAP models to data is partially (but not completely) alleviated. To see this, notice that if $\nu(\cdot)$ is discrete, then

$$\tilde{X}(s, t) = \mathbf{k}^T(s, t) \mathbf{X}^*, \quad (15)$$

where $\mathbf{X}^* = (X(\mathbf{r}_1, u_1), \dots, X(\mathbf{r}_{M^*}, u_{M^*}))^T$ are a vector of covariates at a set of locations (“knots”) and $\mathbf{k}^T(s, t) = (K(\mathbf{r}_1 - s, u_1 - t | \boldsymbol{\xi}(s, t)), \dots, K(\mathbf{r}_{M^*} - s, u_{M^*} - t | \boldsymbol{\xi}(s, t)))$ is a vector of weights (summing to one) and $K(\mathbf{r} - s, u - t | \boldsymbol{\xi}(s, t)) = 0$ if $u > t$. From (15), $E Y(s, t) = \beta_0 + \beta_1 \mathbf{k}^T(s, t) \boldsymbol{\mu}_{X^*}$ where $\boldsymbol{\mu}_{X^*} = (\mu_X(\mathbf{r}_1, u_1), \dots, \mu_X(\mathbf{r}_{M^*}, u_{M^*}))^T$, and (11) and (12) become

$$\mathbb{C}_{\tilde{X}}(s, s'; t, t') = \mathbf{k}^T(s, t) \boldsymbol{\Sigma}_{X^*} \mathbf{k}(s', t') \quad (16)$$

$$\mathbb{C}_{X\tilde{X}}(s, s'; t, t') = \sigma_{X^*}^T(s, t) \mathbf{k}(s', t') \quad (17)$$

where $\boldsymbol{\Sigma}_{X^*} = \{\rho_X(\mathbf{r}_i, \mathbf{r}_j; u_i, u_j)\}_{ij}$ and $\sigma_{X^*}^T(s, t) = (\rho_X(s, \mathbf{r}_1; t, u_1), \dots, \rho_X(s, \mathbf{r}_{M^*}; t, u_{M^*}))$. To emphasize, assuming ν is a discrete measure does not completely alleviate the computational difficulty of fitting the STKAP model. Specifically, if M^* is large then the matrix multiplication in (15), (16), and (17) and normalizing $\mathbf{k}^T(s, t)$ by dividing by its sum can be taxing.

The set of knots $\{\mathbf{r}_m, u_m\}$ need not cover the entire region of interest but can be strategically chosen to be dense near observed $Y(s, t)$. A convenient choice is to let $M^* = N_Y M$ where M is the number of knots that lie within a certain distance of each $(s_{Y,i}, t_{Y,i})$. Note, also, that $\{\mathbf{r}_m, u_m\}$ need not be $\{s_{X,i}, t_{X,i}\}$ but if $\{s_{X,i}, t_{X,i}\} \subset \{\mathbf{r}_m, u_m\}$ then this can further decrease the computational burden of fitting the STKAP model.

If ν is discrete as above, then the likelihood of $(\mathbf{X}, \mathbf{Y})^T$ is Gaussian with mean $(\mu_X^T, \beta_0 + \beta_1(\mathbf{K}\boldsymbol{\mu}_{X^*})^T)^T$ and variance matrix

$$\mathbf{V} = \begin{pmatrix} \sigma_X^2 \boldsymbol{\Sigma}_X & \sigma_X^2 \beta_1 \boldsymbol{\Sigma}_{X\tilde{X}} \mathbf{K}^T \\ \sigma_X^2 \beta_1 \mathbf{K} \boldsymbol{\Sigma}_{X\tilde{X}}^T & \sigma_X^2 \beta_1^2 \mathbf{K} \boldsymbol{\Sigma}_{X^*} \mathbf{K}^T + \sigma_Y^2 \boldsymbol{\Sigma}_Y + \tau_Y^2 \mathbf{I}_{N_Y} \end{pmatrix},$$

where μ_X^T is the $N_X \times 1$ vector of means associated with $\mathbf{X}$, $\mathbf{K}$ is an $N_Y \times M^*$ matrix with rows $\mathbf{k}^T(s_{Y,i}, t_{Y,i})$, $\boldsymbol{\Sigma}_{X\tilde{X}}$ is an $N_X \times M^*$ matrix with rows $\sigma_{X^*}^T(s_{X,i}, t_{X,i})$, $\boldsymbol{\Sigma}_Y$ is $N_Y \times N_Y$ with ij^{th} element $\rho_Y(s_{Y,i}, s_{Y,j}; t_{Y,i}, t_{Y,j})$ and $\mathbf{I}_{N_Y}$ is the rank N_Y identity matrix.

4. Specifying The Smoothing Kernel

4.1. General Guidelines

Only parametric kernels are considered here, in particular, kernels which have a known functional form given a vector of parameters $\boldsymbol{\xi}(s, t)$ which are allowed to vary over space and time (as opposed to constant as in Heaton and Gelfand 18). Thus, from here on, $K_s(\mathbf{r} - s; u - t)$ is written as $K(\mathbf{r} - s, u - t | \boldsymbol{\xi}(s, t))$. The vector $\boldsymbol{\xi}(s, t)$ will be unknown in practice and, in order to properly account for uncertainty, will be assigned a prior distribution with

inference conditioned on the data. As is explicit in this notation, $\xi(s, t)$ is allowed to vary over the spatio-temporal domain of interest. In this way, the shape of K may also vary over this domain leading to heterogeneous spatio-temporal lagged effects.

