A Kernel-Based Spectral Model for Non-Gaussian Spatio-Temporal Processes

Christopher K. Wikle*

Department of Statistics, University of Missouri

Accepted: *Statistical Modelling: An International Journal*

July 26, 2002

*Christopher K. Wikle, Department of Statistics, University of Missouri, 222 Math Science Building, Columbia, MO 65211; wikle@stat.missouri.edu

Abstract

Spatio-temporal processes can often be written as hierarchical state-space processes. In situations with complicated dynamics such as wave propagation, it is difficult to parameterize state transition functions for high-dimensional state processes. Although in some cases prior understanding of the physical process can be used to formulate models for the state transition, this is not always possible. Alternatively, for processes where one considers discrete time and continuous space, complicated dynamics can be modeled by stochastic integro-difference equations in which the associated redistribution kernel is allowed to vary with space and/or time. By considering a spectral implementation of such models, one can formulate a spatio-temporal model with relatively few parameters that can accommodate complicated dynamics. This approach can be developed in a hierarchical framework for non-Gaussian processes, as demonstrated on cloud intensity data.

Key Words: Bayesian, dilation, dynamic models, hierarchical, integro-difference equations, translation

1 Introduction

There has been much interest in recent years in modeling spatio-temporal processes in the environmental and physical sciences. Methods have considered geostatistical approaches (see the review by Kyriakidis and Journel, 1999), multivariate time series approaches (e.g., Bennett 1979), space-time dynamic model approaches (e.g., Goodall and Mardia, 1994; Guttorp, Meiring, and Sampson, 1994; Mardia et al., 1998; Meiring, Guttorp, and Sampson, 1998; Wikle and Cressie, 1999), and hierarchical approaches (e.g., Wikle, Berliner and Cressie, 1998; Berliner, Wikle and Cressie, 2000; Wikle et al. 2001).

The interest here is in processes that have coherent dynamical interactions such as exhibited by geophysical or ecological processes with wave behavior. One approach to accounting for realistic dynamic structure in such complicated spatio-temporal settings is to consider underlying explicit theoretical relationships. For example, in an atmospheric/oceanic context one has well-specified deterministic relationships (partial differential equations, PDEs) that describe, to an extent, the process of interest. Such information can be used in traditional state-space settings as demonstrated in the "data assimilation" literature in atmospheric/ocean science (e.g., Ghil et al. 1981). Alternatively, one can consider the physics as "approximate" and actually use the PDEs as a framework for developing prior distributions in a hierarchical model (Royle et al., 1999; Wikle et al., 2001; Wikle 2002a). This approach works very well when one has an understanding of the underlying physical relationships from which to develop the PDEs and thus, the priors. However, it is often the case that the scientific knowledge for a specific problem is not well-developed and one does not have easily-described physically-based priors. In that case, one may still need efficient methodologies that can model dynamical behavior with relatively few parameters.

Wikle (2001) suggested that a kernel-based spectral approach based on stochastic integro-difference equations can be used to model complicated dynamical processes. In particular, if a hierarchical formulation is considered, this approach can easily allow the redistribution kernel to vary with space and/or time. This methodology makes use of recent developments in nonstationary spatial modeling using convolution kernels (e.g., Higdon, 1998), as well as theoretical results concerning deterministic integro-difference equations and their utility in modeling dynamical processes (e.g., Kot et al. 1996). In particular, the kernels are used to define the propagator (or transition) matrix, which allows the state variables to evolve differently in different parts of the spatial domain. Non-Gaussian observations are incorporated by conditioning on a continuous latent variable, which is assumed to follow this non-stationary spatio-temporal process. Parameter and state estimation is carried out using MCMC methods. The model has wide applicability for several reasons. First, it allows for arbitrary observation types. Second, it easily handles missing data. Third, it scales to high dimensional problems, since it uses spectral decompositions of the spatial kernels and the state process. Finally, it provides a flexible and interpretable kernel representation for the underlying state process. Our purpose in this paper is to show how one can use this approach in a hierarchical Bayesian setting to model non-Gaussian spatio-temporal dynamic processes. The methodology is demonstrated on a cloud intensity process.

2 Background

Although an active area of research, it is extremely difficult to specify realistic covariance models for complicated spatio-temporal processes, at least from a joint-distribution perspective (Cressie and Huang, 1999; Gneiting, 2002). However, if one considers such processes from a hierarchical perspective, a series of relatively simple conditional models can lead to complicated joint models. As described in Wikle et al. (1998), the general hierarchical framework for spatio-temporal models consists of three general stages:

$$\text{(Stage 1) Data model: } [data|process, parameters],$$

$$\text{(Stage 2) Process model: } [process|parameters],$$

$$\text{(Stage 3) Parameter Model: } [parameters],$$

where we use the bracket notation to represent a distribution. Given such a framework, one can get the posterior distribution $[process, parameters|data]$ through Bayes' rule. Each stage can be further factored into series of conditional models. For example, dynamical models are often considered from a state-space perspective, and the state process is factored in a Markovian fashion in the second stage. Parameters associated with the dynamics are then given distributions in the last stage.

As shown by Diggle, Tawn, and Moyeed (1998) one can use this hierarchical framework to consider non-Gaussian spatial models as well. In that case, the non-Gaussian data process is conditional on some latent spatial process, which might be modeled as a Gaussian random field at the second stage. The parameters of this spatial model are then given distributions at the third stage. Wikle (2002b) has shown that by considering the spatial process from a spectral point of view, this approach can be implemented very efficiently for high-dimensional non-Gaussian data. Furthermore, Wikle (2002a) has shown that this approach can be considered for non-Gaussian spatio-temporal processes as well. In that study, ecological abundance (count) data were assumed to be Poisson, conditional on a spatio-temporal intensity process. This intensity process was assumed to be log-normal, with a spatio-temporal dynamical component. In that application, the underlying process was diffusive (i.e., the process modeled was that of an invasive species, showing heterogeneous diffusion and exponential growth over time) and a PDE model provided the basis for the dynamical evolution. However, if one could not assume that the process was diffusive *a priori* then an arbitrary dynamical evolution model would have to be considered. In that case, the dynamical propagator matrix (e.g., transition matrix) would be extremely high-dimensional, and difficult to estimate, even from a hierarchical perspective. Alternatively, one could consider a stochastic integro-difference equation-based methodology in this framework. Such an approach for non-Gaussian spatio-temporal processes is described below. The procedure is illustrated on cloud intensity data from a regional climate model.

2.1 Scientific Problem and Data

The propagation of clouds is a complicated nonlinear function of various atmospheric state processes. In fact, cloud dynamics are still the subject of intense research in the atmospheric science community as cloud parameterizations are a fundamental component of atmospheric General Circulation Models (GCM's) used to study climate and weather. Our interest here is whether the dynamics of such cloud processes can be modeled adequately by a non-Gaussian hierarchical spatio-temporal model. To examine this, we obtained cloud intensity information from a regional climate model as discussed in Pan et al. (2001).

