Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Statistical Planning and Inference 130 (2005) 149–166

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Spatio-temporal change-point modeling

Anandamayee Majumdar[a], Alan E. Gelfand[b,*], Sudipto Banerjee[c]

[a] University of Connecticut, Storrs, CT 06269-4120, USA
[b] Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, USA
[c] Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA

Received 16 June 2003; received in revised form 31 July 2003; accepted 5 August 2003
Available online 3 August 2004

Abstract

There is by now a substantial literature on spatio-temporal modeling. However, to date, there exists essentially no literature which addresses the issue of process change from a certain time. In fact, if we look at change points for purely time series data, the customary form is to propose a model involving a mean or level shift. We see little attempting to capture a change in association structure. Part of the concern is how to specify flexible ways to bridge the association across the time point and still ensure that a proper joint distribution has been defined for all of the data. Introducing a spatial component evidently adds further complication. We want to allow for a change-point reflecting change in both temporal and spatial association. In this paper we propose a constructive, flexible model formulation through additive specifications. We also demonstrate how computational concerns benefit from the availability of temporal order. Finally, we illustrate with several simulated datasets to examine the capability of the model to detect different types of structural changes.

Keywords: Hierarchical model; Markov chain Monte Carlo; Separable covariance function; Stationary process

1. Introduction

The topic of this paper is a particular version of a change point problem. In fact, we are interested in clarifying the notion of a change point in the context of a model for spatio-temporal data. More precisely, the structure we seek is the following. Let $\mathbf{s}$ be a location in some region of interest D where $D \in R^p$ and let t be a time in an interval, say $[0, T]$. Then

* Corresponding author.
E-mail address: alan@stat.duke.edu (A.E. Gelfand).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2003.08.022

if we consider the process $Y(\mathbf{s}, t)$ for $(\mathbf{s}, t) \in D \times [0, T]$, we want $Y(\mathbf{s}, t)$ to come from one spatio-temporal process for $t \leqslant t_0$ and a different process for $t > t_0$. t_0 is referred to as the changepoint and is assumed to be unknown. The change in process can involve change in mean structure, a change in variability and also a change in association structure.

Allowing change in association structure leads to the primary technical issue which is how to *bridge* the transition from the first process to the second. That is, how shall we model the dependence between $Y(\mathbf{s}, t)$ and $Y(\mathbf{s}', t')$ for $\mathbf{s}, \mathbf{s}' \in D$ when $t \leqslant t_0$ and $t' > t_0$? The concern is to ensure that for any set of locations $\mathbf{s_1}, \ldots, \mathbf{s_n}$ in D and any set of locations $t_1, \ldots, t_m$ in $[0, T]$ and any $t_0 \in [0, T]$, the covariance matrix for the resulting $\{Y(s_i, t_j)\}$ will be positive definite. We return to this question below.

The change point problem has a rich history in the statistics literature and we offer only a very brief review here. Such problems arise in most fields of study including, for example, signal detection in engineering settings, response to extreme shocks of market prices in financial and economic settings and change in reliability when manufacturing processes move out of control. The earliest work treats the case of a changepoint in the mean for a sequence of i.i.d. variables from a Gaussian distribution. Seminal rigorous results in this regard appear in the classic paper of Chernoff and Zacks (1964). Subsequent work has treated changes in regression structure and changes in variability as well as the non-Gaussian case. See the paper of Zacks (1983) as well as those of Csorgö and Horváth (1988) and Krishnaiah and Miao (1988).

The change point problem can be formulated either sequentially or retrospectively. In the former, we make a decision regarding the occurrence of a change based upon the data to the present time. In the latter, we look back upon a collected sequence of data and try to determine if a change point has occurred in the sequence and, if so, where it occurred. Our approach below will be implemented within the Bayesian framework which requires the retrospective formulation. Noteworthy Bayesian treatments of the i.i.d. case include Smith (1975), Raftery and Akman (1986) and Carlin et al. (1992).

Leaving the i.i.d. case in a change point setting naturally leads to viewing the data sequence as a time series. Now the notion of a change point becomes a special case of general temporal evolution of the model where distribution theory is allowed to change with each time innovation. Examples include the threshold autoregressive (TAR) model as in, e.g., Tong (1983) and the more general functional coefficient autoregressive models in Chen and Tsay (1993). In this regard, the change point version becomes a level shift model. One mean operates before the change point, a different one after with no change in the autoregressive structure (See, e.g., Brockwell and Davis, 1991.) A different class of temporal evolution models introduce stochastic volatility, i.e., temporal change in model (variability), along with ARCH, GARCH, etc. specifications. See, e.g., Taylor (1994) for a broad review and Kim et al. (1998) for a more statistical presentation as well as Jacquier et al. (1994) for such analysis in the Bayesian framework. Recently, Perreault et al. (2000) use Markov chain Monte Carlo methods to fit a model of a hydrometerological time series with a sudden change.

To connect the preceding discussion to the change point problem which is our interest here, we extend the time series setting to envision a spatial process at each $t \in [0, T]$. That is, rather than a single random variable at time t (or even a finite-dimensional multivariate random variable at time t) we now have a stochastic process at each t. Raftery (1994), in the context of modeling changing curves, offers brief discussion of changing spatial surfaces. A very recent paper by Ecker and Isakson (2004) employs space–time land value data to model a land price mean shift as a function of parcel size. The authors introduce time in the regression structure but incorporate spatial dependence in the error structure.

Here, we work with Gaussian process models for the data. Gaussian processes are, of course, widely used and offer many advantages including ease of specification and convenient distribution theory (The latter is particularly helpful with regard to computational needs for Bayesian model fitting as we elaborate below.) However, the model in Section 2.1 can be viewed as a hierarchical model with conditionally independent first stage Gaussian specification. A different exponential family model for the first stage could be adopted (in the spirit of Diggle et al., 1998) but we have not pursued this here.

A last point concerns the treatment of the time scale. Do we view time as discrete or as continuous? If time is discretized to say equally spaced points as in usual time series we can write our model, in general, as

$$Y_t(\mathbf{s}) = \mu_t(\mathbf{s}) + R_t(\mathbf{s}) + \varepsilon_t(\mathbf{s}), \quad t = 1, \ldots, M. \tag{1}$$

In (1), $\mu_t(\mathbf{s})$ denotes the mean structure, $R_t(\mathbf{s})$ is the mean 0 residual spatial process at time t and the $\varepsilon_t(\mathbf{s})$ are i.i.d pure error terms. But then, what do we mean by a change point model? We can easily introduce a change point in the mean surface, e.g., $\mu_t(\mathbf{s}) = \mu_t^{(1)}(\mathbf{s}), t \leqslant t_0, = \mu_t^{(2)}(\mathbf{s}), t > t_0$. Changes in variability can be introduced in either or both of R_t or $\varepsilon_t(\mathbf{s})$. However, for changes in the dependence structure we must employ $R_t(\mathbf{s})$.

If we write $R_t(\mathbf{s}) = U(\mathbf{s}) + \alpha_t$, i.e., an additive form in space and time we can now model α_t using time series change point ideas, as mentioned above. However, no change in spatial association occurs over time so this form is less than what we seek.

Instead, we could consider $R_t(\mathbf{s})$ to be independent across t, assigning a distinct spatial process model to each t. In this way, we could observe temporal evolution of spatial pattern. We could impose a change point structure by assuming the $R_t(\mathbf{s})$ all have the same spatial process model for $t \leqslant t_0$ and they have a different spatial model for $t > t_0$. Regardless, such modeling would not introduce any temporal dependence.

We could adopt a dynamic model for $R_t(\mathbf{s})$, i.e.,

$$R_t(\mathbf{s}) = R_{t-1}(\mathbf{s}) + \eta_t(\mathbf{s}), \tag{2}$$

where in (2) the $\eta_t(\mathbf{s})$ are independent innovations of a spatial process. A change point version of (2) would specify $\eta_t(\mathbf{s})$ to have one spatial process model for $t \leqslant t_0$ and a different spatial model for $t > t_0$. This form of the model is straightforward to work with but implies *explosive* behavior for variance and covariance. That is,

$$\mathrm{Cov}(R_t(\mathbf{s}), R_{t'}(\mathbf{s}')) = \begin{cases} \min(t, t')C^{(1)}(\mathbf{s}, \mathbf{s}'), & t, t' \leqslant t_0, \\ tC^{(1)}(\mathbf{s}, \mathbf{s}'), & t \leqslant t_0, \ t > t_0, \\ t_0 C^{(1)}(\mathbf{s}, \mathbf{s}') & \\ \quad + \min(t - t_0, t' - t_0)C^{(2)}(\mathbf{s}, \mathbf{s}'), & t, t' > t_0, \end{cases}$$

where $C^{(1)}$ and $C^{(2)}$ denote the pre- and post-change point covariance functions. This could be remedied by introducing an auto-regression parameter γ, $|\gamma| < 1$ before $R_{t-1}(\mathbf{s})$ in (2).