Smoothing kernels can be classified as stationary, symmetric, or separable and can exhibit any combination of these properties simultaneously. The following definitions are adopted with the class of kernels used here. A kernel K is said to be *stationary* if $K(\mathbf{r} - \mathbf{s}; u - t | \xi(s, t)) = K(\mathbf{r} - \mathbf{s}; u - t | \xi)$. That is, the shape of the kernel does not change in space and time. Likewise, a kernel can be spatially stationary yet temporally non-stationary, temporally stationary yet spatially non-stationary, or non-stationary. A kernel is said to be (spatially) *symmetric* if $K(\mathbf{r} - \mathbf{s}; u - t | \xi(s, t)) = K(\mathbf{s} - \mathbf{r}; u - t | \xi(s, t))$. Because temporal smoothing only occurs retrospectively, the notion of symmetry in the temporal argument does not exist. Finally, *separable* kernels assume independent smoothing in space and time such that $K(\mathbf{r} - \mathbf{s}; u - t | \xi(s, t)) = K_1(\mathbf{r} - \mathbf{s} | \xi_1(s))K_2(u - t | \xi_2(t))$. Despite their simple and interpretable form, separable kernels offer little computational benefit unless coupled with a separable correlation function. Specifically, if $\rho_X(s, s'; t, t') = \rho_{X1}(s, s')\rho_{X2}(t, t')$ and K is separable then the variance-covariance structure of $(X(s, t), \tilde{X}(s, t))$ given by (12) factors into a purely spatial and purely temporal component allowing computationally efficient inference.

When specifying the smoothing kernel, a convenient and intuitive factorization is split K into spatial and temporal components such that

$$K(\mathbf{r} - \mathbf{s}; u - t | \xi(s, t)) = K_1(\mathbf{r} - \mathbf{s} | u, \xi_1(s, t))K_2(u - t | \xi_2(s, t)) \quad (18)$$

where $K_1(\mathbf{r} - \mathbf{s} | u, \xi_1(s, t))$ is the spatial weight of location $\mathbf{r}$ at time u , $K_2(u - t | \xi_2(s, t))$ is the temporal weight of time u and $\int_{\mathbb{R}^d} K_1(\mathbf{r} - \mathbf{s} | u, \xi_1(s, t)) \nu(d\mathbf{r}) = \int_{-\infty}^0 K_2(u | \xi_2(s, t)) \nu(du) = 1$. In this way, the kernel can be specified by considering temporal and spatial weighting separately. The conditional spatial smoothing kernel $K_1(\mathbf{r} - \mathbf{s} | u, \xi_1(s, t))$ is a density function on $\mathbb{R}^d$ and is characterized by translation, dilation, and orientation parameters. For example, if $K_1(\mathbf{r} - \mathbf{s} | u, \xi_1(s, t))$ is proportional to the d -dimensional Gaussian density function with positive definite covariance matrix $\Xi(s, t)$ and mean $\boldsymbol{\mu}(s, t) = (\mu_1(s, t), \mu_2(s, t))$ then $\boldsymbol{\mu}(s, t)$ is a translation parameter in that it shifts the center of the spatial kernel away from s suggesting that the effect of $X(s + \boldsymbol{\mu}(s, t), u)$ has the largest effect on $Y(s, t)$ such that the spatial direction of influence of $X(\mathbf{r}, u)$ on $Y(s, t)$ will be in the $-\boldsymbol{\mu}(s, t)$ direction. Admittedly, in most applications, $\boldsymbol{\mu}(s, t) = \mathbf{0}$. Dilation parameters control the dilation of the influence of $X(\mathbf{r}, u)$ on $Y(s, t)$ while orientation parameters rotate the principle axes of this dilation. In this example, the diagonal elements of $\Xi(s, t)$ are dilation parameters while the off-diagonal elements are orientation parameters.

As a precaution in specifying the spatial kernel $K_1(\mathbf{r} - \mathbf{s} | u, \xi_1(s, t))$, note that orientation parameters will *not* be identified if the correlation function ρ_X is spatially isotropic (i.e., $\rho_X(s, s'; t, t') = \rho_X(\|s' - s\|; t, t')$). To see this, simply note that if ρ_X is spatially isotropic then, under model (6), so are the covariance and cross-covariance functions $C_{\tilde{X}}$ and $C_{X\tilde{X}}$. Because of this, the value of the likelihood will be equivalent under any rotation of the spatial axes. Under spatial isotropy, orientation parameters should either not be used or should be fixed at known values *a priori*.

The purely temporal kernel $K_2(u - t | \xi_2(s, t))$ is a density function on $(-\infty, t]$ and is characterized by translation and dilation parameters. A translation in $K_2(u - t | \xi_2(s, t))$ away from zero allows historical covariates to have a larger coefficient in $\mathbb{E} Y(s, t)$ than more recent covariates. Temporal translations are appropriate for many environmental and atmospheric problems; for example, measuring the effect of pollution levels on an adverse health event may include a translation parameter because of the lag from time of exposure to

time of death. Dilations in $K_2(u - t | \xi_2(s, t))$ increase the length of time that X has an effect on Y ; i.e. highly dilated kernels have longer temporal covariate lags.

Ultimately, the choice of kernel should hinge on the application. Additionally, the availability of data may highly influence the choice of kernel. For example, if $\xi(s, t)$ is independent (or is even modeled as a transformation of a spatio-temporal GP) across all (s, t) then $\forall t(Y(s, t))$ is different for each (s, t) and repeated measures would be required to estimate the parameters [see 21, Chapter 3]. In spatio-temporal problems, repeated measures at (s, t) are seldom available. Thus, in order to estimate it from the data, $\xi(s, t)$ would need to follow a functional form (e.g. a spline) or simply allowed to vary over either space or time (but not both).

4.2. A Spatially Non-Stationary Kernel for Ozone Concentrations

As proper specification and parameterization of K is a crucial piece of a STKAP model, an example is given here of the kernel used throughout the applications of Section 5. Particularly, the specification of K in this subsection is motivated by the problem of regressing ozone concentration on temperature. Thus, for this subsection, let $Y(s, t)$ denote ozone concentration and $X(s, t)$ denote temperature.