We consider the data $Z_t(s_i)$ at spatial locations s_i, $i = 1, \ldots, n$ and time $t = 1, \ldots, T$. Specifically, we selected $n = 60$ and $T = 80$. The spatial locations are evenly spaced at a resolution of 52 km and represent a 1-D spatial (longitudinal) domain over the central U.S. The temporal sampling is every 6 hours and is representative of the large-scale meteorological forcing in late March 1979. The data values are cloud water content in grams water per kilogram of air (g/kg). The data have non-negative integer support at each spatial location and time. The ultimate goal for considering these "data" is to develop efficient parameterizations of cloud behavior in regional climate models. However, for the purposes of this study, we simply wish to establish that the kernel-based integro-difference equation spatio-temporal methodology can capture the essential dynamics in regional climate model processes. In addition, we wish to demonstrate that the methodology can work well in situations where much of the data are unobservable (for example, as one might experience with polar orbiting satellite observations of clouds, e.g., Wikle et al. 2001). Thus, for this study, we remove 40% of the data by randomly selecting from the 60 spatial locations at each time, assuming the data are missing from all times at the selected pixels. The goal is to model the cloud intensity (cloud water content) and thus predict the cloud water content over the whole domain, including the locations for which data are missing. This will provide validation for the model's performance.

3 A Non-Gaussian Hierarchical Spatio-Temporal Dynamic Model

First, we describe the kernel-based spectral model for spatio-temporal dynamical processes. We then describe how this can be used in a hierarchical framework to model the cloud intensity process.

3.1 Stochastic Integro-Difference Equation Model

Consider the stochastic integro-difference equation (IDE) for an underlying spatio-temporal process $y_t(s)$ which in general is assumed to be continuous in space and discrete in time:

$$y_{t+1}(s) = \gamma \int k_s(r; \boldsymbol{\theta}_s) y_t(r) dr + \tilde{\eta}_{t+1}(s), \tag{3.1}$$

where $k_s(r; \boldsymbol{\theta}_s)$ is the redistribution kernel that describes how the process at time t is redistributed in space at time $t+1$, $\boldsymbol{\theta}_s$ are parameters of the redistribution kernel (that may be spatially dependent), $\tilde{\eta}$ is a spatially-colored noise process that is independent across time, and the parameter γ is used in this context to control (and allow for) explosive growth. Note that there is a substantial literature on deterministic integro-difference equations, particularly related to dispersal of ecological processes (e.g., Kot et. al 1996). Stochastic versions similar to that presented here were considered by Wikle and Cressie

(1999), Brown et al. (2000), and Brown et al. (2001). However, motivated by the non-stationary spatial modeling approach of Higdon (1998), Wikle (2001) showed that by considering spatially varying (heterogeneous) redistribution kernels in this framework, one can model very complicated dynamics, both diffusive and so-called “extra-diffusive”. Such an approach is the focus here.

3.2 Extra-Diffusive Dynamics

It is well-recognized in the ecology literature that the deterministic IDE framework can accommodate diffusive dynamics, and that the behavior of the dynamics is determined from the kernel specification (e.g., Kot et al. 1996). We propose that the method is significantly more powerful in that it can model more complicated dynamical behavior, which we call extra-diffusive propagation. Specifically, we are interested in propagation of spatial features through time. For illustration, consider the one-dimensional Gaussian spatial kernel,

$$k_s(r, \theta_1, \theta_2) = \frac{1}{\theta_2 \sqrt{2\pi}} \exp\{-.5(r - \theta_1 - s)^2 \theta_2\} \tag{3.2}$$

where the kernel is centered at $\theta_1 + s$ and thus is shifted by θ_1 spatial units relative to location s, and θ_2 is the scale parameter. We refer to θ_1 as the translation parameter and θ_2 as the dilation parameter, analogous to the usual translation and dilation in the description of wavelet basis functions. In the IDE kernel context, these parameters influence the dynamical evolution of the y process.

Figure 1 shows a successive integration of (3.1) with a 2-D Gaussian kernel analogous to (3.2), with the kernel translated to the left and down relative to s, identically for all s. In this case the disturbance shows diffusive propagation (i.e., spreading with time) but the center of the disturbance also propagates to the right and up, with speed proportional to the translation distance. As the spread (dilation) of the kernel increases, the process becomes more diffusive; as the translation of the kernel increases, the center of the process propagates more rapidly.

More importantly, consider a spatially-varying kernel, in which the translation parameters are allowed to vary with space. Such models have recently been considered by Higdon (1998) for spatial problems in which a convolution of white noise is used to generate nonstationary spatial covariance models. In the IDE setting, such spatially-varying (heterogeneous) kernels can capture more complicated dynamics than homogeneous kernels. For example, Figure 2 shows arrows that indicate the propagation direction at each location implied by slowly spatially-varying translation parameters. That is, the kernel is translated in the opposite direction (relative to s) shown by the arrows. Figure 3 shows the resulting propagation of a disturbance; note how the disturbance propagates in a quasi-circular fashion.

The simulation shown in Figure 3 suggests that the stochastic IDE methodology with heterogeneous kernels has the potential to accommodate quite complicated dynamical processes. For example, as will be shown below, this method can model the propagation of cloud intensity in strongly dynamic environments. More generally, the procedure could forecast the propagation of coherent radar reflectivities and thus could be used in short-term forecasting of precipitation for severe weather or hydrological purposes. In addition, the procedure could be applied to the problem of predicting the spread of invasive species across heterogeneous landscapes. In general, any dynamical system for which there is substantial lack of certainty as to the underlying deterministic dynamics can be modeled by this approach.

3.3 Dimension Reduction Through Spectral Representation

Most spatio-temporal processes of interest in the geophysical and ecological sciences have very high-dimensional state (spatial) processes. As is the case with traditional state-space models, the stochastic IDE approach is difficult to implement in these settings. Wikle et al. (2001) showed that implementation is greatly facilitated in PDE-based hierarchical spatio-temporal models if the state-process is formulated spectrally.

A similar approach can be used for IDE-based models. This is consistent with other recent work demonstrating the flexibility of spectral approaches in nonstationary spatial modeling (e.g., Nychka et al. 1999; Nychka et al. 2002; Fuentes 2002).

