# Dynamic Models for Spatio-Temporal Data

Jonathan R. Stroud †

*Department of Statistics, University of Chicago, Chicago, IL 60637, U.S.A.*

Peter Müller

*Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, U.S.A.*

Bruno Sansó

*Centro de Estadística y Software Matemático, Universidad Simón Bolívar, Apartado 89000, Caracas 1080-A, Venezuela.*

**Summary**. We propose a model for nonstationary spatio-temporal data. To account for spatial variability, we model the mean function at each time period as a locally-weighted mixture of linear regressions. To incorporate temporal variation, we allow the regression coefficients to change through time. Kriging-based variogram models can also be also incorporated when the process is highly localized in space. The model is cast in a general state-space framework which allows us to include additional temporal components such as trends, seasonal effects, and autoregressions, and permits fast implementation and full probabilistic inference for the parameters, interpolations and forecasts. To illustrate our method, we apply it to two large datasets from the physical sciences: quarterly rainfall levels in Venezuela and water temperatures in the Atlantic Ocean.

*Key Words:* Bayesian inference; Filtering; Kriging; Locally-weighted mixture; Smoothing; Spatio-temporal models; State-space models.

## 1. Introduction

Spatio-temporal models have gained widespread popularity in recent years. One reason for the growth in this area is an abundance of applications in the environmental and health sciences. Typical examples include regional ozone monitoring, disease mapping, and analysis of satellite data. Increased computational power is another reason for this recent surge. Space-time datasets are often large and therefore require substantial computing resources to fit even simple models. However, few existing methods appear to be both computationally efficient and general enough for routine use. In this paper, we present a computationally efficient method which is widely applicable, including to problems with irregular, non lattice-based data, and observation locations changing over time. An important additional strength of the proposed method is that it naturally accommodates time series structures like seasonality and autoregressive components.

Models for spatio-temporal data are often constructed by combining time series models with variogram-based models from spatial statistics. In the time series context, popular approaches include ARMA models (Box, Jenkins and Reinsel, 1994) for stationary data, and state-space models

†*Address for correspondence:* Jonathan R. Stroud, Department of Statistics, University of Chicago, 5734 S. University Ave., Chicago, IL 60637, U.S.A. E-mail: **stroud@galton.uchicago.edu**

(SSMs; West and Harrison, 1997), which allow for nonstationary components such as temporal trends and seasonality. In the spatial setting, much of the literature revolves around isotropic models (Cressie, 1993). These models grew out of applications in geostatistics, where the main objective is prediction or *kriging*; thus we refer to them as kriging-based models. Various methods also exist for nonstationary spatial processes, where the correlation depends on location as well as distance. An example of this is Sampson and Guttorp (1992), who developed approaches based on transformations of the coordinate system.

Early statistical models for space-time data often relied on the assumption of temporal stationarity. For example, the STARMA (Pfeifer and Deutsch, 1980a,b) and STARMAX (Stoffer, 1986) models were constructed by adding spatial covariance matrices to standard vector ARMA models. Haslett and Raftery (1989) used similar models, but relied on fractional differencing to address long-memory features in Irish wind power data. Handcock and Wallis (1994) developed a model with kriging and long-memory dependence to study global warming trends in the Northern United States. Another related approach was taken by Carroll et al. (1997), who estimated ground-level ozone through Gaussian random fields assuming separable space and time correlation functions.

Recently several models have been proposed for nonstationarity and spatial anisotropy in spatio-temporal data. For example, Guttorp, Meiring and Sampson (1994) used the deformation technique of Sampson and Guttorp (1992) to capture spatial anisotropy in a temporally prewhitened series of ozone readings. Higdon (1999) relied on a Bayesian process convolution approach to model temporal and spatial nonstationarity in North Atlantic ocean temperatures. Lavine and Lozier (1999) proposed a Bayesian model for ocean temperatures using a three-dimensional Markov random field. Smith and Robinson (1997) proposed a Gibbs sampling approach for estimating rainfall in the presence of storms, and Dellaportas, Denison and Mallick (1998) developed space-time models for wind speed. Waller et al. (1997) used a hierarchical Bayesian model with CAR priors for spatial effects for mapping of Ohio lung cancer rates. Other approaches involving hierarchical Bayesian models include Wikle et al. (1999), who analyzed monthly maximum atmospheric temperatures, and Gelfand et al. (1998) who estimated a model for real estate prices.