Exploratory analysis of temperatures (see Section 5.2 below for more details) suggest a fully-symmetric yet non-separable form for the covariance function. Let $\rho_X(\cdot)$ be given by,

$$\rho_X(s, s'; t, t') = (\psi_X |t' - t| + 1)^{-(\lambda_X + \eta_X)} \exp \left\{ -\phi_X \frac{\|s' - s\|}{(\psi_X |t' - t| + 1)^{\eta_X/2}} \right\}, \quad (19)$$

where $\psi_X, \lambda_X > 0$ are temporal decay parameters, $\phi_X > 0$ is the spatial decay parameter and $\eta_X \in [0, 1]$ controls the separability of spatial and temporal dimensions [19]. The purely temporal range is the solution to $\rho_X(s, s'; t, t') = 0.05$ in terms of $|t' - t|$; carrying out the algebra yields a temporal range given by,

$$r_2 = \frac{20^{1/(\lambda_X + \eta_X)} - 1}{\psi_X}. \quad (20)$$

Given a temporal separation of $u = |t' - t|$, the spatial range ($r_1(u)$) is the solution to $\rho_X(s, s'; t, t + u) = 0.05$ in terms of $\|s' - s\|$. For (19),

$$r_1(u) = - \frac{\log[0.05(\psi_X u + 1)^{(\lambda_X + \eta_X)}](\psi_X u + 1)^{\eta_X/2}}{\phi_X}. \quad (21)$$

As a special case, the purely spatial range, i.e. $t' = t$, is $r_2(0) \approx 3/\phi_X$ which is the spatial range of the exponential covariance function [see 22, Chapter 2]. We note the recent work of Kent et al. [23] who pointed out a “dimple” property associated with (19). The dimple property was unknown to us at the time this work was done.

To specify K , factor K as in (18) such that the spatial (K_1) and temporal (K_2) kernels can be considered separately. Because ozone is formed by a reaction between nitrogen oxides (NO_x) and volatile organic compounds (VOC) in the presence of sunlight, temperatures (a proxy for sunlight) in close spatio-temporal proximity to (s, t) are likely the most highly correlated with $Y(s, t)$ suggesting that K_1 should be centered at s and K_2 should be centered at 0 (i.e. K_2 reaches a maximum at the current time period t). Ozone concentrations are also highly affected by weather conditions suggesting that spatio-temporal lagged effects of temperature should vary over the spatial region and be constrained to periods of stable

temperatures. From this intuition, consider the temporally stationary (yet spatially non-stationary) kernel given by

$$K_2(u - t | \xi_2(s)) = \frac{1}{\xi_2(s)r_2/3} \exp \left\{ \frac{u - t}{\xi_2(s)r_2/3} \right\} u \leq t \quad (22)$$

$$K_1(\mathbf{r} - \mathbf{s} | u, \xi_1(s)) = \frac{1}{2\pi(\xi_1(s)r_1(t - u)/2)^2} \exp \left\{ -\frac{\|\mathbf{r} - \mathbf{s}\|^2}{2(\xi_1(s)r_1(t - u)/2)^2} \right\} \mathbf{r} \in \mathbb{R}^2 \quad (23)$$

where $\xi_1(s), \xi_2(s) \in (0, 1)$ for all s . The spatial kernel in (23) is Gaussian with center s and common standard deviation $\xi_1(s)r_1(t - u)/2$. Thus, under (23), covariates closer to s have a larger effect on $\mathbb{E} Y(s, t)$ than covariates farther away. The parameter $\xi_1(s)$ governs the scale of the spatial kernel with values of $\xi_1(s)$ close to 1 indicating broader spatial lags. The temporal kernel in (22) arises by assuming that $u^* = t - u$ follows an exponential distribution and allows for the effect of temporally lagged covariates to decrease as u decreases where the rate of decay is governed by the parameter $\xi_2(s)$. In other words, as $\xi_2(s)$ increases, the length of the temporal lag increases.

Notice that K defined by (22) and (23) above is not separable because (22) depends on the temporal piece through $r_1(t - u)$ and spatially non-stationary because $\xi_1(s)$ and $\xi_2(s)$ vary over the spatial domain (see Section 5 for a discussion on the spatial prior distributions used for $\xi_1(s)$ and $\xi_2(s)$). However, K is symmetric because (23) depends only on $\|\mathbf{r} - \mathbf{s}\|$. The K above also incorporates the point predictor model in (1) by allowing $\xi_1(s), \xi_2(s) \rightarrow 0$ and the kernel averaged predictor model of Heaton and Gelfand [18] if $\xi_2(s) = 0$.

5. Application

5.1. Ozone Concentration Data Set

Due to its important implications for public health, spatial modeling of ozone concentrations has been given considerable attention in the literature. Citing only a few pertinent examples, Zhang and Fan [24] and Reich et al. [8] model maximum 8-hour average ozone concentrations without a normality assumption but constrain the relationship between meteorological variables and ozone concentration to a single spatio-temporal coordinate. In contrast, Sahu et al. [25] propose a space-time model which specifies ozone concentration as a function of the yearly change in meteorological conditions but their model doesn't account for short-term and spatial fluctuations in meteorology (the interested reader is directed to the references contained in the above cited articles for further discussion of statistical approaches to ozone modeling). Here, a STKAP model is used to explore the relationship between average daily temperature and daily maximum 8-hour average ozone measurements.

The data set under consideration is comprised of daily maximum 8-hour average ozone concentrations (in parts per billion; ppb) and daily average temperature (in degrees Fahrenheit; °F) measured at various stations across the state of California between July 1, 2010 to July 14, 2010 (the stations are given by the grid squares in Figure 1). The number of ozone monitoring stations varied from day to day from a minimum of 94 to a maximum of 97 ozone stations (98 unique stations) totaling 1346 ozone measurements. In contrast, average daily temperature was available at 238 stations (including the 98 ozone monitoring stations) totaling 3332 temperature measurements. Figure 1 displays the empirical average ozone and temperature level over the temporal domain of interest. Figure 2 (a) displays a scatter plot of the ozone measurements and the aligned temperature measurement and Figure 2 (b) displays the correlation between ozone concentration and average temperature by

spatial location. While Figure 2 (a) displays a clear positive linear relationship between ozone and daily temperature, Figure 2 (b) shows that this relationship changes over the spatial domain. The spatially varying correlation in Figure 2 (b) suggests extending (6) to a spatially varying coefficients model (i.e. $\beta(s, t)$). As stated previously, such models may be difficult to identify since they introduce the product form $\beta(s, t)\tilde{X}(s, t)$ where neither is observed. For this reason, extensions to spatially-varying coefficient models is left for future research and the purpose here is to determine if, on a global scale, the relationship between temperature and ozone concentration is better explained by looking beyond a single spatio-temporal coordinate.