For this paper, we choose to formulate a stationary spatio-temporal process specification. Hence we do not pursue (2) any further. In fact, in the sequel, we work with t being continuous and with stationary space time covariance functions of the form $C(\mathbf{s}-\mathbf{s}', t-t')$. In Section 2, we define our class of models in a constructive way rather than through existence arguments. We present the properties of these models and discuss model fitting and inferential details. This discussion reveals an interesting distinction between the case of a purely spatial model and one that introduces an additional white noise component.

We propose to permit change which can be in the mean, in the variability and in the association structure. It is evident that small changes will be difficult to detect. This is the case in all change point settings; often a minimum magnitude of change is specified. It is also clear that change of one sort can be difficult to distinguish from change of another sort. For example, a change in mean may be difficult to distinguish from a change in variability. A change in variability may be difficult to distinguish from a change of rate of decay in spatial association. A change in temporal decay in association may be difficult to separate from a change of spatial decay in association. Obviously, a large number of locations and a large number of time points will be needed to successfully identify changes. To clarify these issues a bit, in Section 3 we present a small illustrative simulation study. In Section 4, we conclude with some discussion and extensions.

2. An additive spatio-temporal change point model

2.1. The general model

In general, suppose locations $\mathbf{s}$ belong to R^d, the d-dimensional Euclidean space and timepoints t belong to the positive real line R^+. We assume the process to be multivariate and finite second moment. Hence the process is a collection of random variables $Y(\mathbf{s}, t)$ on $R^d \times R^+$. We denote the covariance of the process by $\mathrm{Cov}(Y(\mathbf{s}, t), Y(\mathbf{s}', t'))$ for $\mathbf{s}, \mathbf{s}' \in R^d$ and $t, t' \in R^+$.

For a weakly stationary process we employ separable covariance functions, i.e., $C(\mathbf{s}, \mathbf{s}', t, t') = \sigma^2 \rho^1_\psi(\mathbf{s}-\mathbf{s}')\rho^2_\phi(t-t')$, where ρ^1_ψ and ρ^2_ϕ are valid correlation functions in R^d and R^1. For an isotropic process we write

$$C(\mathbf{s}, \mathbf{s}', t, t') = \sigma^2 \rho^1_\psi(\|\mathbf{s}-\mathbf{s}'\|)\rho^2_\phi(|t-t'|), \tag{3}$$

where $(\|\mathbf{s}-\mathbf{s}'\|)$ is the Euclidean distance between the two locations. The form in (3) is discussed in Mardia and Goodall (1993). See also Banerjee and Gelfand (2002). Non-separable forms are available. See, e.g., Cressie and Huang (1999) and Gneiting (2002). In our examples below, we illustrate with the exponential class where $\rho^1_\psi(\|\Delta_{\mathbf{s}}\|) = \exp(-\psi\|\Delta_{\mathbf{s}}\|)$, $\rho^2_\phi(|\Delta_t|) = \exp(-\phi|\Delta_t|)$. More flexible versions including the Matèrn class and the power exponential (Stein, 1999) could be adopted. Our processes will be Gaussian unless otherwise stated.

Our approach to the problem is to view a spatio-temporal change point model as one which has a mean structure that depends only on the time (i.e., no covariates at the moment) and an additive mean 0 spatio-temporal error structure that changes from before to after the change point.

Letting t_0 denote the change point, the probability model can be described as follows:

$$\begin{aligned} Y(\mathbf{s},t) &= \mu_t + U(\mathbf{s},t) + W(\mathbf{s},t) + \varepsilon(\mathbf{s},t), \quad t \leqslant t_0 \\ &= \mu_t + U(\mathbf{s},t) + V(\mathbf{s},t) + \varepsilon'(\mathbf{s},t), \quad t > t_0, \end{aligned} \tag{4}$$

where $\varepsilon(\mathbf{s},t) \overset{\text{i.i.d.}}{\sim} N(0,\tau_1^2)$, $\varepsilon'(\mathbf{s},t) \overset{\text{i.i.d.}}{\sim} N(0,\tau_2^2)$. Here, we assume that $U(\mathbf{s},t)$ follows a spatio-temporal process with covariance function $\sigma_1^2\rho^1_{\psi_1}(\mathbf{s}-\mathbf{s}')\rho^2_{\phi_2}(t-t')$, $W(\mathbf{s},t)$ follows a spatio-temporal process with covariance function $\sigma_2^2\rho^1_{\psi_2}(\mathbf{s}-\mathbf{s}')\rho^2_{\phi_2}(t-t')$ and $V(\mathbf{s},t)$ follows a spatio-temporal process with covariance $\sigma_3^2\rho^1_{\psi_3}(\mathbf{s}-\mathbf{s}')\rho^2_{\phi_3}(t-t')$. All three processes are assumed independent of each other and independent of the ε's. The sum $U(\mathbf{s},t)+W(\mathbf{s},t)$ denotes the pre-change spatio-temporal random effect, the sum $U(\mathbf{s},t)+V(\mathbf{s},t)$ denotes the post-change effect. Then, given the $\{\mu_t\}$ the resulting covariance structure is

$$\operatorname{Cov}(Y(\mathbf{s},t),Y(\mathbf{s}',t')) = \begin{cases} \sigma_1^2\rho^1_{\psi_1}(\mathbf{s}-\mathbf{s}')\rho^2_{\phi_1}(t-t') \\ \quad +\sigma_2^2\rho^1_{\psi_2}(\mathbf{s}-\mathbf{s}')\rho^2_{\phi_2}(t-t'), & t\leqslant t_0,\ t'\leqslant t_0, \\ \sigma_1^2\rho^1_{\psi_1}(\mathbf{s}-\mathbf{s}')\rho^2_{\phi_1}(t-t'), & t\leqslant t_0,\ t'>t_0, \\ \sigma_1^2\rho^1_{\psi_1}(\mathbf{s}-\mathbf{s}')\rho^2_{\phi_1}(t-t') \\ \quad +\sigma_3^2\rho^1_{\psi_2}(\mathbf{s}-\mathbf{s}')\rho^2_{\phi_2}(t-t'), & t>t_0,\ t'>t_0. \end{cases} \tag{5}$$

Correlations can be calculated directly from (5). When both t and t' are $\leqslant t_0$ or both t and t' are $> t_0$, they emerge as variance weighted convex combinations of correlations associated with the component processes. When t is $\leqslant t_0$ and $t' > t_0$ we obtain a plausible transition.

The role of the $U(\mathbf{s},t)$ process in (4) is to accommodate dependence between measurements before t_0 and after t_0 as evidenced in (5). The $W(\mathbf{s},t)$ and $V(\mathbf{s},t)$ represent the pre- and post-change point adjustments to the spatio temporal model. As a result, neither the pre-change point or post-change point processes for the Y's have separable covariance structure. In this regard, the independence assumption sacrifices little flexibility. Also, the constructice form in (4) ensures that for any set of locations and time points, the resulting joint covariance matrix will be positive definite. With respect to the μ_t, we set $\mu_t=\mu_1$, $t\leqslant t_0$, and $\mu_t=\mu_2$, $t>t_0$. In the presence of covariate information at each location more flexible choices might be $\mu(\mathbf{s},t)=X^{\mathrm{T}}(\mathbf{s})\beta_1$, $t\leqslant t_0$, and $\mu(\mathbf{s},t)=X^{\mathrm{T}}(\mathbf{s})\beta_2$, $t>t_0$.

Let us define $M(t)$ to be the largest j such that $M(t_j)\leqslant t$. Thus, $M(t_0)$ is the number of time points before the change-point and $M-M(t_0)$ the number after. Let $W^{\mathrm{T}}_{t^-} = (W^{\mathrm{T}}_{t_1}, W^{\mathrm{T}}_{t_2}, \ldots, W^{\mathrm{T}}_{t_0})$ and $W^{\mathrm{T}}_{t^+} = (W^{\mathrm{T}}_{t_0+1}, W^{\mathrm{T}}_{t_0+2}, \ldots, W^{\mathrm{T}}_{M})$. Analogously, we partition $V^{\mathrm{T}} = (V^{\mathrm{T}}_{t^-}, V^{\mathrm{T}}_{t^+})$, $U^{\mathrm{T}} = (U^{\mathrm{T}}_{t^-}, U^{\mathrm{T}}_{t^+})$, and $Y^{\mathrm{T}} = (Y^{\mathrm{T}}_{t^-}, Y^{\mathrm{T}}_{t^+})$. Hence, the vector U has an $MN\times MN$ covariance matrix $\sigma_1^2R(\phi_1)\otimes R(\psi_1)$. Similarly, V has $MN\times MN$ covariance matrix $\sigma_2^2R(\phi_2)\otimes R(\psi_2)$ and W has $MN\times MN$ covariance matrix $\sigma_3^2R(\phi_3)\otimes R(\psi_3)$. Here, $((R(\psi_l)))_{i,i'} = \rho^1_{\psi_l}(\mathbf{s}_i-\mathbf{s}'_i)$ and $((R(\phi_l)))_{j,j'} = \rho^2_{\phi_l}(t_j-t'_j)$.