Another modeling approach for time series is through the state-space framework. Sansó and Guenni (2000) assumed a parametric covariance function for the observation errors, and fitted a model using Markov chain Monte Carlo (MCMC) simulation. Tonellato (1997) proposed a vector AR(p) model with spatially correlated time shocks for fixed observation stations, and applied it in Tonellato (1998) for wind power prediction. The method of Tonellato (1997) is limited by the fact that the dimension of the state vector grows in the number of observation stations. Non-Bayesian space-time SSM approaches have also been suggested. Mardia et al. (1998) combined kriging and SSMs to model spatial covariance and time dependence. They used iterative numerical optimization to estimate variogram parameters and the Kalman filter to estimate temporal trend fields. Similar non-Bayesian approaches include Wikle and Cressie (1999) and Huang and Cressie (1996).

In this paper, we propose a state-space modelling framework for space-time data which imposes minimum constraints on the format of the data and at the same time allows for closed form posterior computations without resorting to MCMC simulations. To account for spatial variability, we model the mean function at each time as a locally-weighted mixture of regression surfaces. Temporal variability is incorporated by allowing the component surfaces to evolve through time. Temporal

trends, cyclical components, and exogenous regressors can all be explored within the SSM framework. Using the Kalman filter and smoothing algorithms, we can obtain posterior and posterior predictive distributions in closed form. This allows quick and straightforward implementation of the model, and provides full probabilistic inference for the parameters, interpolations and forecasts.

The important features of our approach are as follows. It produces space-time maps of interesting quantities along with associated uncertainties. It requires no restrictive assumptions such as stationarity, isotropy, or separability of the space-time correlation functions. The method applies to non-lattice data, and for data stations that move over time. And, finally, computations can be performed at a minimal cost.

In Section 2, we introduce our spatial model and discuss the construction of locally-weighted mixtures. In Section 3, we extend the model to the dynamic case, allowing the parameters to vary with time. The space-time model is then cast in a state-space framework and extensions such as random weight functions and kriging-based structures are discussed. The approach is illustrated in Section 4 using two large datasets: tropical rainfall levels and North Atlantic water temperatures. Conclusions are given in Section 5.

## 2. Spatial Modelling

We assume a spatial model of the form

$$Y(\mathbf{x}) = S(\mathbf{x};\boldsymbol{\beta}) + \epsilon(\mathbf{x}) \tag{1}$$

where $Y(\mathbf{x})$ is an observation taken at location $\mathbf{x}$ in domain $D$, $S(\mathbf{x};\boldsymbol{\beta})$ is a spatial mean process with parameter $\boldsymbol{\beta}$, and $\epsilon(\mathbf{x})$ is a Gaussian noise process. The mean function $S(\mathbf{x};\boldsymbol{\beta})$ is defined as a locally-weighted mixture of linear regressions:

$$S(\mathbf{x};\boldsymbol{\beta}) \;=\; \sum_{j=1}^{J} \pi_j(\mathbf{x})\mathbf{f}_j'(\mathbf{x})\boldsymbol{\beta}_j, \tag{2}$$

where $\mathbf{f}_j(\mathbf{x}) = \{f_{j1}(\mathbf{x}), \ldots, f_{jq}(\mathbf{x})\}'$ is a set of known basis functions, $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jq})'$ is a vector of unknown, random parameters, $\pi_j(\mathbf{x})$ is a non-negative weighting kernel centered at location $\boldsymbol{\mu}_j$, and $J$ denotes the total number of mixture components.