First, expand the kernel and the process in terms of spectral basis functions $\phi_i(s)$:

$$k_s(r;\boldsymbol{\theta}_s) = \sum_i b_i(s;\boldsymbol{\theta}_s)\phi_i(r), \tag{3.3}$$

$$y_t(s) = \sum_j \alpha_j(t)\phi_j(s), \tag{3.4}$$

where the basis functions are complete and orthonormal. If the process of interest has non-trivial dynamics, then the redistribution kernel has significant spread and thus can be represented as the linear combination of a relatively small set of spectral basis functions. In that case, the sum in (3.3) is truncated at I and upon substitution of (3.3) and (3.4) into (3.1), we get

$$y_{t+1}(s) = \gamma \mathbf{b}'(s;\boldsymbol{\theta}_s)\boldsymbol{\alpha}_t^{(1)} + \tilde{\eta}_{t+1}(s) \tag{3.5}$$

where $\mathbf{b}(s;\boldsymbol{\theta}_s) \equiv [b_1(s;\boldsymbol{\theta}_s)\ldots b_I(s;\boldsymbol{\theta}_s)]'$ and $\boldsymbol{\alpha}_t^{(1)} \equiv [\alpha_1(t)\ldots\alpha_I(t)]'$. So, for spatial locations $\{s_1,\ldots,s_n\}$,

$$\mathbf{y}_{t+1} = \gamma \mathbf{B}'_{\boldsymbol{\theta}}\boldsymbol{\alpha}_t^{(1)} + \tilde{\boldsymbol{\eta}}_{t+1}$$

where $\mathbf{y}_{t+1} \equiv [y_{t+1}(s_1)\ldots y_{t+1}(s_n)]'$ and $\mathbf{B}_{\boldsymbol{\theta}} = [\mathbf{b}(s_1;\boldsymbol{\theta}_{s_1})\ldots\mathbf{b}(s_n;\boldsymbol{\theta}_{s_n})]$. Thus, since from (3.4), $\mathbf{y}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\alpha}_{t+1}$, it follows that:

$$\boldsymbol{\alpha}_{t+1}^{(1)} = \boldsymbol{\Phi}'_{(1)}\mathbf{B}'_{\boldsymbol{\theta}}\boldsymbol{\alpha}_t^{(1)} + \boldsymbol{\eta}_{t+1}^{(1)} \tag{3.6}$$

$$\boldsymbol{\alpha}_{t+1}^{(2)} = \boldsymbol{\Phi}'_{(2)}\mathbf{B}'_{\boldsymbol{\theta}}\boldsymbol{\alpha}_t^{(1)} + \boldsymbol{\eta}_{t+1}^{(2)}, \tag{3.7}$$

where $\boldsymbol{\alpha}_t = [\boldsymbol{\alpha}_t^{(1)'}\ \boldsymbol{\alpha}_t^{(2)'}]'$, $\boldsymbol{\alpha}_t^{(2)} \equiv [\alpha_{I+1}(t)\ \ldots\ \alpha_n(t)]'$, $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_{(1)}\ \boldsymbol{\Phi}_{(2)}]$, with $\boldsymbol{\Phi}_{(1)} \equiv [\boldsymbol{\phi}_1\ \ldots\ \boldsymbol{\phi}_I]$, and $\boldsymbol{\Phi}_{(2)} \equiv [\boldsymbol{\phi}_{I+1}\ \ldots\ \boldsymbol{\phi}_n]$. If we assume that $\tilde{\boldsymbol{\eta}}_t \sim N(\mathbf{0},\mathbf{C}_{\tilde{\eta}})$, then $\boldsymbol{\eta}_t^{(1)} \sim N(\mathbf{0},\mathbf{C}_\eta^{(1)})$, $\boldsymbol{\eta}_t^{(2)} \sim N(\mathbf{0},\mathbf{C}_\eta^{(2)})$, where $\mathbf{C}_\eta^{(j)} \equiv \boldsymbol{\Phi}'_{(j)}\mathbf{C}_{\tilde{\eta}}\boldsymbol{\Phi}_{(j)}$ for $j = 1, 2$.

Several comments are in order. First, the spatial locations $\{s_1, \ldots, s_n\}$ need not correspond to data locations if there is a data model that conditions on the process $\mathbf{y}_t$. Furthermore, we note that the evolution of $\boldsymbol{\alpha}^{(2)}$ in (3.7) depends on the past value of the $\boldsymbol{\alpha}^{(1)}$ process rather than the past values of the $\boldsymbol{\alpha}^{(2)}$ process. Depending on the dimension reduction (i.e., the spread of the kernel) and the underlying process, one might assume that $\boldsymbol{\alpha}^{(2)}$ is a non-dynamic spatio-temporal component without much loss of predictive power (e.g., Wikle and Cressie 1999).

3.3.1 Kernel Representation

The spectral-based IDE methodology outlined above does not depend explicitly on the functional form of the kernel, with the exception of the assumption that the kernel can be modeled reasonably well as a finite sum of orthogonal basis functions. Indeed, since the final model formulation is in spectral space, an estimation procedure need only find estimates for the elements of $\mathbf{B}$ (where we drop the θ subscript in this case). One may consider this problem using moment-based estimators in a Kalman filter framework as in Wikle and Cressie (1999). Likelihood estimates obtained from E-M algorithm approaches can be used as well, provided the number of parameters is reasonably small (e.g., Shumway and Stoffer, 1982).

A parsimonious representation of $\mathbf{B}_{\boldsymbol{\theta}}$ can be obtained if one considers a specific parametric kernel function. This approach has the additional advantage that one can control the dynamics by altering the translation (shift) and dilation (spread) of the kernel in terms of a relatively few number of parameters. Although easily implemented in a two-dimensional spatial setting, for ease of illustration we focus on the one-dimensional spatial case here.

Consider the Gaussian kernel given by (3.2). Let $\phi_i(s)$ be Fourier basis functions. Of course, the Fourier transform of the Gaussian kernel is just its characteristic function and

$$
\begin{aligned}
b_j(s;\theta_1(s),\theta_2(s)) &= \exp\{i\omega_j(\theta_1(s)+s) - .5\omega_j^2\theta_2(s)\} \\
&= \cos\{\omega_j(\theta_1(s)+s)\}\exp\{-.5\omega_j^2\theta_2(s)\} \\
&\quad +i\sin\{\omega_j(\theta_1(s)+s)\}\exp\{-.5\omega_j^2\theta_2(s)\},
\end{aligned} \tag{3.8}
$$

where ω_j is the spatial frequency. Thus, the real and imaginary coefficients of the characteristic function correspond to cosine and sine Fourier basis functions, respectively. These coefficients are completely determined if we know the kernel parameters $\theta_1(s)$ and $\theta_2(s)$ at each spatial location s. That is, we need to specify the spatial fields $\boldsymbol{\theta}_1 = [\theta_1(s_1)\ldots\theta_1(s_n)]'$ and $\boldsymbol{\theta}_2 = [\theta_2(s_1)\ldots\theta_2(s_n)]'$. For example, we might assume $\boldsymbol{\theta}_1 \sim N(\boldsymbol{\mu}_1, \mathbf{C}_{\theta_1})$ and $\boldsymbol{\theta}_2 \sim LN(\boldsymbol{\mu}_2, \mathbf{C}_{\theta_2})$, where LN refers to a log-normal distribution. Since we expect the dynamics to be relatively slowly varying over space, the spatial structure in these fields should be relatively simple.