Also, W_{t^-} has an $M(t_0)N\times M(t_0)N$ spatial covariance matrix $\sigma_2^2R^{M(t_0)}(\phi_2)\otimes R(\psi_2)$ and V_{t^+} has an $(M-M(t_0))N\times(M-M(t_0))N$ spatial covariance matrix $\sigma_3^2R^{M-M(t_0)}(\phi_3)\otimes R(\psi_3)$, where $R^{M(t_0)}(\phi_2)$ denotes the upper left $M(t_0)N\times M(t_0)N$ matrix of $R(\phi_2)$ and $R^{M-M(t_0)}(\phi_3)$ denotes the lower right $M(t_0)N\times M(t_0)N$ matrix of $R(\phi_3)$.

2.2. Model fitting and inferential details

We work in dimension $d=2$. For simplicity, we assume a constant mean μ_1 over the region of interest D before the change point and mean μ_2 over the region after the change point. Letting $\theta=(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \psi_1, \psi_2, \psi_3, \phi_1, \phi_2, \phi_3, t_0)$, the conditional log-likelihood given U, W and V and the change point t_0 is

$$\begin{aligned} & l(\mu_1, \mu_2, \tau_1^2, \tau_2^2, t_0; U, W, V, Y) \\ & \quad \propto -\frac{M(t_0)N}{2}\log(\tau_1^2) - \frac{(M-M(t_0))N}{2}\log(\tau_2^2) \\ & \qquad - (Y_{t^-} - \mu_1 1 - U_{t^-} - W_{t^-})^{\mathrm{T}}(Y_{t^-} - \mu_1 1 - U_{t^-} - W_{t^-})/2\tau_1^2 \\ & \qquad - (Y_{t^+} - \mu_2 1 - U_{t^+} - V_{t^+})^{\mathrm{T}}(Y_{t^+} - \mu_2 1 - U_{t^+} - V_{t^+})/2\tau_2^2. \end{aligned} \tag{6}$$

Incorporating the distribution of U, V and W with a prior π for θ the full Bayesian model becomes

$$\begin{aligned} & \log(f(U, V, W, \theta|Y)) \\ & \quad \propto l(\mu_1, \mu_2, \tau_1^2, \tau_2^2, t_0; U, W, V, Y) - \{U^{\mathrm{T}}R(\phi_1)^{-1} \\ & \qquad \otimes R(\psi_1)^{-1}U/2\sigma_1^2\} - \{W^{\mathrm{T}}R(\phi_2)^{-1} \otimes R(\psi_2)^{-1}W/2\sigma_2^2\} - \{V^{\mathrm{T}}R(\phi_2)^{-1} \\ & \qquad \otimes R(\psi_2)^{-1}V/2\sigma_3^2\} - NM/2\log(\sigma_1^2) - NM/2\log(\sigma_2^2) - NM/2\log(\sigma_3^2) \\ & \qquad - M/2\log|R(\psi_1)| - N/2\log|R(\phi_1)| - M/2\log|R(\psi_2)| \\ & \qquad - N/2\log|R(\phi_2)| - M/2\log|R(\psi_3)| - N/2\log|R(\phi_3)| + \log(\pi(\theta)). \end{aligned} \tag{7}$$

Customarily, we assume prior independence for the components of θ. Then we would propose proper fairly non-informative priors for these components, e.g., vague normal for the μs, vague inverse Gamma for σ^2s and τ^2s, e.g., $\sigma_l^2 \sim \mathrm{IG}(2, \beta_l), l = 1, 2, 3, \tau_1^2 \sim \mathrm{IG}(2, \beta_4), \tau_2^2 \sim \mathrm{IG}(2, \beta_5)$, and Gamma's for ϕ_l and ψ_l. For t_0 we assume that $0 < t_0 < t_M$, noting that $t_0 < t_1$ means no change point. In fact, we can discretize the parameter space for t_0 to $\{t_1, t_2, \ldots, t_M\}$ defining "$t_0 = t_j$" to be $t_{j-1} < t_0 \leqslant t_j$. This follows because the likelihood does not change as t_0 varies over this interval. Therefore, we can use a discrete prior $p_{t_0}; t_0 = t_1, t_2, \ldots, t_M$. The full conditional distribution for t_0 is discrete with probabilities proportional to p_t times (6) evaluated at $t_0 = t$.

Fitting of model (5) can only be attempted using Gibbs sampling. In principle, we could marginalize (4) to $L(\mu_1, \mu_2, \tau_1^2, \tau_2^2, t_0; Y)$ by integrating over the U, V and the W thereby substantially reducing dimensionality. However, if we retain the intermediary U, V, W, the model in (6) can be viewed as a hierarchical specification. If we fit the model hierarchically, i.e., by generating U, V, and W from their full conditional multivariate normal distributions in the Gibbs sampler then the computations are easier to manage. The full conditionals are supplied in Appendix A. We note that the full conditionals for μ_1 and μ_2 follow normal distributions while the τ^2's and σ^2's follow inverse Gammas. The full conditionals of the ϕ and the ψ's are awkward. However, the reader can verify from the appendix that, in updating the ψ's and the ϕ's we can preserve the Kronecker forms in (5). Rather than having to work with $MN \times MN$ matrices, we can work with $M \times M$ and $N \times N$ matrices. So we use a Metropolis Hastings update in this context. The appendix reveals that the full conditional distribution for U, W and V do necessitate working with $MN \times MN$ matrices but the required normal draws can be made through Cholesky decomposition of the available inverse matrix. As a result, the output of the MCMC fitting will provide essentially posterior samples $\theta_l^*, l = 1, 2, \ldots, L$ from $f(\theta|\mathbf{Y})$.

2.3. The purely spatial model

Inclusion of the white noise terms $\varepsilon(\mathbf{s}, t)$ in (4) enables the model to incorporate error which is non-spatial. Such heterogeneity is usually attributed to measurement error or microscale variability. Evidently, the model in (4) can be specified without such error terms, i.e.,

$$
\begin{aligned}
Y(\mathbf{s}, t) &= \mu_1 + U(\mathbf{s}, t) + W(\mathbf{s}, t), \quad t \leqslant t_0 \\
&= \mu_2 + U(\mathbf{s}, t) + V(\mathbf{s}, t), \quad t > t_0.
\end{aligned} \tag{8}
$$

So, we are rid of the parameters τ_1^2 and τ_2^2 in the model. Hence, we have $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \psi_1, \psi_2, \psi_3, \phi_1, \phi_2, \phi_3, t_0)$. Now we could use the conditional log-likelihood given U and t_0 which takes advantage of the independence of W and V to yield

$$
\begin{aligned}
& l(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \psi_2, \psi_3, \phi_2, \phi_3; Y, U) \\
& \quad \propto -\{(Y_{t^-} - \mu_1 \mathbf{1} - U_{t^-})^{\mathrm{T}} R(\phi_2)^{-1} \otimes R(\psi_2)^{-1} (Y_{t^-} - \mu_1 \mathbf{1} - U_{t^-})/2\sigma_2^2\} \\
& \qquad - \{(Y_{t^+} - \mu_2 1 - U_{t^+}^{\mathrm{T}}) R(\phi_3)^{-1} \otimes R(\psi_3)^{-1} (Y_{t^+} - \mu_2 \mathbf{1} - U_{t^+})/2\sigma_3^2\} \\
& \qquad - M(t_0)N/2 \log(\sigma_2^2) - (M - M(t_0))N/2 \log(\sigma_3^2) - M(t_0)/2 \log|R(\psi_2)| \\
& \qquad - N/2 \log|R(\phi_2)|(M - M(t_0))/2 \log|R(\psi_3)| - N/2 \log|R(\phi_3)|.
\end{aligned} \tag{9}
$$