For one realization of the process, we observe $\mathbf{Y} = (Y(\mathbf{x}_1), \ldots, Y(\mathbf{x}_n))'$, at locations $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Define $\boldsymbol{\pi}_j = (\pi_j(\mathbf{x}_1), \ldots, \pi_j(\mathbf{x}_n))'$ as the vector of weights for the $j$th mixture component and $\mathbf{X}_j = [\mathbf{f}_j(\mathbf{x}_1) \;\cdots\; \mathbf{f}_j(\mathbf{x}_n)]'$ as the design matrix for the corresponding component. The full model can be written as a simple linear regression: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{X} = [\mathrm{diag}(\boldsymbol{\pi}_1)\mathbf{X}_1 \;\cdots\; \mathrm{diag}(\boldsymbol{\pi}_J)\mathbf{X}_J]$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_J')'$.

Fig. 1 shows an example of a simple mixture model (2) with $J = 2$ components. Here the spatial process $S(\mathbf{x};\boldsymbol{\beta})$ is constructed using Gaussian weight kernels and linear surfaces. The kernels are defined by $\pi_j(\mathbf{x}) \propto |\boldsymbol{\Sigma}_j|^{-1/2} \exp\{-(\mathbf{x}-\boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_j^{-1}(\mathbf{x}-\boldsymbol{\mu}_j)/2\}$. The linear surfaces are defined by $\mathbf{f}_j(\mathbf{x}_i) = (1, \mathbf{x}_{i1}, \mathbf{x}_{i2})'$, where $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$ denote the first and second coordinates of location $\mathbf{x}_i$.

In general, the process $S(\mathbf{x};\boldsymbol{\beta})$ is nonstationary. If we assume the parameter $\boldsymbol{\beta}$ has a prior distribution with zero mean and covariance matrix $\mathbf{W}$, then the process $S(\mathbf{x};\boldsymbol{\beta})$ has covariance

$$K(\mathbf{x},\mathbf{y}) \;=\; \mathrm{cov}\{S(\mathbf{x})\,,\, S(\mathbf{y})\} \;=\; \sum_{i=1}^{J}\sum_{j=1}^{J} \pi_i(\mathbf{x})\,\pi_j(\mathbf{y})\,\mathbf{f}_i'(\mathbf{x})\,\mathbf{f}_j(\mathbf{y})\,W_{ij}, \tag{3}$$

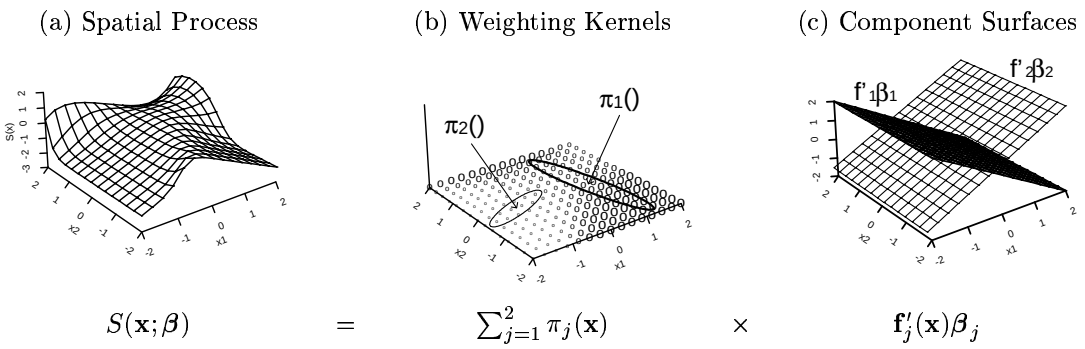


**Fig. 1.** Construction of the locally-weighted mixture. Panel (a): Spatial process generated from a mixture with $J = 2$. Panel (b): 10% HPD contours for the Gaussian weighting kernels and circles representing mixture weights, $\pi_1(\mathbf{x})$, for each point on a spatial grid. Panel (c): linear regression surfaces.