5.2. Exploratory Analysis and Model Specification

Let $Y(s, t)$ represent the ozone concentration at site s and day t and $X(s, t)$ represent the average temperature on day t at site s . Let the model for $X(s, t)$ be given by (4) where $\mu_X(s, t) = \mu_X$ for all s and t . The form of the covariance function $\rho_X(\cdot)$ is important in that $\rho_X(\cdot)$ governs the spatio-temporal correlation of average temperatures and, marginally, the spatio-temporal correlation of ozone concentrations. Hence, assessing the validity of assumptions such as stationarity, full-symmetry, and separability of the covariance function is necessary. Prior to analyzing the ozone data set described above, an exploratory analysis of a training data set consisting of average temperatures for the same region as that shown in Figure 1 for the last two weeks of June 2010 was used to assess assumptions of full-symmetry and separability. For simplicity, only stationary covariance functions were considered. Both graphical methods, such as those suggested by Gneiting et al. [2], and the tests described in Li et al. [26] suggest a fully-symmetric yet, possibly, non-separable covariance function for temperature (the exploratory results were less conclusive for separability than for full-symmetry). These findings are not too surprising given that the temporal domain of interest is only the first two weeks of July when weather and atmospheric conditions are relatively stable. Thus, let $\rho_X(\cdot)$ be defined as in (19).

Let the model for $Y(s, t)$ be given by (6) where $\tilde{X}(s, t)$ is the kernel averaged predictor given by (15) and $\sigma_Y^2=0$ such that all spatio-temporal correlation in $Y(s, t)$ is explained by $\tilde{X}(s, t)$. A total of 10,062 knot locations were used for the vector $\mathbf{X}^*$. The first 3332 of these knot locations were those locations where $X(s, t)$ was observed directly. The remaining $10,062 - 3332 = 6,730 = 1346 \times 5$ locations were chosen by placing 5 additional knots near the 1346 locations where $Y(s, t)$ was observed.

Numerous studies have observed changes in ozone cycles, temporal trends, and formation due to changes in elevation [see, e.g. 27, 28, 29, and the references therein]. Given the diffuse geography of California from beaches of San Diego to rugged mountains of Lake Tahoe, the lagged effect of temperature on ozone will certainly vary over the state. Let the kernel K be that given in (22) and (23). The models for $\xi_1(s)$ and $\xi_2(s)$ are given by,

$$\xi_i(s) = \Phi(\gamma_{i0} + \gamma_{i1}E(s)) \quad (24)$$

where $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function, $E(s)$ is the elevation of site s and $i = 1, 2$. In this way, the parameter γ_{i0} represents the extent of the spatial ($i = 1$) and temporal ($i = 2$) lagged effects for a station at sea level. That is, as γ_{10} (γ_{20}) increases, the breadth of the spatial (temporal) lag also increases. The parameter γ_{i1} controls the change in lagged effects with elevation. For example, if γ_{11} (γ_{21}) is negative then as elevation increases, the breadth of the spatial (temporal) lag decreases. Of course, if $\gamma_{i1} = 0$ then the extent of the spatio-temporal lags is constant over the spatial region.

By (24), notice that a degree of non-identifiability arises in the parameters γ_{i0} and γ_{i1} which determine $\xi_i(s)$. For example, if $\gamma_{i0} \ll 0$ then the data will contain no information about the

value of γ_{i1} if $\gamma_{i1} < 0$ (because the value of $\xi_A(s)$ would change very little over space) but the data could inform about the value of γ_{i1} if $\gamma_{i1} > 0$ (because the value of $\xi_A(s)$ would vary over space). Thus, to avoid the non-identifiability, $\{\gamma_{i0}, \gamma_{i1}\}_{i=1}^2$ are constrained such that $\Phi(\gamma_{i0} + \gamma_{i1}E(s)) \in (0.005, 0.995)$ for all $E(s)$.

In order to complete the model specification, prior distributions are required for each unknown parameter. Based on experience associated with the simulation studies that follow, informative prior distributions should be used for parameters associated with ρ_X and K . To see why this is the case, notice that, after integrating out $\tilde{X}$, $\xi(s)$ and ϕ_X become parameters associated with the covariance function of $Y(s, t)$. As discussed in Zhang [20], parameters associated with covariance functions are not necessarily consistently estimable (but some functions of them are). To deal with this, the approach here is to use discrete prior distributions for these parameters. Each of the prior distributions below were motivated by the results of the exploratory analysis of the last two weeks of June 2010 described above. The details are as follows.

The value of λ_X was fixed at 10 and discrete uniform prior distributions were assumed for each of $[\psi_X | \eta_X]$, $[\eta_X]$ and $[\phi_X]$. Specifically, equal mass was assumed for each of $\eta_X \in \{0, 0.3, 0.5, 0.8, 1\}$. Discrete uniform prior distributions for ϕ_X and ψ_X were induced by placing equal mass on the purely spatial and temporal ranges of $r_1(0) \in \{0.1 \times D_X, 0.3 \times D_X, 0.5 \times D_X, 0.7 \times D_X, 0.9 \times D_X\}$ and $r_2 \in \{1, 2, 4, 7, 10\}$, respectively, where D_X is the maximum observed distance between temperature stations.

Following the discussion above, γ_{i0} $i = 1, 2$ were assumed, *a priori*, to follow a discrete uniform prior with masses at $\Phi^{-1}(0.01)$, $\Phi^{-1}(0.1)$, $\Phi^{-1}(0.3)$, $\Phi^{-1}(0.5)$, $\Phi^{-1}(0.7)$, and $\Phi^{-1}(0.9)$. Finally, the prior distribution for γ_{i1} was assumed to be $\gamma_{i1} | \gamma_{i0} \sim \mathcal{U}((\Phi^{-1}(0.005) - \gamma_{i0})/E_{\max}, (\Phi^{-1}(0.995) - \gamma_{i0})/E_{\max})$ where $E_{\max} = \max_s E(s)$ which gives the appropriate constraints. Finally, vague (yet proper) Gaussian prior distributions were used for β_0 and β_1 ; likewise, vague (but proper) inverse-gamma prior distributions were assumed for τ_y^2 and σ_x^2 .