Notice that there is no way to condition so that the Y's are independent as in (4). Adding the distribution of U, with a prior on θ, $\pi(\theta)$ the full model yields

$$
\begin{aligned}
\log(f(U, \theta|Y)) \propto & - \{(Y_{t^-} - \mu_1 \mathbf{1} - U_{t^-})^{\mathrm{T}} R(\psi_2)^{-1} \otimes R(\phi_2)^{-1} (Y_{t^-} - \mu_1 \mathbf{1} \\
& - U_{t^-})/2\sigma_2^2\} - \{(Y_{t^+} - \mu_2 \mathbf{1} - U_{t^+})^{\mathrm{T}} R(\psi_3)^{-1} \otimes R(\phi_3)^{-1} \\
& \times (Y_{t^+} - \mu_2 \mathbf{1} - U_{t^+})/2\sigma_3^2\} - \{U^{\mathrm{T}} R(\psi_1)^{-1} \otimes R(\phi_1)^{-1} U/2\sigma_1^2\} \\
& - M/2 \log|R(\psi_1)| - N/2 \log|R(\phi_1)| - M(t_0)/2 \log|R(\psi_2)| \\
& - N/2 \log|R(\phi_2)| - (M - M(t_0))/2 \log|R(\psi_3)| \\
& - N/2 \log|R(\phi_3)| - MN \log(\sigma_1) - M(t_0)N/2 \log(\sigma_2^2) \\
& - (M - M(t_0))N/2 \log(\sigma_3^2) + \log(\pi(\theta)).
\end{aligned} \tag{10}
$$

The full conditional distributions associated with (10) are provided in Appendix B. Again, the ψ's and the ϕ's can be updated retaining the Kronecker form. Again, the full conditional distribution for U is normal involving an $MN \times MN$ covariance matrix.

3. A simulation study

We undertake a modest simulation study in order to see how well our model is able to identify different sorts of changes. In fact, as noted in the introduction, there are many types of changes that can be introduced, e.g., mean change, variance change, association change, spatial correlation change but no temporal association change or vice versa. Moreover, more than one of these changes could occur at t_0. In addition, what should the magnitude of these changes be? The price of a flexible model such as (4) is that a thorough simulation study is not feasible.

So instead we consider four different simulations, each with a particular objective. We generate 2400 points for this study under each of the four simulations. In particular, we set the number of spatial points $N = 40$ and the number of temporal points $M = 60$. The spatial locations are selected at random from the $[0, 10] \times [0, 10]$ square where as the time points are selected at random from the $[0, 10]$ interval.

We keep the change point t_0 fixed at $t_0 = 4.614$, which happens to be the thirtieth largest observation from the 60 points that we randomly selected from the $[0, 10]$ interval. In the first simulation we study the effect of change in decay of the spatial and temporal association. Letting the variances and the means remain the same, we let the spatial and the temporal association parameters change across the change point. That is, in (4), we let the decay of spatial and temporal association of $W(\mathbf{s}, t)$ and $V(\mathbf{s}, t)$ differ considerably. Effectively, we increased both the spatial range and the temporal range by a multiple of 4. In the third simulation, our objective is similar to the first simulation. But here we study the change in decay of the spatial and temporal association, along with a mean shift. Our objective in the second simulation is to study the effect of a change in variance, in fact, a four fold increase in variance. So we keep the decay parameters of $W(\mathbf{s}, t)$ and $V(\mathbf{s}, t)$ the same and assume zero mean shift across the change point. And finally, our objective of the fourth simulation, is similar to that of the second simulation, except that along with studying the effect of change in variance we also look at the change in mean or level shift across time. In choosing the decay parameters we keep in mind that the range of the spatial and the temporal processes involved are reasonable. We also keep in mind that the variance corresponding to the noise in (4) is not too large (which could obscure other changes) or too small (which leads to unstable computation if we are close to (8)) compared to the spatio-temporal variances. To quantify the objectives, we present in Table 1 the values of the parameters used in each simulation.

The summaries of the computational findings are given below. As a summary of the posterior distributions, obtained from running a Gibbs sampler, in Table 2 we display the median, the 2.5th and 97.5th percentile for each component of θ. The original values of the parameters are also presented.

From the results of the output in Table 2, we can see that the posterior inference has captured all components of θ in all simulations except μ_1 being slightly negative and μ_2 being slightly positive.

Fig. 1 shows a histogram of the posterior samples for the change point for each of the simulations. Noteworthy is the concentration of these histograms about the true value. Fig. 2 chooses a location ($\mathbf{s} = (5, 5)$, without loss of generality) and for a fine grid in the time scale plots the posterior predictive mean vs. t along with .95 predictive intervals for each of the simulations. The mean shift in simulations 3 and 4 is clearly detected and the change is to essentially the correct magnitude. Note also, the widening of the predictive intervals for simulations 3 and 4, in accordance with the increase in spatial variance from before the change point to after. For the same location and the same grid, Fig. 3 plots the variance of the posterior predictive distribution for an observation at this location. The variance change is clearly detected and the overall process variances are of the correct magnitudes.

Table 1
The parameter values across the four simulations

	Simulation 1	Simulation 2	Simulation 3	Simulation 4
μ_1	0	0	0	0
μ_2	0	0	5	5
σ_1^2	1	1	1	1
σ_2^2	1	4	1	4
σ_3^2	1	1	1	1
τ_1^2	1	1	1	1
τ_2^2	1	1	1	1
ϕ_1	3	3	3	3
ψ_1	3	3	3	3
ϕ_2	1.5	3	1.5	3
ψ_2	1.5	3	1.5	3
ϕ_3	6	3	6	3
ψ_3	6	3	6	3
t_0	4.614	4.614	4.614	4.614

Finally, in Figs. 4–7 we attempt to reveal a change in association structure. Change in covariance will be masked by change in σ_2^2 and σ_3^2. In fact, results in (5) indicate that process correlations arise as variance weighted combinations of correlations associated with the component processes. Hence, if $\phi_2 = \phi_3$ but different from ϕ_1 and/or $\psi_2 = \psi_3$ but different from ψ_1, i.e., "no change" in association structure, a change from σ_2^2 to σ_3^2 will imply a change in correlation for $Y(\mathbf{s}, t)$ across t_0. So, change in association structure in the presence of change in variance structure will be difficult to detect. Due to assumed isotropy given $\mathbf{s}$ we only need fix $\mathbf{s}'$ to determine $\|\mathbf{s} - \mathbf{s}'\|$ and then select a Δt. Then, for a fixed $\mathbf{s}$ and $\mathbf{s}'$ we obtained the posterior correlation and posterior covariance between $Y(\mathbf{s}, t)$ and $Y(\mathbf{s}', t+\Delta t)$ for the same grid of t values as above. In particular we set $\mathbf{s} = (5, 5)$ and $\mathbf{s}' = (5.14, 5.14)$ so that $\|\mathbf{s} - \mathbf{s}'\| = .2$ in Figs. 4 and 6. We also set $\Delta t = .1$. And in Figs. 5 and 7 we set $\mathbf{s}' = (5.071, 5.071)$ so that $\|\mathbf{s} - \mathbf{s}'\| = .1$. The values of $\|\mathbf{s} - \mathbf{s}'\|$ and Δt ensure non-negligible association in space and time. We note that in Figs. 4 and 5 the posterior correlations decrease after the change point in simulations 1 and 3 and increase in simulations 2 and 4. Likewise in Figs. 6 and 7 the posterior covariance decrease after the change point in simulations 1 and 3 and increase in simulations 2 and 4. Consistent with the foregoing discussion we see a change point in association structure in all four simulations.