Thus, the model described by (3.6) and (3.7), along with the choice of a Gaussian kernel and Fourier basis functions, provide a hierarchical formulation for the spatio-temporal dynamic process from the IDE perspective. This corresponds to the second level of the hierarchical framework above. Furthermore, if we specify distributions for the kernel dilation and translation parameters, they correspond to the third stage in the general hierarchy. These can be combined with an appropriate data model to complete the hierarchy.

The methodology outlined here is not limited to Fourier basis functions. Any orthogonal set of basis functions could be considered (e.g., wavelets, empirical orthogonal functions). The relative advantages and disadvantages of these alternative basis functions will be investigated elsewhere. It is clear, however, that the use of Fourier basis functions has the direct benefit that the Fourier transform of distribution-based kernels is known analytically (i.e., it is the characteristic function of the distribution).

The methodology is not limited to Gaussian kernels. In fact, the only limitation is that the one must be able to easily obtain the analytical Fourier transform of the kernel. Previous research suggests the advantage of certain "heavy-tail" kernels in the deterministic IDE framework for cases where diffusion is the only dynamical process of interest (e.g., Kot et al., 1996). However, in cases where the IDE is stochastic and extra-diffusive dynamics are of primary interest, there doesn't seem to be much advantage to having kernels with heavy tails. However, note that by shifting the kernels in the extra-diffusive case, one effectively is considering skewed kernels.

3.4 Complete Hierarchical Model

The regional climate model cloud data described previously has non-negative integer support. Thus, similar to the non-Gaussian spatial modeling approach of Diggle et al. (1998) and the spatio-temporal approach of Wikle (2002a), we assume that conditional on a Poisson intensity process at all spatial and temporal locations of interest, the data are distributed as independent Poisson random variables,

$$Z_t(s_i)|\boldsymbol{\lambda}_t \sim Poi(\mathbf{k}'_{i,t}\boldsymbol{\lambda}_t) \tag{3.9}$$

where for all i and t in the domain of interest, $Z_t(s_i)$ is the cloud intensity at location s_i and time t, $\boldsymbol{\lambda}_t \equiv [\lambda_t(s_1), \ldots, \lambda_t(s_n)]'$ is the Poisson intensity process at all spatial locations for time t, and $\mathbf{k}_{i,t}$ is an incidence vector indicating whether a prediction location has an associated observation. To simplify notation, let $\mathbf{u}_t \equiv \log(\boldsymbol{\lambda}_t)$ and assume

$$\mathbf{u}_t|\mu, \nu, \mathbf{y}_t, \sigma_\epsilon^2 \sim N(\mu\mathbf{1} + \nu\boldsymbol{\Phi}\boldsymbol{\alpha}_t, \sigma_\epsilon^2\mathbf{I}), \tag{3.10}$$

where μ is the overall mean effect, ν is a scaling parameter, and σ_ϵ^2 represents extra-Poisson variability. We then make use of the model (3.6) and (3.7),

$$\begin{aligned}\boldsymbol{\alpha}_t^{(1)}|\boldsymbol{\alpha}_{t-1}^{(1)}, \mathbf{B}_\theta, \mathbf{C}_\eta^{(1)} &\sim N(\boldsymbol{\Phi}_{(1)}'\mathbf{B}_\theta'\boldsymbol{\alpha}_{t-1}^{(1)}, \mathbf{C}_\eta^{(1)}) \\ \boldsymbol{\alpha}_t^{(2)}|\boldsymbol{\alpha}_{t-1}^{(1)}, \mathbf{B}_\theta, \mathbf{C}_\eta^{(2)} &\sim N(\boldsymbol{\Phi}_{(2)}'\mathbf{B}_\theta'\boldsymbol{\alpha}_{t-1}^{(1)}, \mathbf{C}_\eta^{(2)}).\end{aligned}$$

Note that we have assumed $\gamma = 1$ in this example. One of our primary interests with these data is to determine the overall tendency of the spatially-varying translation parameter. Thus, we assume $\boldsymbol{\theta}_1$ is relatively smooth and let $\boldsymbol{\theta}_1 = \boldsymbol{\Psi}\mathbf{f}$, with $\mathbf{f} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Psi}$ are the first p ($p = 5$) eigenvectors of an exponential correlation matrix with relatively strong spatial dependence ($c_\theta(h) = \exp(-h/30)$ for distances $h = 0, \ldots, 60$) and $\boldsymbol{\Sigma}$ is the associated diagonal matrix. We also assume that $\log(\boldsymbol{\theta}_2) = \boldsymbol{\Psi}\mathbf{g}$, with $\mathbf{g} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Furthermore, we let $\mathbf{C}_\eta^{(j)} = \boldsymbol{\Phi}_{(j)}'\mathbf{C}_\eta\boldsymbol{\Phi}_{(j)}$ where $\mathbf{C}_\eta$ is a Gaussian covariance matrix with fixed parameters ($c_\eta(h) = \exp(-h^2/200)$ for $h = 0, \ldots, 60$). We also specify uniform priors for $\mu \sim Unif[-10, 10]$ and $\nu \sim Unif[.5, 5]$ and let $\sigma_\epsilon^2 \sim IG(q_\epsilon, r_\epsilon)$, where $q_\epsilon = 3$ and $r_\epsilon = 5$. Finally, we specify a prior distribution for the initial state for the $\boldsymbol{\alpha}^{(1)}$ process; we let $\boldsymbol{\alpha}_0^{(1)} \sim N(\mathbf{0}, .05\mathbf{I})$.

Implementation was via a Gibbs sampler, utilizing straightforward conjugate updating with the exception of $\mathbf{u}_t$, $\mathbf{f}$, and $\mathbf{g}$, which were updated via Metropolis-Hastings steps. In particular, $\mathbf{u}_t$ was updated individually for each spatial location and time by a random-walk Metropolis-Hastings algorithm with random walk variance equal to 0.2. Similarly, $\mathbf{f}$ and $\mathbf{g}$ were updated by random walk Metropolis-Hastings steps with random walk variances of 0.3. The spectral dynamic process $\boldsymbol{\alpha}_t^{(1)}$ for $t = 0, \ldots, T$ was updated by its conjugate multivariate normal full-conditionals; similarly for $\boldsymbol{\alpha}_t^{(2)}$, $t = 1, \ldots, T$. The "regression" parameters μ, ν were updated jointly by their truncated Gaussian full-conditional distributions. Finally, σ_ϵ^2 was updated by its conjugate inverse gamma full-conditional distribution. Samples of $\mathbf{y}_t$ were obtained by simply multiplying the samples of $\boldsymbol{\alpha}_t$ by the spectral basis matrix $\boldsymbol{\Phi}$ (i.e., taking the inverse Fourier transform of $\boldsymbol{\alpha}_t$).