5.3. Simulations

In order to verify the ability to retrieve the model parameters, the STKAP model was first fit to two groups of 100 simulated data sets. The first group of 100 data sets were simulated from the STKAP model described in Section 5.2 assuming $\mu_X = 0$, $\beta_0 = 0$, $\beta_1 = 1$, $\sigma_x^2 = 1$, $\tau_y^2 = 0.5$, $\eta_X = 0.5$, $r_2 = 4$, $r_1(0) = 0.5 \times D_X$, $\gamma_{10} = \gamma_{20} = \Phi^{-1}(0.9)$ and $\gamma_{11} = \gamma_{21} = -0.0058$. The value of τ_y^2 was chosen such that $\tilde{X}(s, t)$ explained approximately 50% of the variance in $Y(s, t)$. The second group of 100 data sets were simulated from the point predictor (PtP) model $Y(s, t) = \beta_0 + \beta_1 X(s, t)$ using the same parameter values as the first group (note that under the PtP model $\xi_1(s) = \xi_2(s) = 0$ for all s). For both groups of simulated data sets, the spatio-temporal locations were assumed to be the same as those observed in the ozone data set displayed in Figure 1 and described in Section 5.1. For comparison, the PtP model was also fit to each group of data sets.

Table 1 compares the fit of the STKAP and PtP models across the first group of simulated data sets (i.e. those simulated under the STKAP model assumptions) in terms of the empirical bias, root mean square error (RMSE), and coverage probabilities of a 95% central credible interval for the continuous parameters. In general, the STKAP model correctly estimates the parameters. The PtP model, however, had a large negative bias for β_1 . This result was argued in Section 2.2. Furthermore, the PtP model had a positive bias for τ_y^2 suggesting that the kernel averaged predictor $\tilde{X}(s, t)$ explains a larger percent of the variance in $Y(s, t)$ than the point predictor $X(s, t)$.

Table 2 compares the fit of the STKAP and PtP models to the second group of simulated data sets (i.e. those simulated under the PtP model assumptions) in terms of bias, RMSE, and 95% credible interval coverage probabilities. For this group of data sets, the values in Table 2 suggest that both models perform comparably. This is because the STKAP model correctly estimated small spatio-temporal lagged effects (i.e. $\xi_1(s) = \xi_2(s) \approx 0$ for all s). Specifically, the STKAP model correctly placed all posterior mass at $\gamma_{10} = \gamma_{20} = \Phi^{-1}(0.01)$ and $\gamma_{11} = \gamma_{21} \approx 0$ suggesting that spatio-temporal lagged effects were negligible. This performance is encouraging in that the data can correctly inform the STKAP model as to whether or not spatio-temporal lagged effects are present.

5.4. Ozone Production Data Set Analysis

To analyze the California ozone data set, both the STKAP and PtP models were fit and the deviance information criterion (DIC) was calculated for each model [see 30]. Here, DIC is an appropriate model comparison diagnostic because the PtP model can be considered to be a “reduced” model from the “full” STKAP model where the difference in model dimension is given by the dimension of ξ (the number of parameters of the kernel). The DIC values were 14027 and 20188 with 8.99 and 3.88 effective parameters (p_D) for the STKAP and PtP models, respectively. This suggests that the STKAP model produces a superior fit with only a small increase in the number of effective parameters.

Figure 3 displays several diagnostic plots for the assumptions used to fit the STKAP model. Figure 3 (a) displays the residuals vs. fitted values plot where residuals are defined as

$Y(s, t) = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}(s, t)$ and the “hat” notation denotes the posterior mean. By Figure 3 (a), the variability of the residuals increases with ozone concentration. This makes physical sense in that high levels of ozone are attributed to many other sources beyond temperature alone such as pollution emissions, traffic density, etc. Figure 3 (b) displays the QQ-plot of the residuals and points to violations in the normality assumption used here. Specifically, ozone concentrations, after controlling for temperature, have a heavier right tail than is captured by the Gaussian distribution.

One assumption used here is that the spatio-temporal variability of ozone is completely explained by average daily temperature (i.e. $\sigma_Y^2 = 0$). However, the residuals showed extra spatial variability not accounted for by temperature. Specifically, an independent likelihood ratio test was performed on the residuals for each time period where the “full” model included spatial correlation and the “reduced” model did not. For all time periods, the null hypothesis of “no spatial correlation” was rejected at the 0.05 level. Figure 3 (c) displays the histogram of lag-1 autocorrelations of the residuals at each ozone monitoring station. The empirical autocorrelations are generally small suggesting that temperature did explain most of the temporal correlation of ozone. The methodology outlined in the previous sections discusses how to account for the extra spatial variability in $Y(s, t)$ beyond that explained by $X(s, t)$. However, the primary goal of STKAP models is to quantify the explanatory relationship between X and Y . In this light, including such extra spatial correlation is peripheral to the primary objective of the analysis.

Considering the covariance function ρ_X , the marginal posterior distributions of the spatial range $r_1(0)$, temporal range r_2 , and η_X were similar for both the STKAP and PtP models. The fact that both the STKAP model and the PtP model estimated a similar covariance function is not surprising because the model for $X(s, t)$ is the same in both cases. In both cases, $[\eta_X = 0 \mid Y, X] = 1$ such that a separable space-time covariance function was used for average daily temperature. As mentioned in the exploratory analysis above, because the temporal domain of interest is only two weeks during the middle of the summer, atmospheric conditions are relatively stable such that a separable space-time covariance

function is appropriate. For longer temporal domains, however, a non-separable covariance function may be needed. As the marginal posterior distributions for $r_1(0)$ and r_2 were similar for both the STKAP and PtP models, Figure 4 displays the marginal posterior distributions for $r_1(0)$ and r_2 from the STKAP model. From Figure 4, the data is clearly informing about $r_1(0)$ and r_2 as their respective posterior distributions do not match the uniform prior distribution. The posterior means for the purely spatial and purely temporal ranges were approximately 550 miles and 5 days for both the STKAP and PtP models.

Table 3 compares posterior quantiles of β_0 , β_1 , τ_r^2 , and σ_x^2 across both models. First, notice that the estimate of β_1 in the STKAP model is larger than that in the PtP model. This is in accord with the results in Section 2.2 and also realized in the simulation study in Section 5.3. Also notice from Table 3 that the posterior quantiles for σ_x^2 are essentially identical across the two models because in both the STKAP and PtP models, the marginal distribution of $X(s, t)$ is the same.