4. Discussion and extensions

Change point detection for a spatio-temporal process is a complex problem. One can envision change in mean structure, change in variance structure and change in association. We have attempted only a rudimentary investigation here, working with a convenient constructive model form. We have demonstrated through simulation that changes can be detected. However small changes will be difficult to reveal and concurrent changes of different types may be difficult to separate and may even mask each other. The simulation study could be

Table 2
The posterior sample summaries of the parameters compared with the original values

	Simulation 1	Simulation 2	Simulation 3	Simulation 4
μ_1	−0.178 (−0.272, −0.084) 0	−0.141 (−0.222, −0.058) 0	−0.176 (−0.271, −0.084) 0	−0.139 (−0.221, −0.057) 0
μ_2	0.208 (0.114,0.302) 0	0.174 (0.079,0.270) 0	5.210 (5.114,5.303) 5	5.174 (5.079,5.270) 5
σ_1^2	0.954 (0.583,1.818) 1	0.764 (0.383,1.206) 1	0.996 (0.569,1.984) 1	0.685 (0.404,1.004) 1
σ_2^2	1.062 (0.586,1.960) 1	0.627 (0.335,1.003) 1	1.157 (0.750,2.014) 1	0.697 (0.417,1.105) 1
σ_3^2	1.112 (0.816,2.221) 1	4.321 (3.372,6.398) 4	0.885 (0.501, 1.404) 1	3.778 (2.962, 5.387) 4
τ_1^2	0.796 (0.632,1.019) 1	0.691 (0.439,1.139) 1	0.692 (0.498,1.009) 1	0.701 (0.449,1.196) 1
τ_2^2	1.247 (0.778,2.036) 1	0.801 (0.626,1.306) 1	0.771 (0.563,1.162) 1	0.767 (0.544,1.233) 1
ϕ_1	4.173 (2.274,5.553) 3	2.750 (1.882,6.173) 3	3.062 (2.414,4.419) 3	3.581 (2.077,7.322) 3
ψ_1	3.564 (2.117,7.407) 3	2.255 (1.747,3.632) 3	2.119 (1.732,3.306) 3	3.542 (2.276,4.772) 3
ϕ_2	2.681 (1.357,5.686) 1.5	2.724 (1.872,4.623) 3	1.513 (1.179,2.317) 1.5	2.018 (1.514,4.132) 3
ψ_2	2.293 (1.431,5.375) 1.5	2.458 (1.196,5.206) 3	1.427 (1.171,2.064) 1.5	2.958 (1.903,5.196) 3
ϕ_3	5.007 (2.685,12.372) 6	2.224 (1.398,4.045) 3	3.108 (2.198,7.096) 6	2.172 (1.676,4.465) 3
ψ_3	5.595 (2.782,9.139) 6	2.236 (1.628,4.177) 3	7.201 (3.073,14.191) 6	2.118(1.353,5.150) 3
t_0	4.787 (1.538,8.368) 4.614	4.987 (1.368,8.938) 4.614	4.614 (2.574,5.380) 4.614	4.614 (2.594,5.544) 4.614

extended to "*what if*" scenarios where we simulate under one change point model and fit under another. The foregoing caveats may be even more appropriate.

Alternative constructions could be offered; further investigation for a discrete time scale could be undertaken. However, in our limited experience, most critical is to have observations at a large number of locations and a large number of time points. But, working with process models this introduces dramatically increased computational burden associated with evaluation of the likelihood and/or the joint distribution of the spatial random effects.

Appendix A. Full conditional distribution theory for the additive model incorporating random noise

Here, we employ the full conditional distribution employed in the Gibbs sampling to fit this model. We suppress the variables being conditioned on.

$$\text{for } \mu_1, \mu_1 \sim N(\tilde{\mu_1}, \sigma^2_{\mu,1}) \text{ with } \sigma^2_{\mu,1} = \tau_1^2/M(t_0)N$$

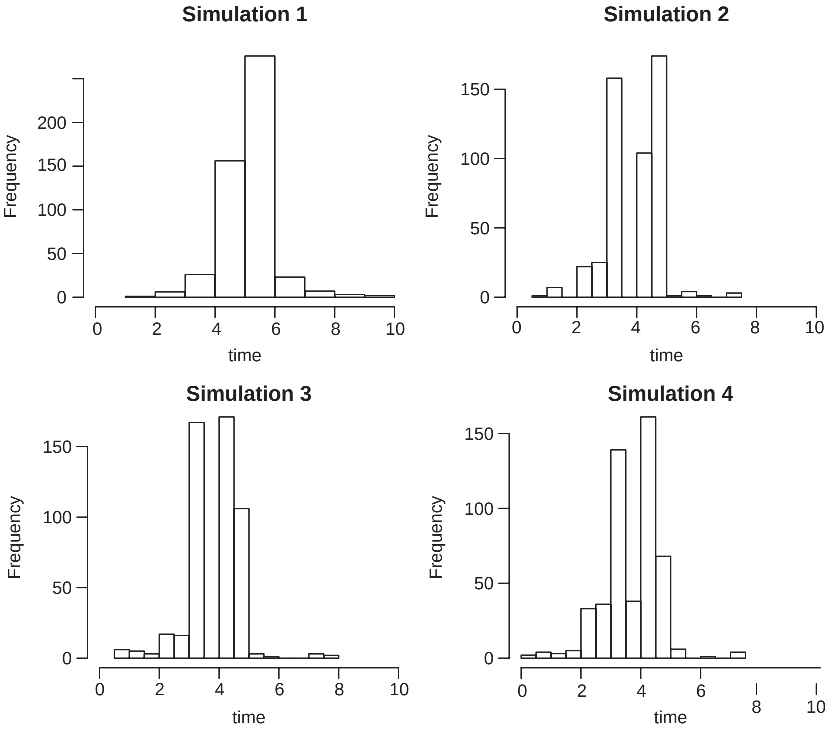

Fig. 1. Posterior histograms for t_0.

and

$$\tilde{\mu_1} = \sigma^2_{\mu,1} 1^{\mathrm{T}}(Y_{t^-} - U_{t^-} - W_{t^-})/\tau_1^2;$$

for μ_2, $\mu_2 \sim N(\tilde{\mu_2}, \sigma^2_{\mu,2})$ with $\sigma^2_{\mu,2} = \tau_2^2/(M - M(t_0))N$

and

$$\tilde{\mu_2} = \sigma^2_{\mu,2} 1^{\mathrm{T}}(Y_{t^+} - U_{t^+} - V_{t^+})/\tau_2^2;$$

for σ_1^2, $\sigma_1^2 \sim \mathrm{IG}\left(2 + \dfrac{MN}{2}, \beta_1 + U^{\mathrm{T}} R(\psi_1)^{-1} \otimes R(\phi_1)^{-1} U/2\right)$;

for σ_2^2, $\sigma_2^2 \sim \mathrm{IG}\left(2 + \dfrac{NM(t_0)}{2}, \beta_2 + W^{\mathrm{T}} R(\psi_2)^{-1} \otimes R(\phi_2)^{-1} W/2\right)$;

for σ_3^2, $\sigma_3^2 \sim \mathrm{IG}\left(2 + \dfrac{N(M - M(t_0))}{2}, \beta_3 + V^{\mathrm{T}} R(\psi_3)^{-1} \otimes R(\phi_3)^{-1} V/2\right)$;

for ψ_1, $f(\psi_1)$

$$\propto \exp\{-(U^{\mathrm{T}} R(\psi_1)^{-1} \otimes R(\phi_1)^{-1} U/2\sigma_1^2) - \psi_1 \beta_{\psi_1}\} \psi_1^{\alpha_{\psi_1 - 1}} / |R(\psi_1)|^{\frac{M}{2}};$$

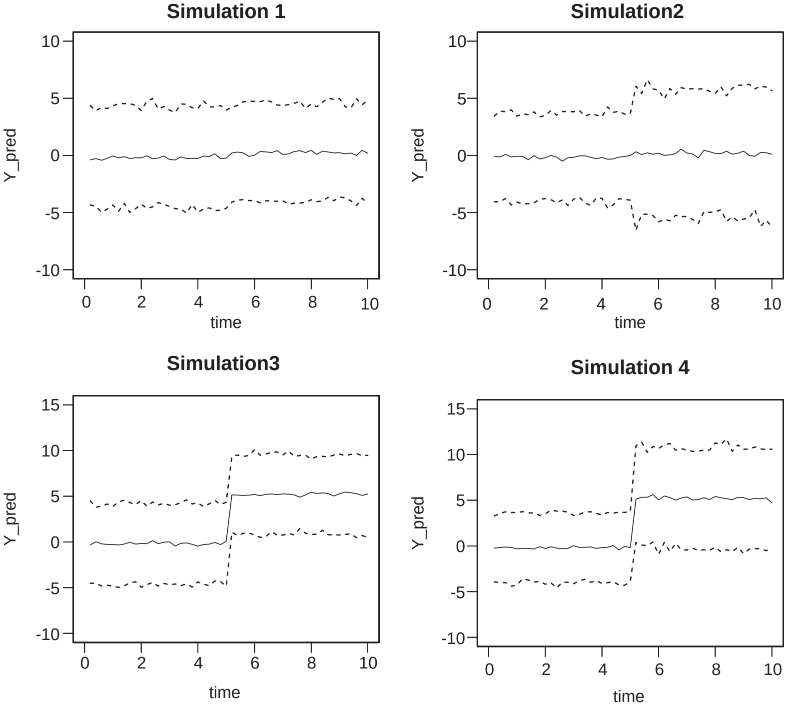

Fig. 2. Posterior predictive mean at location $\mathbf{s} = (5,5)$ and associated .95 predictive intervals across t for each simulation.