The Gibbs sampler was run for 10000 iterations after a 5000 iteration burn-in. Standard errors were computed by batching, to account for the correlation in the Markov chain. The results described below were not overly sensitive to the choices of the the fixed parameters in the aforementioned prior distributions.

3.5 Results

Figure 4 shows the posterior mean of the intensities $\boldsymbol{\lambda}_t$ as well as the "data" $\mathbf{Z}_t$ and the climate model truth. These plots only consider 1-D space and are interpreted by considering the x-axis as the spatial axis (e.g., "longitude") and the y-axis as the time axis, such that time increases from top to bottom. Thus, if one sees a diagonal stripe slanted to the right, it suggests propagation to the right (or east since the x-axis represents longitude); similarly, a left-slanted stripe suggests propagation to the left (west).

The model is able to fill in the missing information in a dynamically reasonable fashion. In addition, the posterior mean image is smoother in the sense that the information has been "blurred" a bit. Although this can be controlled via the dilation parameter, it is a common (and often desirable) feature of the stochastic IDE model (e.g., Brown et al. 2000). The posterior mean of the underlying Gaussian spatio-temporal process $\mathbf{y}_t$ (where $\mathbf{y}_t = \boldsymbol{\Phi}\boldsymbol{\alpha}_t$) is shown in Figure 5 along with the associated standard deviation and the intensity truth. This latent process captures the eastward propagation of the clouds. In addition, it is able to capture the spatially-heterogeneous dynamics suggested in the truth field. Specifically, visual inspection of the truth suggests that the speed of eastward propagation is less in the western and eastern portion of the domain than in the central portion. Figure 6 shows the posterior mean of the $\boldsymbol{\theta}_1$ (translation parameters) and $\boldsymbol{\theta}_2$ (dilation parameter) spatial processes. The posterior mean for the translation parameter confirms that the dynamical propagation is to the right over the entire domain (recall that a negative translation parameter suggests propagation to the right), but is slower over the eastern and western portions of the domain, as suggested visually from the truth fields.

Correspondingly, the dilation parameter is larger over the eastern and western portions of the domain than in the center, implying that the diffusive aspect of the process is greater in the eastern and western portions of the domain as well.

4 Discussion

We have presented a spatio-temporal hierarchical model for a non-Gaussian process that is based on a stochastic integro-difference equation with heterogeneous redistribution kernels. By letting the kernels exhibit spatially-varying translation and dilation, complicated non-separable and non-homogeneous dynamics can be modeled. This is analogous to the use of such kernel-based methods in the modeling of non-stationary spatial processes (e.g., Higdon 1998). The methodology was demonstrated on cloud intensity data from a regional climate model.

Although the modeling approach works well for the cloud water content data, one can imagine that over longer time-spans the kernels should change with time as well. For example, the dynamics in the mid-latitudes are different in the spring than in the summer due to the annual migration of the jet stream and associated semi-permanent high and low pressure systems. Thus, the kernel translation and dilation should change accordingly. As discussed in Wikle (2001), the hierarchical modeling approach outlined here can accommodate time-varying kernels. That is, in addition to letting θ_1 and θ_2 vary spatially, we allow them to vary temporally as well (e.g., $\boldsymbol{\theta}_1(t)$ and $\boldsymbol{\theta}_2(t)$). In this case, the model (3.6) becomes $\boldsymbol{\alpha}_{t+1}^{(1)} = \boldsymbol{\Phi}'_{(1)}\mathbf{B}'_{\theta_t}\boldsymbol{\alpha}_t^{(1)} + \boldsymbol{\eta}_{t+1}$. The θ processes might be modeled using one of the recently developed classes of space-time covariance functions (Cressie and Huang 1999, Gneiting 2002), if the dimensionality is not an issue. Alternatively, one might consider spatio-temporal dynamical models for these parameters. However, one must be careful in this setting not to simply replace one complicated spatio-temporal problem by another of equal (or greater) complexity! Typically, one would expect that the spatio-temporal dynamics of the parameters are substantially less complicated than the dynamics of the original process. In that case, dimensionality might be reduced by another spectral decomposition of these parameter processes.

We note that one might allow the kernels to be influenced by other variables and processes. In the present application, this suggests a possible approach to dynamic parameterization of clouds in climate models, where other atmospheric variables dictate the likely spatio-temporal distribution of the kernel parameters. This will be explored elsewhere. In general, such ideas can be extended to other applications in which the physical dimensionality is large and yet prior knowledge of explicit dynamical relationships is weak. Potential applications include predicting the spread of invasive species in Ecology, forecasting radar and satellite imagery, and the prediction of functional MRI imagery.

Acknowledgments

This research was made possible by a grant from the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) program, Assistance Agreement No. R827257-01-0. The author would like to thank Dr. Z. Pan for providing the cloud data and anonymous reviewers for helpful suggestions on an early draft.

References

Bennett, R.J., (1979). *Spatial Time Series*. London, Pion Limited.

Berliner, L.M. (1996), "Hierarchical Bayesian time series models," K. M. Hanson and R. N. Silver (Eds.) *Maximum Entropy and Bayesian Methods,* 15-22, Kluwer Academic Publishers.

Berliner, L.M., Wikle, C.K., and Cressie, N., (2000). Long-lead prediction of Pacific

SSTs via Bayesian dynamic modeling. *Journal of Climate*, **13**, 3953-3968.

Brown, P.E., Karesen, K.F., Roberts, G.O., and Tonellato, S., (2000). Blur-generated non-separable space-time models. *J.R. Statistical Society, Series B*, **62**, 847-860.

Brown, P.E., Diggle, P.J., Lord, M.E., and Young, P.C., (2001). Space-time calibration of radar rainfall data. *Applied Statistics*, **50**, 221-241.

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

Diggle P.J., Tawn J.A., and Moyeed, R.A. (1998). Model-based geostatistics (with discussion). *Applied Statistics* **47**, 299-350.

Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika*, **89**, 197-210.

Ghil, M., Cohn, S.E., Tavantzis, J., Bube, K. and Isaacson, E., (1981). Applications of estimation theory to numerical weather prediction. *Dynamic meteorology: Data assimilation methods*, L. Bengtsson, M. Ghi l, and E. Kallen, eds. New York: Springer-Verlag, 139-224.

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

Guttorp, P., Meiring, W. and Sampson, P.D., (1994). A space-time analysis of ground-level ozone data. *Environmetrics* **5**, 241-254.