Figure 5 displays the posterior means of $\xi_1(s)r_1(0)/2$ (the standard deviation of the purely spatial kernel in miles) and $\xi_2(s)r_2/3$ (the mean of the exponential temporal kernel in days). As is displayed in Figure 5 the width of the spatial kernel decreases with elevation while the temporal kernel is approximately constant with elevation (95% central credible intervals for γ_{11} and γ_{21} are $(-0.003, -0.001)$ and $(-0.009, 0.005)$, respectively). At sea level, high temperatures within a distance of approximately 80 miles (2 standard deviations) will increase the concentration of ozone (see Figure 5 (a)). Likewise, if temperatures have been high over the past 1 or 2 days, ozone concentrations are also expected to be high.

As a final point in this analysis, Figure 6 displays the $R^2(s)$ statistic (i.e. the proportion of variation in $Y(s, t)$ explained by $\tilde{X}(s, t)$) at the observed locations. Under the STKAP model, the $R^2(s)$ statistic is spatially varying but not temporally varying because the kernel K only varies in space. According to Figure 6, for areas of higher elevation and much of Southern California, spatially and temporally lagged daily temperature explains more of the variability in ozone as compared to temperature at a given station and day.

6. Summary

This article extended the work of Heaton and Gelfand [18] to the spatio-temporal setting by developing statistical models which account for spatial and temporal lagged effects between a covariate and a response. Specifically, the stationary kernels of Heaton and Gelfand [18] were extended to be parametric yet non-stationary to allow the relationship between the response and covariates to vary over the region. Importantly, the STKAP model presented here contains the traditional point predictor models as well as the models of Heaton and Gelfand [18] as special cases and simulation studies effectively showed that the data can indicate the presence or absence of spatio-temporal lagged effects (provided that the point predictor model is not ruled out *a priori*).

In fitting the STKAP model to a data set of ozone concentrations and daily temperatures in California, new scientific insight emerged in that the relationship between temperature and ozone concentration may not be confined to a single spatial location and time period (as is typically assumed in statistical models). However, for sake of clarity in illustration, several scientific aspects of tropospheric ozone were not included in the analysis. For example, as one reviewer pointed out, sunlight (solar radiation) plays the critical role of ozone formation and temperature is merely a surrogate measure. Thus, to emphasize, care should be used when interpreting the slope parameter (β_1) in the ozone application as β_1 relates to the effect of average daily temperature and *not* solar radiation. Indeed, the complex spatio-temporal

relationship between average daily temperature and ozone concentration discovered in Section 5.4 may not be equivalent to the relationship between solar radiation and ozone.

Ozone formation and destruction is also tied to rural versus urban landscapes. While not included here, a rural versus urban landscape indicator variable could be included as either a fixed (i.e. not lagged over space and time) covariate or included as an additional covariate determining the shape of the kernel itself. This raises a particularly critical issue regarding STKAP models in that the shape of the kernel can, and should, be tied to scientific understanding of the variables being modeled. Furthermore, due to the interpretability of the kernel, eliciting subject expert opinions is possible.

As a final point regarding ozone modeling, daily maximum or average daytime temperature may be a preferred covariate to average daily temperature because nighttime temperatures are not indicative of sunlight. Original stages of this work explored the use of daily maximum temperature as average daytime temperature was not available for the dataset under consideration. However, due to the highly right skewed nature of daily maximum temperatures, the Gaussian assumption of Equation (4) was inappropriate and did not accurately capture the covariate surface. Future extensions of STKAP models will consider non-Gaussian data.

The STKAP models considered here are “simple linear” STKAP models in that they involved only a univariate response and a single predictor in the Gaussian setting. Extensions to multivariate responses and multiple predictors present several challenges such as the use of different kernels for different predictors with parameters that are, potentially, correlated. Furthermore, when dealing with multivariate data (or space-time data in general), the size of the data sets increases very quickly. Methods for dealing with large spatio-temporal data sets have recently been given considerable attention [see, e.g., 31, 32, 33, for a review] and such methods would need to be adapted for STKAP models.

Acknowledgments

This work was partially supported by NSF DMS 0934595.

Appendix A. Derivation of Distributional Results

Let $\mathbf{Z}(s, t) = (X(s, t), \tilde{X}(s, t), Y(s, t))^T$ which, from (4), (6), (9), and (10) can be written as,

$$\mathbf{Z}(s, t) = \begin{pmatrix} \mu_X(s, t) \\ \mu_{\tilde{X}}(s, t) \\ \beta_0 + \beta_1 \mu_{\tilde{X}}(s, t) \end{pmatrix} + \begin{pmatrix} \sigma_X & 0 & 0 \\ 0 & \sigma_X & 0 \\ 0 & \beta_1 \sigma_X & \sigma_Y \end{pmatrix} \begin{pmatrix} w_X(s, t) \\ w_{\tilde{X}}(s, t) \\ w_Y(s, t) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \tau_Y \varepsilon_Y(s, t) \end{pmatrix} \quad (\text{A.1})$$

where each of $w_X(s, t)$, $w_{\tilde{X}}(s, t)$ and $w_Y(s, t)$ follow mean zero Gaussian processes with covariance functions given by (5), (11), and $\rho_Y(\cdot)$, respectively. While $w_Y(s, t) \perp w_X(s, t)$ and, by extension, $w_Y(s, t) \perp w_{\tilde{X}}(s, t)$ where “ $\perp$ ” denotes “independence”, $w_X(s, t)$ and $w_{\tilde{X}}(s, t)$ are dependent with cross-covariance function given by (12). Finally, $\varepsilon_Y(s, t)$ denotes a mean 0, unit variance white noise process. Because each of $w_X(s, t)$, $w_{\tilde{X}}(s, t)$, $w_Y(s, t)$, and $\varepsilon_Y(s, t)$ are distributed as a Gaussian distribution, $\mathbf{Z}(s, t)$ also follows a Gaussian distribution with mean $(\mu_X(s, t), \mu_{\tilde{X}}(s, t), \beta_0 + \beta_1 \mu_{\tilde{X}}(s, t))^T$ and covariance matrix given by,