$$\text{for } \phi_1,\ f(\phi_1) \propto \exp\{-(U^{\mathrm{T}}R(\psi_1)^{-1} \otimes R(\phi_1)^{-1}U/2\sigma_1^2) - \phi_1\beta_{\phi_1}\}\phi_1^{\alpha_{\phi_1-1}}/|R(\phi_1)|^{\frac{N}{2}};$$

$$\text{for } \psi_2,\ f(\psi_2) \propto \exp\{-(W^{\mathrm{T}}R(\psi_2)^{-1} \otimes R(\phi_2)^{-1}W/2\sigma_2^2) - \psi_2\beta_{\psi_2}\}\psi_2^{\alpha_{\psi_2-1}}/|R(\psi_2)|^{\frac{M}{2}};$$

$$\text{for } \phi_2,\ f(\phi_2) \propto \exp\{-(W^{\mathrm{T}}R(\psi_1)^{-1} \otimes R(\phi_1)^{-1}W/2\sigma_2^2) - \phi_2\beta_{\phi_2}\}\phi_2^{\alpha_{\phi_2-1}}/|R(\phi_2)|^{\frac{N}{2}};$$

$$\text{for } \psi_3,\ f(\psi_3) \propto \exp\{-(V^{\mathrm{T}}R(\psi_3)^{-1} \otimes R(\phi_3)^{-1}V/2\sigma_3^2) - \psi_3\beta_{\psi_3}\}\psi_3^{\alpha_{\psi_3-1}}/|R(\psi_3)|^{\frac{(M)}{2}};$$

$$\text{for } \phi_3,\ f(\phi_3) \propto \exp\{-(V^{\mathrm{T}}R(\psi_3)^{-1} \otimes R(\phi_3)^{-1}V/2\sigma_3^2) - \phi_3\beta_{\phi_3}\}\phi_3^{\alpha_{\phi_3-1}}/|R(\phi_3)|^{\frac{N}{2}};$$

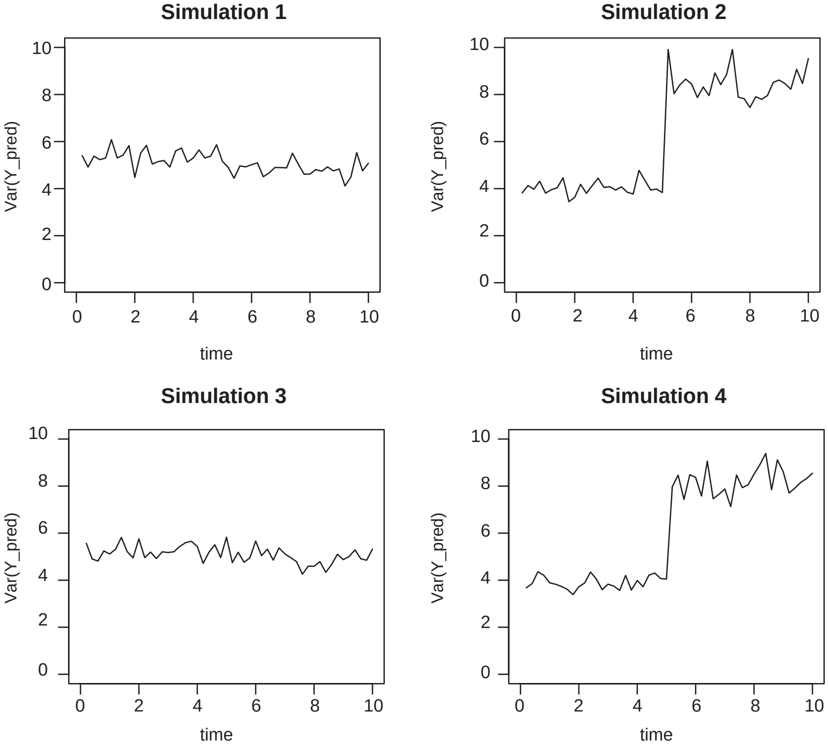

Fig. 3. Variance of posterior predictive distribution for an observation at location $\mathbf{s} = (5, 5)$ across t for each simulation.

$$\text{for } \tau_1^2, \tau_1^2 \sim \text{IG}\left(2 + \frac{M(t_0)N}{2}, \beta_4 + (Y_{t^-} - \mu_1\mathbf{1} - U_{t^-} - W_{t^-})^{\mathrm{T}} \times (Y_{t^-} - \mu_1\mathbf{1} - U_{t^-} - W_{t^-})/2\right);$$

$$\text{for } \tau_2^2, \tau_2^2 \sim \text{IG}\left(2 + \frac{(M - M(t_0))N}{2}, \beta_5 + (Y_{t^+} - \mu_2 1 - U_{t^+} - V_{t^+})^{\mathrm{T}} \times (Y_{t^+} - \mu_2\mathbf{1} - U_{t^+} - V_{t^+})/2\right);$$

for $U, U \sim N(\tilde{\mu_U}, \Sigma_U^\star)$ with

$$\Sigma_U^\star = (D + R(\psi_1)^{-1} \otimes R(\phi_1)^{-1})^{-1}/\sigma_1^2,$$

where D is a $MN \times MN$ diagonal matrix with the first $M(t_0)N$ diagonal entries $1/\tau_1^2$ and the next $(M - M(t_0))N$ diagonal entries $1/\tau_2^2$, i.e., D depends on τ_1^2, τ_2^2 and t_0.

$$\tilde{\mu_U} = \Sigma_U^\star((Y_{t^-} - \mu_1\mathbf{1} - U_{t^-} - W_{t^-})^{\mathrm{T}}/\tau_1^2, (Y_{t^+} - \mu_2 1 - U_{t^+} - V_{t^+})^{\mathrm{T}}/\tau_2^2)^{\mathrm{T}};$$

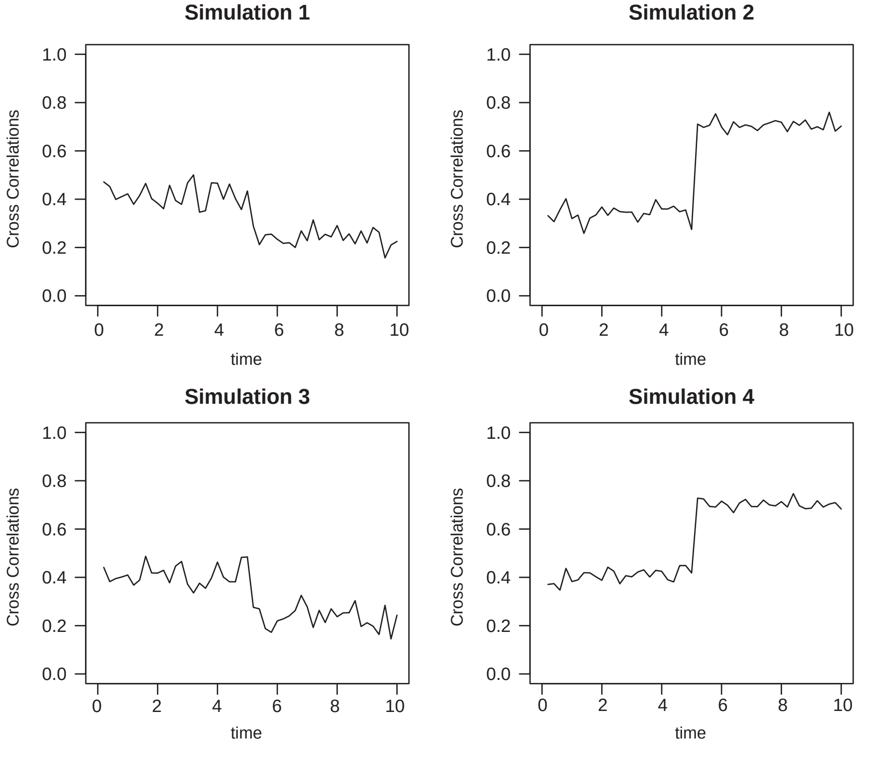

Fig. 4. Posterior correlation between $Y(\mathbf{s},t)$ and $Y(\mathbf{s}',t)$ for $\mathbf{s}=(5,5)$ and $\mathbf{s}'=(5.14,5.14)$, $t'-t=.1$, plotted vs. t.

$$\text{for } W,\ W \sim N(\tilde{\mu_W}, \Sigma_W^\star) \text{ with}$$

$$\Sigma_W^\star = (D + R(\psi_2)^{-1} \otimes R(\phi_2)^{-1}/\sigma_2^2)^{-1}$$

$$\tilde{\mu_W} = \Sigma_W^\star (Y_{M(t_0)} - \mu_1 \mathbf{1} - U_{M(t_0)} - W_{M(t_0)})/\tau_1^2;$$

$$\text{for } V,\ V \sim N(\tilde{\mu_V}, \Sigma_V^\star) \text{ with}$$

$$\Sigma_V^\star = (D + R(\psi_3)^{-1} \otimes R(\phi_3)^{-1}/\sigma_3^2)^{-1}$$

and

$$\tilde{\mu_V} = \Sigma_V^\star (Y_{t+} - \mu_2 \mathbf{1} - U_{t+} - V_{t+})/\tau_2^2.$$