Higdon, D., (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. **5**, 173-190.

Kot, M., Lewis, M.A., and van den Driessche, P., (1996). Dispersal data and the spread of invading organisms. *Ecology*, **77**, 2027-2042.

Kyriakidis, P.C., and Journel, A.G., (1999). Geostatistical space-time models: A review. *Mathematical Geology*, **31**, 651-684.

Mardia, K.V., Goodall, C.R., Redfern, E.J. and Alonso, F.J. (1998). The Kriged Kalman filter (with discussion). *Test*, **7**, 217-285.

Meiring, W., Guttorp, P. & Sampson, P.D. (1998). Space-time estimation of grid-cell hourly ozone levels for assessment of a deterministic model. *Envir. Ecol. Statist.* **5**, 197-222.

Neubert, M., Kot, M., and Lewis, M. (1995). Dispersal and pattern formation in a discrete-time predator-prey model. *Theoretical Population Biology*, **48**, 7-43.

Nychka, D., Wikle, C.K., and Royle, J.A. (1999). Large spatial prediction problems and nonstationary random fields. Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, CO, research report (http://www.cgd.ucar.edu/stats/manuscripts/krig5.ps).

Nychka, D., Wikle, C.K., and Royle, J.A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling: An International Journal*, to appear.

Pan, Z., Christensen, J.H., Arritt, R.W., Gutowski, W.J., Jr., Takle, E.S., and Otieno, F., (2001). Evaluation of uncertainties in regional climate change simulations. Submitted to *Journal of Geophysical Research*.

Royle, J.A., Berliner, L.M., Wikle, C.K., and Milliff, R. (1999). A hierarchical spatial model for constructing wind fields from scatterometer data in the Labrador sea. *Case Studies in Bayesian Statistics IV* , Springer-Verlag, 367-382.

Shumway, R.H. and Stoffer, D. (1982). An approach to time series smoothing and forecasting using the EM algorithm, *Journal of Time Series Analysis*, **3**, 253-264.

Wikle, C.K. (2001). A kernel-based spectral approach for spatio-temporal dynamic models. *Proceedings of the 1st Spanish Workshop on Spatio-Temporal Modelling of Environmental Processes (METMA)*, Benicassim, Castellon (Spain), 28-31 October 2001, pp. 167-180.

Wikle, C.K. (2002a) Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, to appear.

Wikle, C.K. (2002b) Spatial modeling of count data: A case study in modeling breeding bird survey data on large spatial domains. In *Spatial Cluster Modelling*, A. Lawson and D. Denison, eds. Chapman and Hall, 199-209.

Wikle, C.K., Berliner, L.M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, **5**, 117-154.

Wikle, C.K. and Cressie, N. (1999). A dimension reduced approach to space-time Kalman filtering. *Biometrika* , **86** , 815-829.

Wikle, C.K., Milliff, R.F., Nychka, D., and Berliner, L.M. (2001). Spatio-temporal hierarchical Bayesian Modeling: Tropical ocean surface wind data. *Journal of the American Statistical Association*, **96**, 382-397.

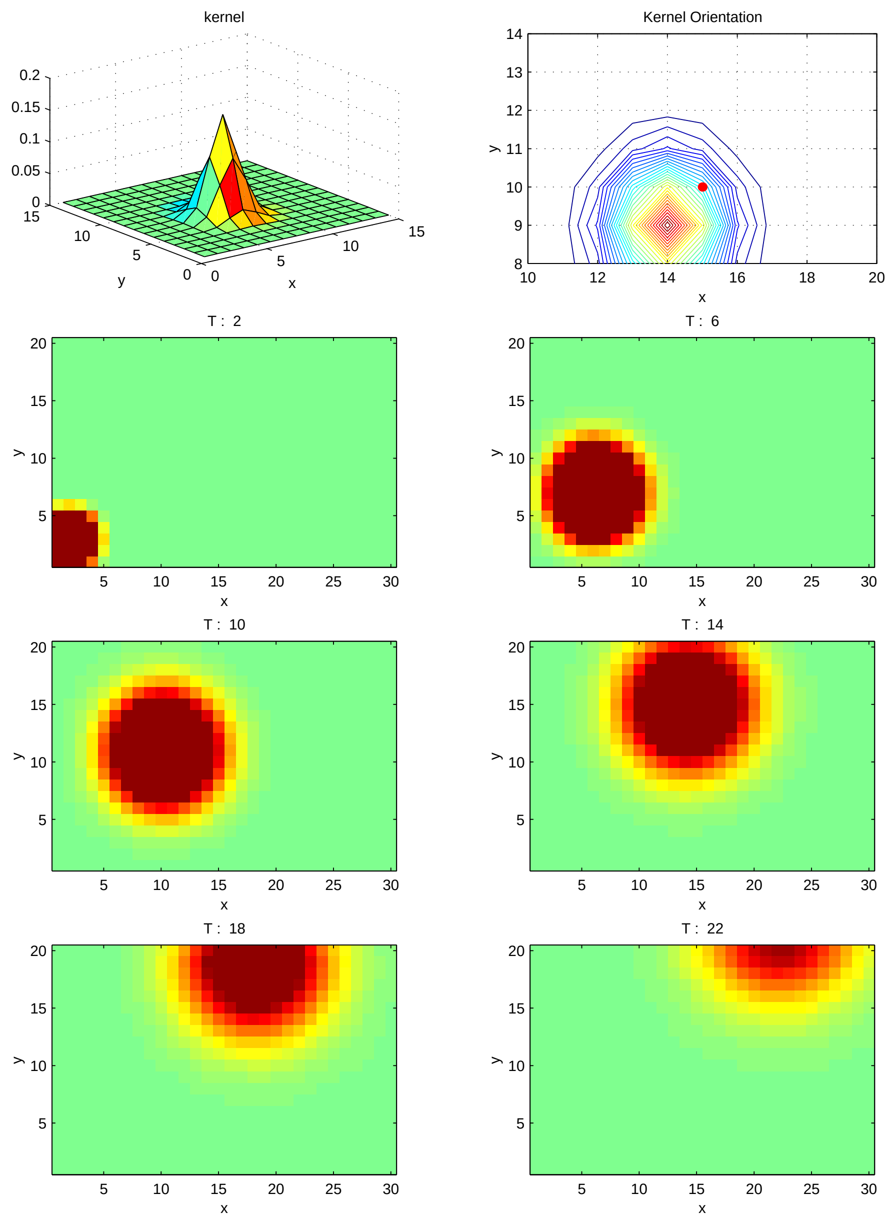

Figure 1: Simulation using a kernel in 2-d space with kernel translated to the left and down, implying diffusion and propagation up and to the right.