where $\mathbf{x}$ and $\mathbf{y}$ are locations, and $W_{ij}$ denotes the $ij$th element of $\mathbf{W}$. If we choose constant surfaces, i.e., $q = 1$ and $\mathbf{f}'(\mathbf{x}) = 1$, then (3) reduces to $K(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{J} \sum_{j=1}^{J} \pi_i(\mathbf{x})\, \pi_j(\mathbf{y})\, W_{ij}$. If we also choose an i.i.d. prior for the coefficients with covariance matrix $\mathbf{W} = \tau^2\, \mathbf{I}$, then we obtain $K(\mathbf{x}, \mathbf{y}) = \tau^2 \sum_{j=1}^{J} \pi_j(\mathbf{x})\, \pi_j(\mathbf{y})$.

This formulation is similar to the process convolution approach of Higdon (1999). His spatial model can be thought of as a special case of (2), where the local surfaces are constant, and the number of kernels $J$ is large. Another kernel-based approach in a spatial setting is given by Fuentes (2000).

### 2.1. Choosing the Surface and Kernel Types

The choice of the weight functions $\pi_j(\mathbf{x})$ provides a mechanism to include prior information about the unknown spatial process. For example, if the weights are Gaussian kernels as in Fig. 1, then the locations and covariance matrices of the kernels have to be chosen. In many applications prior information is related to the level of spatial smoothness, but the prior distribution over the unknown surface is isotropic, i.e. invariant with respect to shift and rotation of the coordinates. This prior can be represented (approximately) with an equally-spaced grid of spherical kernels. In some problems more specific prior information is available. For example, in Section 4.2 we consider an application where the surface is known to change faster in some regions than in others. Placing more kernels in regions of faster change represents such prior information.

The proposed model includes some standard spatial models as limiting cases. For example, if we choose equally-sized spherical kernels, constant local surfaces, and an i.i.d. prior for the surface coefficients, we obtain an isotropic spatial process as the number of kernels $J$ goes to infinity. Similarly, a geometrically anisotropic process can be obtained if we choose identical non-spherical Gaussian weighting kernels and an i.i.d. prior on the coefficients. Any case involving non-constant surfaces, correlated coefficients, or spatially-varying weighting kernels results in a nonstationary process.

Fig. 2 shows two possible kernel configurations, along with the corresponding covariance function and a simulated spatial process. The kernels are Gaussian, and we assume constant local surfaces

with a white noise prior on the coefficients, $\boldsymbol{\beta} \sim N(0, 10\,\mathbf{I})$. For the covariance plots, we fix $\mathbf{y} = \mathbf{0}$ and graph $K(\mathbf{x}, \mathbf{y})$ as a function of $\mathbf{x}$. Row one shows an example with an equally-spaced grid of $J = 25$ spherical kernels. The correlation function is isotropic and approximately Gaussian. The simulated process is fairly smooth. Row two shows an example of a nonstationary process, with equally-spaced nonspherical kernels with the same orientation but different scales. The covariance is strongest in the southwest-northeast direction, along the major axis of the ellipses. The covariance function falls off faster in the southwest direction than the northeast since the kernels are smaller in the southwest. The effect of the increasing kernel size is also apparent in the simulated process, since it becomes smoother as we move from southwest to northeast.