Appendix B. The distribution theory for the additive model without random noise

The full conditional distributions for the parameters arising under (10) are as follows:

$$\text{for } \mu_1,\ \mu_2 \sim N(\tilde{\mu_1}, \sigma_{\mu,1}^2) \text{ with } \sigma_{\mu,1}^2 = \sigma_2^2/(\mathbf{1}^{\mathrm{T}} R(\psi_2)^{-1} \otimes R(\phi_2)^{-1} \mathbf{1})$$

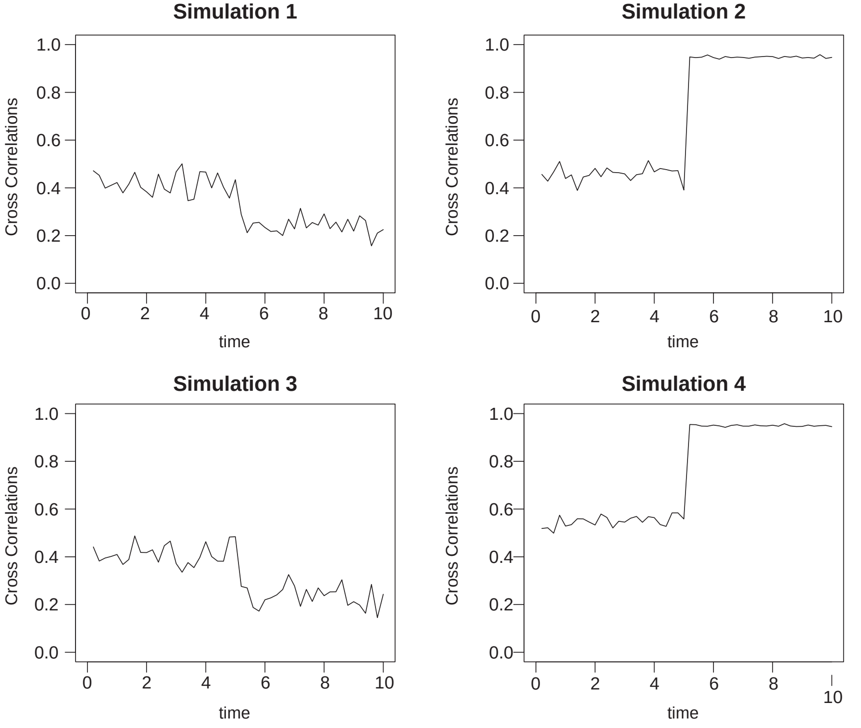

Fig. 5. Posterior correlation between $Y(\mathbf{s}, t)$ and $Y(\mathbf{s}', t)$ for $\mathbf{s} = (5, 5)$ and $\mathbf{s}' = (5.071, 5.071)$, $t' - t = .1$, plotted vs. t.

and

$$\tilde{\mu_1} = \sigma^2_{\mu,1} 1^{\mathrm{T}} R(\psi_2)^{-1} \otimes R(\phi_2)^{-1}(Y_{t^-} - U_{t^-})/\sigma_2^2;$$

for μ_2, $\mu_2 \sim N(\tilde{\mu_2}, \sigma^2_{\mu,2})$ with $\sigma^2_{\mu,2} = \sigma_3^2/(\mathbf{1}^{\mathrm{T}} R(\psi_3)^{-1} \otimes R(\phi_3)^{-1}\mathbf{1})$;

and

$$\tilde{\mu_2} = \sigma^2_{\mu,2} 1^{\mathrm{T}} R(\psi_3)^{-1} \otimes R(\phi_3)^{-1}(Y_{t^+} - U_{t^+})/\sigma_3^2;$$

for σ_1^2, $\sigma_1^2 \sim \mathrm{IG}\left(2 + \dfrac{MN}{2}, \beta_1 + U^{\mathrm{T}} R(\psi_1)^{-1} \otimes R(\phi_1)^{-1} U/2\right)$;

for σ_2^2, σ_2^2

$$\sim \mathrm{IG}(2 + M(t_0)N/2, \beta_2 + (Y_{t^-} - \mu_1\mathbf{1} - U_{t^-})^{\mathrm{T}} R(\psi_2)^{-1} \otimes R(\phi_2)^{-1} \times (Y_{t^-} - \mu_1\mathbf{1} - U_{t^-})/2);$$

for σ_3^2, σ_3^2

$$\sim \mathrm{IG}(2 + (M - M(t_0))N/2, \beta_3 + (Y_{t^+} - \mu_2\mathbf{1} - U_{t^+})^{\mathrm{T}} R(\psi_3)^{-1} \otimes R(\phi_3)^{-1} \times (Y_{t^+} - \mu_2\mathbf{1} - U_{t^+})/2);$$

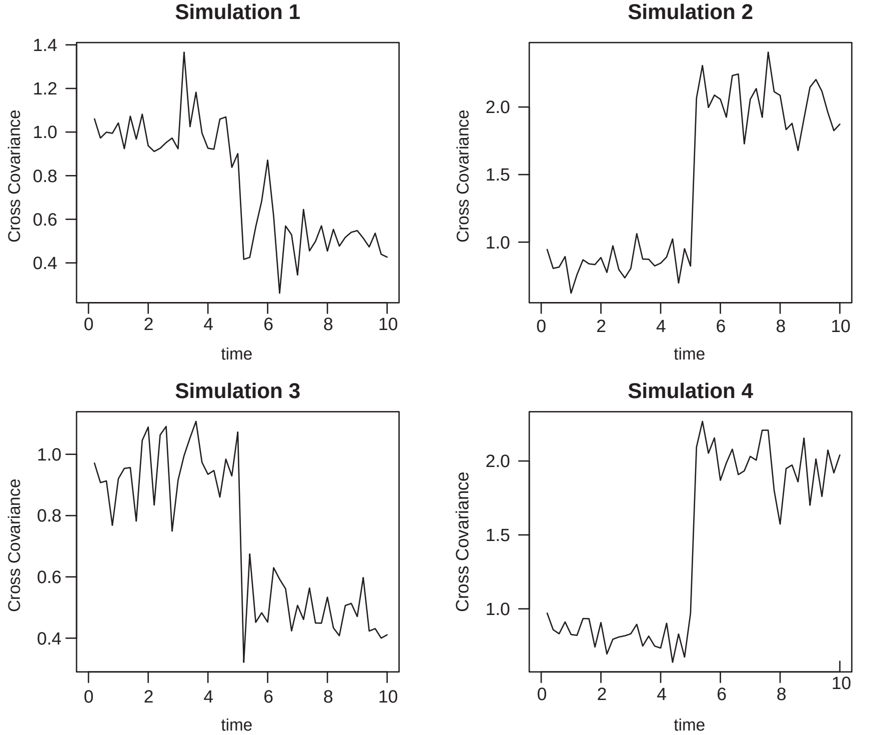

Fig. 6. Posterior covariance between $Y(\mathbf{s},t)$ and $Y(\mathbf{s}',t)$ for $\mathbf{s}=(5,5)$ and $\mathbf{s}'=(5.14,5.14)$, $t'-t=.1$, plotted vs. t.