$$\begin{pmatrix} \sigma_X^2 & \sigma_X^2 \mathbb{C}_{X\tilde{X}}(s, s; t, t) & \sigma_X^2 \beta_1 \mathbb{C}_{X\tilde{X}}(s, s; t, t) \\ \sigma_X^2 \mathbb{C}_{X\tilde{X}}(s, s; t, t) & \sigma_X^2 \mathbb{C}_{\tilde{X}\tilde{X}}(s, s; t, t) & \sigma_X^2 \beta_1 \mathbb{C}_{\tilde{X}\tilde{X}}(s, s; t, t) \\ \sigma_X^2 \beta_1 \mathbb{C}_{X\tilde{X}}(s, s; t, t) & \sigma_X^2 \beta_1 \mathbb{C}_{\tilde{X}\tilde{X}}(s, s; t, t) & \sigma_X^2 \beta_1^2 \mathbb{C}_{\tilde{X}\tilde{X}}(s, s; t, t) + \sigma_Y^2 + \tau_Y^2 \end{pmatrix} \quad (\text{A.2})$$

which is a special case of (14) where $(s, t) = (s', t')$. From (A.2), note

$$0 \leq \text{Corr}(X(s, t), \tilde{X}(s, t)) = \mathbb{C}_{XX}(s, s; t, t) / \sqrt{\mathbb{C}_{XX}(s, s; t, t)} \leq 1 \text{ which implies}$$

$$0 \leq \mathbb{C}_{XX}^2(s, s; t, t) \leq \mathbb{C}_{XX}(s, s; t, t) \leq 1 \text{ where the last inequality arises because } \mathbb{C}_{XX}(s, s; t, t) = \mathbb{E}_{(r, u, v, m)}(\rho_X(r, v; u, m)) \text{ where } (r, u) \text{ and } (v, m) \text{ are iid with distribution } K \text{ and } 0 \leq \rho_X(s, s'; t, t') \leq 1.$$

Proof of Results. Using properties of the multivariate Gaussian distribution, the implied distribution of $Y(s, t) | X(s, t)$ is Gaussian with mean

$$\beta_0 + \beta_1 \mu_X(s, t) + \beta_1 \mathbb{C}_{XX}(s, s; t, t)(X(s, t) - \mu_X(s, t)).$$

Hence, the coefficient of $X(s, t)$ on $Y(s, t)$ is $\beta_1 \mathbb{C}_{XX}(s, s; t, t)$ and $0 \leq \rho_{XX}(s, s; t, t) \leq 1$ is a multiplicative bias term which proves Result 1. Again by the properties of the multivariate Gaussian distribution,

$$\text{Var}(Y(s, t) | X(s, t)) = \sigma_X^2 \beta_1^2 (\mathbb{C}_{XX}(s, s; t, t) - \mathbb{C}_{XX}^2(s, s; t, t)) + \sigma_Y^2 + \tau_Y^2 \geq \sigma_Y^2 + \tau_Y^2 = \text{Var}(Y(s, t) | \tilde{X}(s, t))$$

leading to the proof of Result 2.

References

1. Stein ML. Space-time covariance functions. *Journal of the American Statistical Association*. 2005; 100:310–321.
2. Gneiting, T.; Genton, MG.; Guttorp, P. Geostatistical space-time models: Stationarity, separability, and full symmetry. In: Finkenstadt, B.; Held, L.; Isham, V., editors. *Statistical Methods for Spatio-temporal Systems*. Boca Raton, FL: Chapman and Hall/CRC; 2007. p. 151-175.
3. Gelfand AE, Kim HJ, Sirmans CF, Banerjee S. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*. 2003; 98:387–396.
4. Stroud JR, Müller P, Sansó B. Dynamics Models for Spatiotemporal Data. *Journal of the Royal Statistical Society Series B*. 2001; 63:673–689.
5. Huerta G, Sansó B, Stroud JR. A Spatiotemporal Model for Mexico City Ozone Levels. *Applied Statistics*. 2004; 53:231–248.
6. Gelfand AE, Banerjee S, Gamerman D. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*. 2005; 16:465–479.
7. Abdul-Wahab SA, Bakheit CS, Al-Alawi SM. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling and Software*. 2005; 20:1263–1271.
8. Reich B, Fuentes M, Dunson DB. Bayesian spatial quantile regression. *Journal of the American Statistical Association*. 2011; 106
9. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology*. 2000; 11:320–326. [PubMed: 10784251]
10. Welty LJ, Zeger SL. Are the acute effects of particulate matter on mortality in the national morbidity, mortality, and air pollution study the result of inadequate control for weather and season? A sensitivity analysis using flexible distributed lag models. *American Journal of Epidemiology*. 2005; 162:80–88.
11. Welty LJ, Peng RD, Zeger SL, Dominici F. Bayesian distributed lag models: Estimating the effects of particulate matter air pollution on daily mortality. *Biometrics*. 2009; 65:282–291. [PubMed: 18422792]
12. Ravines RR, Schmidt AM, Migon HS. Revisiting distributed lag models through a Bayesian perspective. *Applied Stochastic Models in Business and Industry*. 2006; 22:193–210.