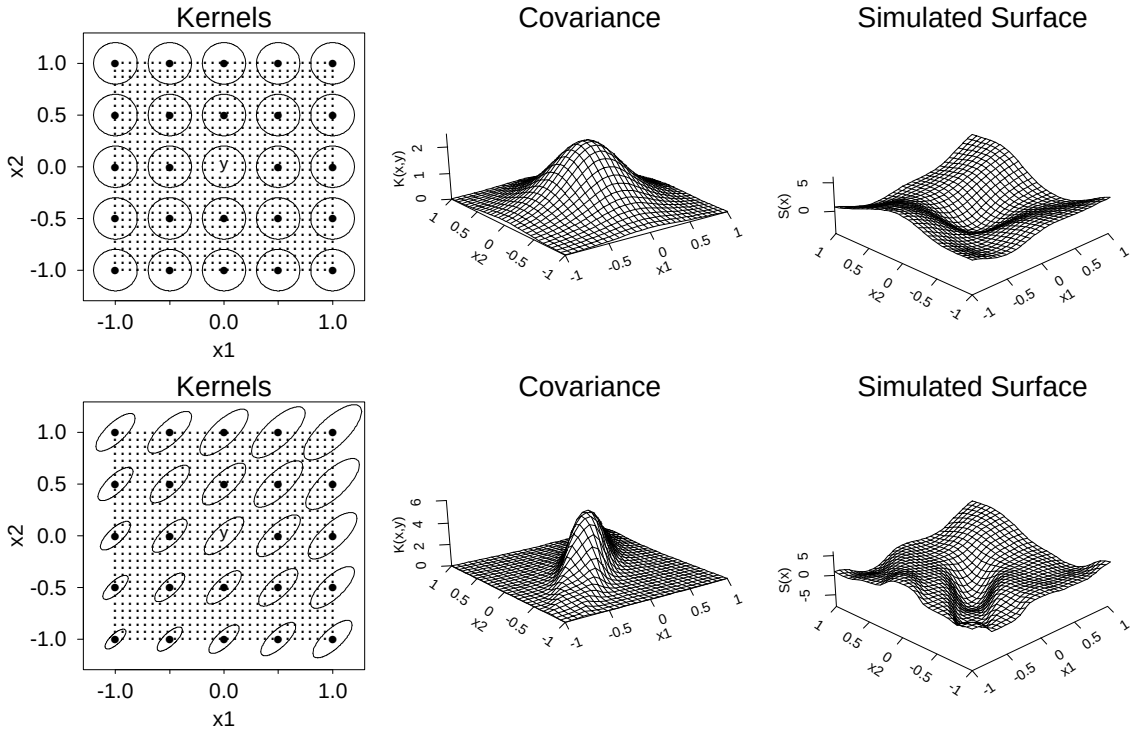


**Fig. 2.** Possible kernel configurations. First column: kernel locations - ellipses represent the central 20% regions for the weight functions. Second column: Prior covariance function at the center of the domain, i.e. $K(\mathbf{x}, \mathbf{y})$, where $\mathbf{y} = \mathbf{0}$ is fixed. Third column: Simulated process $S(\mathbf{x}; \boldsymbol{\beta})$, with constant surfaces and $\boldsymbol{\beta} \sim N(0, 10\,\mathbf{I})$.

We consider Gaussian weighting kernels throughout the paper. Equally-spaced Gaussian kernels lead to a Gaussian covariogram as $J$ goes to infinity. See Higdon, Swall and Kern (1999) for the form of $K(\mathbf{x}, \mathbf{y})$ in this case. With a finite number of kernels, the covariogram is not exactly Gaussian. However, for equally-spaced kernels and reasonably large $J$, a plot of the covariance function $K(\mathbf{x}, \mathbf{y})$ looks indistinguishable from a bivariate Gaussian kernel (see Fig. 2). Besides normal densities, other kernels types could also be used. Kern (2000) provides a catalog of kernels, including Epanechnikov, exponential, and half-sphere, and their corresponding covariance functions. Each of these kernels implies a different level of spatial smoothness for the underlying process.

## 3. Spatio-Temporal Modelling

Suppose now that we have observations collected in space and time. The spatial model (1) can be easily augmented to include a temporal component as follows. First, define $\boldsymbol{\beta}_t = (\boldsymbol{\beta}'_{t1}, \ldots, \boldsymbol{\beta}'_{tJ})'$, as the full parameter vector at time $t$, for $t = 1, \ldots, T$. Next, specify a probability model $p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})$ that links the parameters $\beta_t$ over time. Applying the locally-weighted mixture (2) to each $\boldsymbol{\beta}_t$, we obtain a time series of surfaces $S(\mathbf{x};\boldsymbol{\beta}_1), \ldots, S(\mathbf{x};\boldsymbol{\beta}_T)$ such as those shown in Fig. 3. Assuming a linear evolution equation for $p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})$ the space-time model is