$$\text{for } \psi_1,\ f(\psi_1) \propto \exp\{-(U^{\mathrm{T}}R(\psi_1)^{-1}\otimes R(\phi_1)^{-1}U/2\sigma_1^2) - \psi_1\beta_{\psi_1}\} \times \psi_1^{\alpha_{\psi_1-1}}/|R(\psi_1)|^{\frac{M}{2}};$$

$$\text{for } \phi_1,\ f(\phi_1) \propto \exp\{-(U^{\mathrm{T}}R(\psi_1)^{-1}\otimes R(\phi_1)^{-1}U/2\sigma_1^2) - \phi_1\beta_{\phi_1}\} \times \phi_1^{\alpha_{\phi_1-1}}/|R(\phi_1)|^{\frac{N}{2}};$$

$$\text{for } \psi_2,\ f(\psi_2) \propto \exp\{-((Y-\mu_1\mathbf{1}-U)^{\mathrm{T}}R(\psi_2)^{-1}\otimes R(\phi_2)^{-1} \times (Y-\mu_1\mathbf{1}-U)/2\sigma_2^2) - \psi_2\beta_{\psi_2}\}\psi_2^{\alpha_{\psi_2-1}}/|R(\psi_2)|^{\frac{M(t_0)}{2}};$$

$$\text{for } \phi_2,\ f(\phi_2) \propto \exp\{-((Y_{t^-}-\mu_1\mathbf{1}-U_{t^-})^{\mathrm{T}}R(\psi_2)^{-1}\otimes R(\phi_2)^{-1} \times (Y_{t^-}-\mu_1\mathbf{1}-U_{t^-})/2\sigma_2^2) - \phi_2\beta_{\phi_2}\}\phi_2^{\alpha_{\phi_2-1}}/|R(\phi_2)|^{\frac{N}{2}};$$

$$\text{for } \psi_3,\ f(\psi_3) \propto \exp\{-((Y_{t^+}-\mu_2\mathbf{1}-U_{t^+})^{\mathrm{T}}R(\psi_3)^{-1}\otimes R(\phi_3)^{-1} \times (Y_{t^+}-\mu_2 1-U_{t^+})/2\sigma_3^2) - \psi_3\beta_{\psi_3}\}\psi_3^{\alpha_{\psi_3-1}}/|R(\psi_3)|^{\frac{M-M(t_0)}{2}};$$

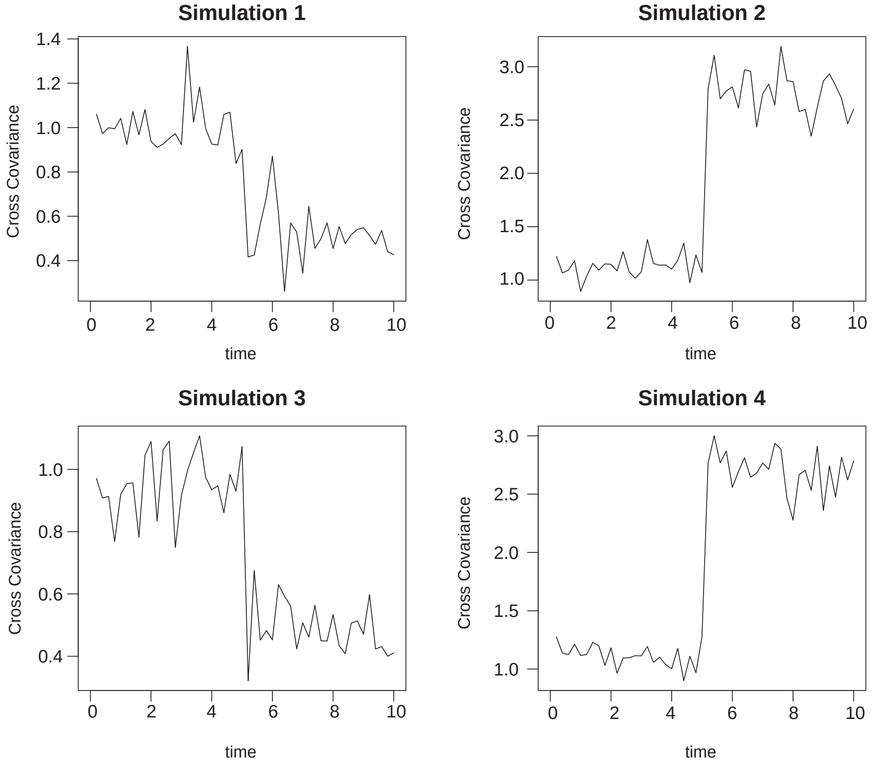

Fig. 7. Posterior covariance between $Y(\mathbf{s}, t)$ and $Y(\mathbf{s}', t)$ for $\mathbf{s} = (5, 5)$ and $\mathbf{s}' = (5.071, 5.071)$, $t' - t = .1$, plotted vs. t.

$$\text{for } \phi_3,\ f(\phi_3) \propto \exp\{-((Y_{t^+} - \mu_2\mathbf{1} - U_{t^+})^{\mathrm{T}} R(\psi_3)^{-1} \otimes R(\phi_3)^{-1} \times (Y_{t^+} - \mu_2\mathbf{1} - U_{t^+})/2\sigma_3^2) - \phi_3\beta_{\phi_3}\}\phi_3^{\alpha_{\phi_3 - 1}}/|R(\phi_3)|^{\frac{N}{2}}.$$

for U, $U \sim N(\mu\tilde{U}, \Sigma_U^\star)$ with

$$\Sigma_U^\star = (\tilde{R} + R(\psi_1)^{-1} \otimes R(\phi_1)^{-1})^{-1}/\sigma_1^2,$$

where $\tilde{R}$ is a $MN \times MN$ block diagonal matrix with the first $M(t_0)N \times M(t_0)N$ diagonal block $R(\psi_2)^{-1} \otimes R(\phi_2)^{-1}$ and the next $(M - M(t_0))N \times (M - M(t_0))N$ diagonal block $R(\psi_3)^{-1} \otimes R(\phi_3)^{-1}$.

Finally, $\mu\tilde{U} = \Sigma_U^\star(R(\psi_2)^{-1} \otimes R(\phi_2)^{-1}(Y_{t^-} - \mu_1\mathbf{1} - U_{t^-} - W_{t^-})^{\mathrm{T}}, R(\psi_3)^{-1} \otimes R(\phi_3)^{-1}(Y_{t^+} - \mu_2 1 - U_{t^+} - V_{t^+})^{\mathrm{T}})^{\mathrm{T}}$.

References

Banerjee, S., Gelfand, A.E., 2002. Prediction, interpolation and regression for spatially misaligned data points. Sankhya 64, 227–245.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. second ed. Springer, New York.
Carlin, B.P., Gelfand, A.E., Smith, A.F.M., 1992. Hierarchical Bayesian analysis of change-point problems. Appl. Statist. 389–405.
Chen, R., Tsay, R.S., 1993. Functional-coefficient autoregressive models. J. Amer. Statist. Assoc. 88, 298–308.
Chernoff, H., Zacks, S., 1964. Estimating the current mean of a normal distribution which is subjected to change in time. Ann. Math. Statist. 35, 999–1018.
Cressie, N.A.C., Huang, H.C., 1999. Classes of nonseparable spatiotemporal stationary covariance functions. J. Amer. Statist. Assoc. 94, 1330–1340.
Csorgö, M., Horváth, L., 1988. Nonparametric methods for change-point problems. Krishnaiah, P.R., Rao, C.R. (Eds.), Handbook of Statistics, vol. 7. Elsevier, Amsterdam, pp. 403–425.
Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model based Geostatistics (with discussion). Appl. Statist. 47, 299–350.
Ecker, M.D., Isakson, H.R., 2004. A unified convex-concave model of urban land. Regional Sci. Urban Econom., to appear.
Gneiting, T., 2002. Nonseparable stationary covariance functions for space–time data. J. Amer. Statist. Assoc. 97, 590–600.
Jacquier, E., Polson, N.E., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models (with discussion). J. Business Econom. Statist. 12, 371–419.
Kim, S.N., Shephard, N., Chen, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Rev. Econom. Stud. 65, 361–393.
Krishnaiah, P.R., Miao, B.Q., 1988. Review about estimation of change-points. Krishnaiah, P.R., Rao, C.R. (Eds.), Handbook of Statistics, vol. 7. Elsevier, Amsterdam, pp. 375–402.
Mardia, K.V., Goodall, C., 1993. Spatiotemporal analyses of multivariate environmental monitoring data. Patil, G.P., Rao, C.R. (Eds.), Multivariate Environmental Statistics. Elsevier, Amsterdam, pp. 347–386.
Perreault, L., Bernier, J., Bobee, B., Parent, E., 2000. Bayesian change-point analysis in hydrometerological time series. Part 1. The normal model revisited. J. Hydrol. 235, 221–241.
Raftery, A.E., 1994. Change point and change curve modeling in stochastic processes and spatial statistics. Appl. Statist. Sci. 1, 403–424.
Raftery, A.E., Akman, V.E., 1986. Bayesian analysis of a Poisson process with a change point. Biometrika 73, 85–89.
Smith, A.F.M., 1975. A Bayesian approach to inference about a change-point in a sequence of random variables. Biometrika 62, 407–416.
Stein, M.L., 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.
Taylor, S.J., 1994. Modeling stochastic volatility: a comparative study. Math. Finance 4, 183–204.
Tong, H., 1983. Threshold Models in Nonlinear Time Series Analysis. Lecture Notes in Statistics, vol. 21. Springer, New York.
Zacks, S., 1983. Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. Recent Advances in Statistics: Herman Chernoff Festschrift. Academic Press, New York, pp. 245–269.