13. Knorr-Held L, Richardson S. A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Journal of the Royal Statistical Society Series C*. 2003; 52:169–183.
14. Almon S. The distributed lag between capital appropriations and expenditures. *Econometrica*. 1965; 33:178–196.
15. Koyck, LM. *Distributed Lags and Investment Analysis*. North-Holland: Amsterdam; 1954.
16. Frances, PH.; van Oest, R. Economic Institute. Erasmus University Rotterdam; 2004. On the econometrics of the Koyck model, Technical Report.
17. Zanobetti A, Wand MP, Schwartz J, Ryan LM. Generalized additive distributed lag models: Quantifying mortality displacement. *Biostatistics*. 2000; 1:279–292. [PubMed: 12933509]
18. Heaton MJ, Gelfand AE. Spatial regression using kernel averaged predictors. *Journal of Agricultural, Biological and Environmental Statistics*. 2011; 16:233–252.
19. Gneiting T. Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*. 2002; 97:590–600.
20. Zhang H. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*. 2004; 99:250–261.
21. Heaton, MJ. Ph.D. thesis. Duke University; 2011. Kernel Averaged Predictors for Space and Space-Time Processes.
22. Banerjee, S.; Carlin, BP.; Gelfand, AE. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC; 2004.
23. Kent JT, Mohammadzadeh M, Mosammam AM. The Dimple in Gneiting's Spatial-Temporal Covariance Model. *Biometrika*. 2011; 98:489–494.
24. Zhang K, Fan W. Forecasting skewed biased stochastic ozone days: Analyses, solutions and beyond. *Knowledge and Information Systems*. 2008; 14:299–326.
25. Sahu SK, Gelfand AE, Holland DM. High resolution space-time ozone modeling for assessing trends. *Journal of the American Statistical Association*. 2007; 102:1221–1234. [PubMed: 19759840]
26. Li B, Genton M, Sherman M. A nonparametric assessment of properties of spacetime covariance functions. *Journal of the American Statistical Association*. 2007; 102:736–744.
27. Aneja VP, Li Z, Das M. Ozone Case Studies at High Elevation in the Eastern United States. *Chemosphere*. 1994; 29:1711–1733.
28. Chevalier A, Gheusi F, Delmas R, Ordóñez C, Sarrat C, Zbinden R, Thouret V, Athier G, Cousin JM. Influence of Altitude on Ozone Levels and Variability in the Lower Troposphere: A Ground-Based Study for Western Europe Over the Period 2001–2004. *Atmospheric Chemistry and Physics*. 2007; 7:4311–4326.
29. Brodin M, Helming D, Oltmans S. Seasonal Ozone Behavior Along an Elevation Gradient in the Colorado Front Range Mountains. *Atmospheric Environment*. 2010; 44:5305–5315.
30. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society Series B*. 2002; 64:583–639.
31. Finley AO, Sang H, Banerjee S, Gelfand AE. Improving the performance of predictive process modeling for large datasets. *Computational Statistics and Data Analysis*. 2009; 53:2873–2884. [PubMed: 20016667]
32. Wikle, CK. Low-rank representations for spatial processes. In: Gelfand, AE.; Diggle, PJ.; Fuentes, M.; Guttorp, P., editors. *Handbook of Spatial Statistics*. Chapman and Hall/CRC; 2010. p. 107–118.
33. Sun, Y.; Li, B.; Genton, M. Geostatistics for large datasets. In: Montero, JM.; Porcu, E.; Schlather, M., editors. *Advances and Challenges in Space-time Modelling of Natural Events*. Springer; 2011. p. 55–78.

Figure 1. Empirical (a) average daily ozone level and (b) average daily temperature for July 1, 2010 – July 14, 2010 across California. Each square on the grid corresponds to a single station.

Figure 2.

(a) Scatterplot of ozone concentration vs. average temperature at aligned temperature and ozone monitoring stations during July 2010 in California. (b) Correlation between ozone concentration and average temperature by spatial location. The asterisks indicate a negative correlation.

Figure 3. Several validation diagnostics of the STKAP model for the ozone production study. (a) Residuals vs. fitted values plot for O_3 . (b) QQ-plot for O_3 residuals. (c) Histogram of empirical lag 1 temporal autocorrelations from each ozone station.

Figure 4. Marginal posterior distributions for (a) the purely spatial range $r_1(0)$ in miles and (b) the purely temporal range r_2 in days. The discrete values for $r_1(0)$ represent 10%, 30%, 50%, 70% and 90% of the maximum spatial distance between temperature stations.

Figure 5.

Posterior mean of (a) $\xi_1(s)r_1(0)/2$ (the standard deviation of the purely spatial kernel) in miles and (b) $\xi_2(s)r_2/3$ (the mean of the exponential temporal kernel) in days. Ninety-five percent central credible intervals for γ_{11} and γ_{21} are $(-0.003, -0.001)$ and $(-0.009, 0.005)$, respectively, and suggest that the width of the spatial kernel decreases with elevation while the spread of temporal kernel is constant with elevation.

Figure 6.

Posterior mean of the $R^2(s)$ statistic according to the STKAP model. The posterior mean of R^2 under the PtP model was 0.35. For areas of higher elevation and much of Southern California, spatially and temporally lagged daily temperature explains more of the variability in ozone as compared to temperature at a given station and day.

Table 1

Comparison of empirical bias, root mean square error (RMSE), and 95% credible interval coverage for the STKAP and PtP models fit to data simulated from the STKAP model (NA signifies “not applicable”).

Parameter	Bias		RMSE		Coverage	
	STKAP	PtP	STKAP	PtP	STKAP	PtP
β_1	-0.027	-0.467	0.059	0.470	0.91	0.00
τ_y^2	-0.020	0.101	0.031	0.101	0.98	0.05
σ_x^2	0.003	0.008	0.036	0.037	0.94	0.92
γ_{11}	-0.07×10^{-3}	NA	0.72×10^{-3}	NA	0.98	NA
γ_{21}	-0.40×10^{-3}	NA	0.96×10^{-3}	NA	0.96	NA

Table 2

Comparison of empirical bias, root mean square error (RMSE), and 95% credible interval coverage for the STKAP and PtP models fit to data simulated from the PtP model (NA signifies “not applicable”).

Parameter	Bias		RMSE		Coverage	
	STKAP	PtP	STKAP	PtP	STKAP	PtP
β_1	-0.004	0.001	0.026	0.026	0.98	0.98
τ_y^2	0.016	0.001	0.029	0.023	0.93	0.96
σ_x^2	0.-0.001	-0.002	0.033	0.033	0.98	0.97
γ_{11}	-0.32×10^{-3}	NA	0.42×10^{-3}	NA	1.00	NA
γ_{21}	-0.24×10^{-3}	NA	0.27×10^{-3}	NA	1.00	NA

Table 3

Posterior quantiles for common parameters in the California ozone study.

Quantile	β_0		β_1		τ_y^2		σ_x^2	
	STKAP	PtP	STKAP	PtP	STKAP	PtP	STKAP	PtP
0.025	45.22	44.73	1.14	1.02	109.00	111.72	54.80	54.94
0.500	45.81	45.32	1.22	1.08	117.39	120.54	57.46	57.46
0.975	46.40	45.91	1.29	1.14	126.77	129.49	60.27	60.36