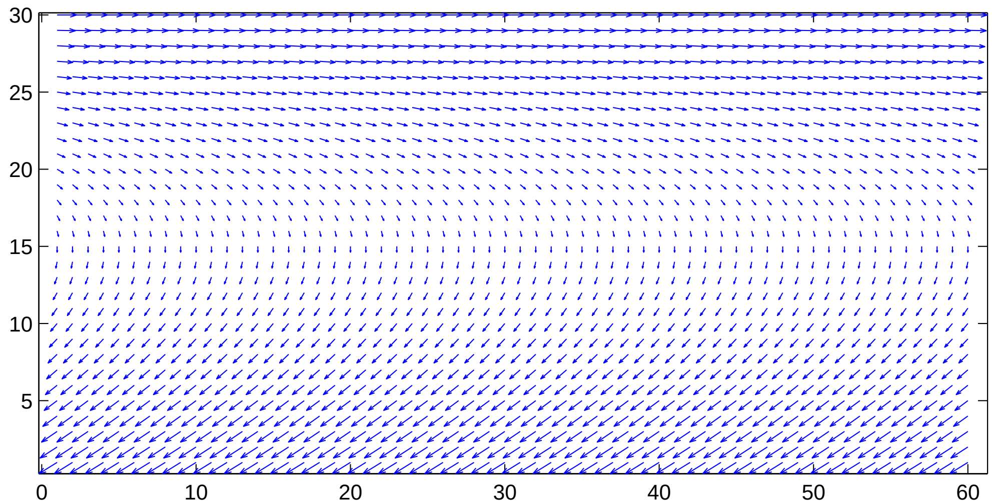

Figure 2: Vectors indicating spatial variation in translation parameter in a 2-d spatial kernel. Note that the arrow points in the direction in which the propagation is suggested by the translation (i.e., actual translation vector is in the opposite direction as shown).

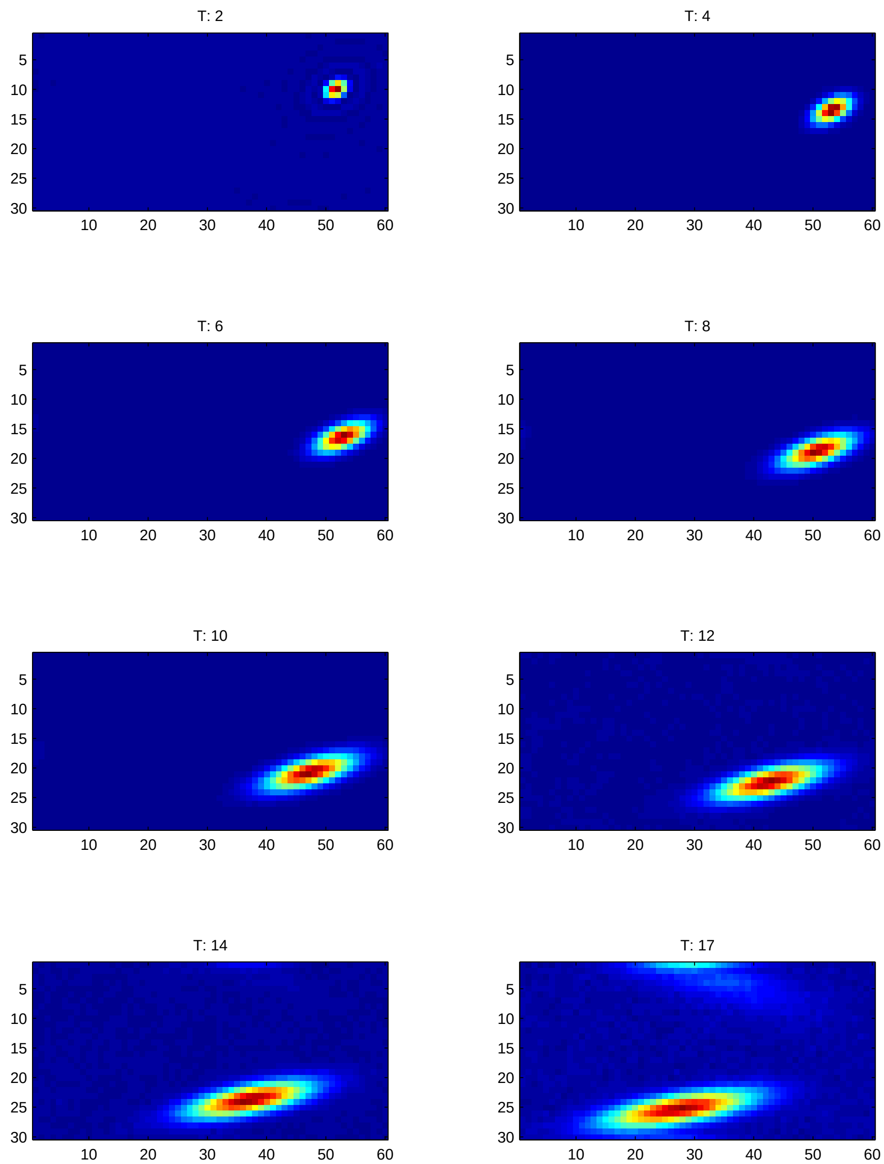

Figure 3: Propagation and diffusion suggested by the spatially-varying translation parameters shown in Figure 2.

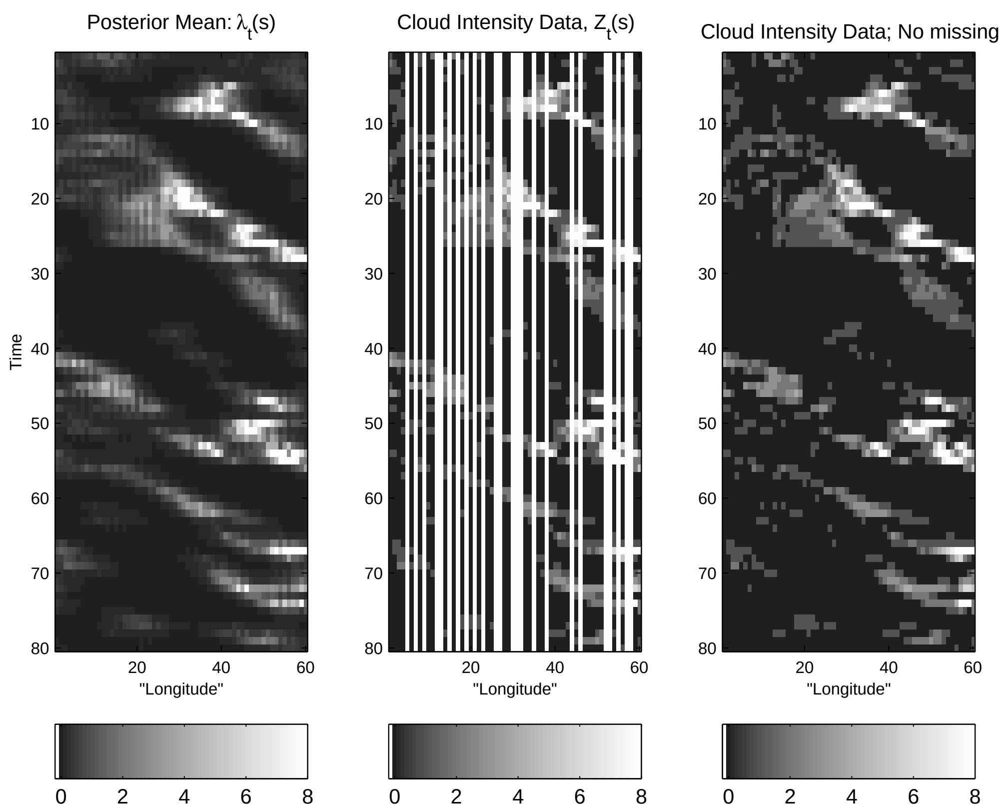

Figure 4: Left Panel: Posterior mean of $\boldsymbol{\lambda}_t$; Center Panel: Cloud intensity "Data" in which 40% of the spatial locations have randomly been chosen to have missing observations (indicated by the vertical white lines); Right Panel: Cloud intensity "truth" (i.e., no missing data) from the regional climate model simulation.

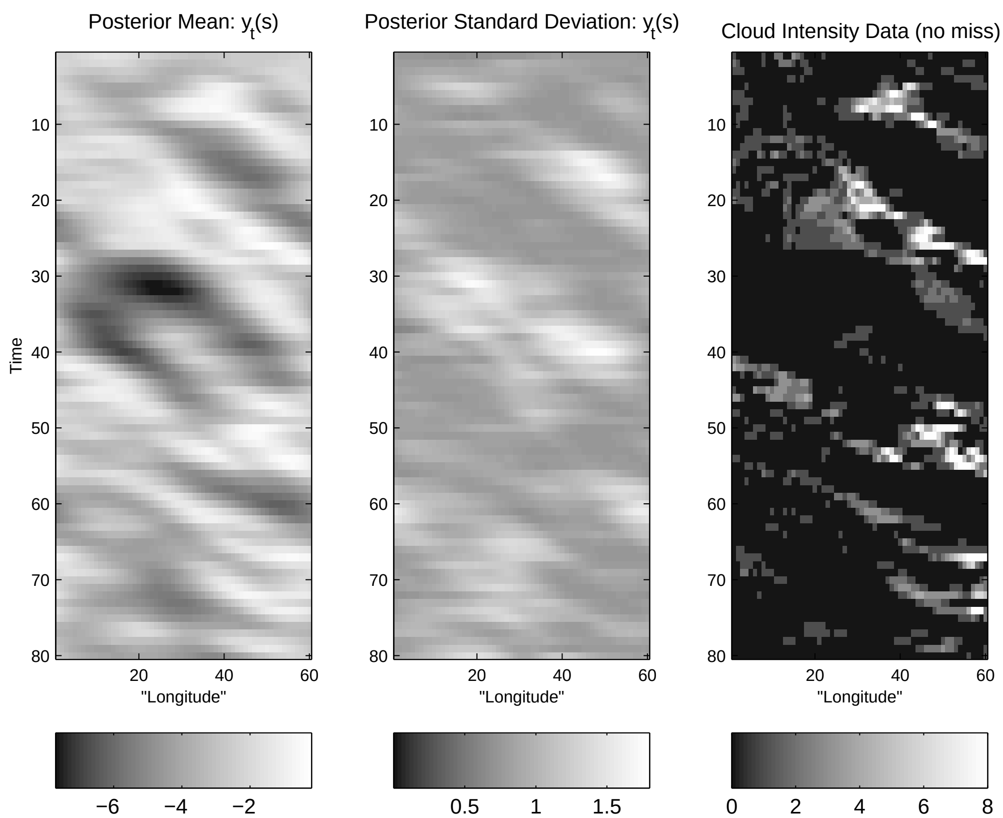

Figure 5: Left Panel: Posterior mean of $\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{\alpha}_t$; Center Panel: Posterior standard deviation for $\boldsymbol{y}_t$ process ; Right Panel: Cloud intensity "truth".

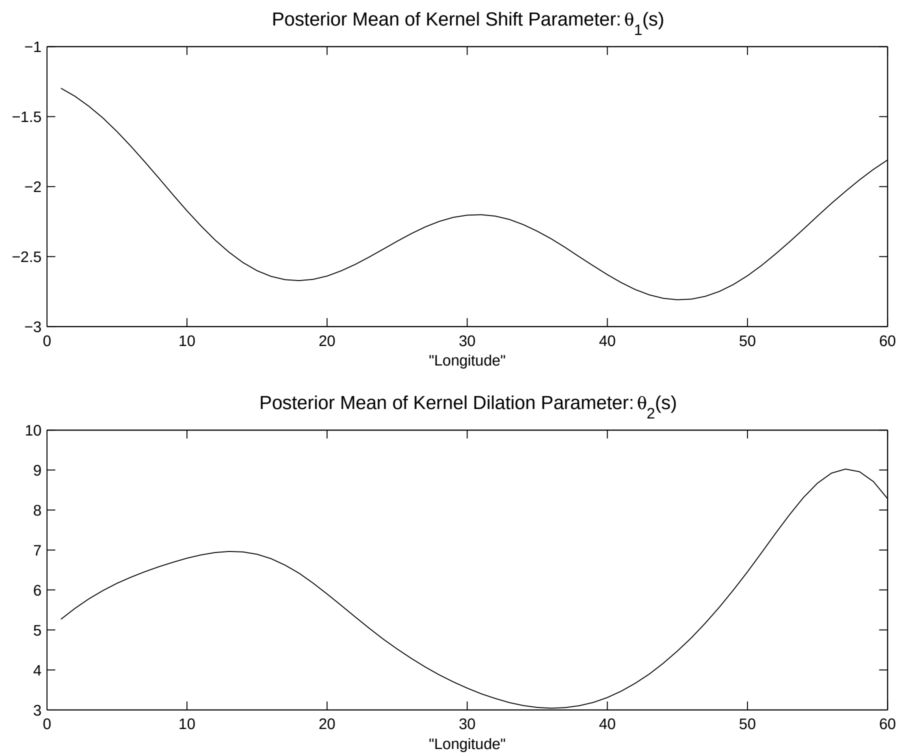

Figure 6: Top Panel: Posterior mean of $\boldsymbol{\theta}_1$, the kernel translation parameter. Bottom Panel: Posterior mean of $\boldsymbol{\theta}_2$, the kernel dilation parameter.