$$\mathbf{Y}_t = \mathbf{X}_t\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{V}_t), \tag{4}$$

$$\boldsymbol{\beta}_t = \mathbf{G}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t), \tag{5}$$

where $\mathbf{Y}_t = (Y(\mathbf{x}_{t1}), \ldots, Y(\mathbf{x}_{tn_t}))'$ are the observations taken at locations $\mathbf{x}_{t1}, \ldots, \mathbf{x}_{tn_t}$. Here, $\mathbf{X}_t$ is the time-dependent design matrix, constructed as in Section 2, whose dimension and elements are determined by the number and locations of the observations at time $t$. Equation (4) is an extension of the spatial model (1), with additional subscripts $t$ to signify the dependence on time, and (5) formalizes change over time of the regression parameters, $\boldsymbol{\beta}_t$. Model (4)-(5) is a Gaussian state-space model (GSSM). In the next section we define the general GSSM and explain how the model fits into this framework.

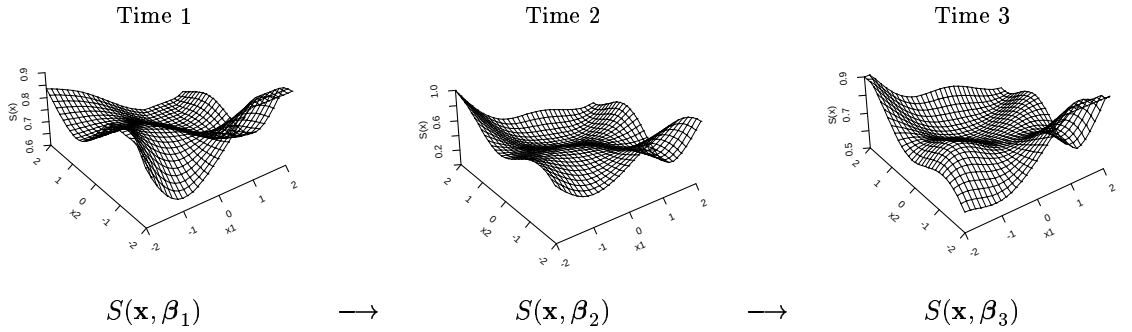


**Fig. 3.** A time series of spatial mean fields. The surfaces were obtained through a locally-weighted mixture of $J = 6$ components with fixed weighting kernels and dynamic regression coefficients.

### *3.1. The State-Space Framework*

A Gaussian state-space model takes the form

$$\mathbf{Y}_t = \mathbf{F}_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{V}_t), \tag{6}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t), \tag{7}$$

for $t = 1, \ldots, T$. The observation equation (6) relates the $n_t \times 1$ data vector $\mathbf{Y}_t$ to the unobserved $p \times 1$ state vector, $\boldsymbol{\theta}_t$. The evolution equation (7) links the states over time. Here, $\mathbf{F}_t$ and $\mathbf{G}_t$ are known as the observation and evolution matrices, respectively, and $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\omega}_t$ are independent white noise sequences with variance-covariance matrices $\mathbf{V}_t$ and $\mathbf{W}_t$, respectively. The model is completed with a prior on the initial state, $\boldsymbol{\theta}_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$.

We now cast the locally-weighted mixture model (4)-(5) in the state-space framework. This allows us to use computationally efficient posterior inference for the GSSM in Model (6)-(7), and to include additional latent temporal components such as trends, seasonal effects and autoregressions. To begin, define the state vector as $\boldsymbol{\theta}_t = (\boldsymbol{\theta}'_{t1}, \ldots, \boldsymbol{\theta}'_{tk})'$, where $k$ is the total number of components and $\boldsymbol{\theta}_{t1} = \boldsymbol{\beta}_t$. Next, write the design matrix $\mathbf{F}_t = [\mathbf{F}_{t1} \ \ldots \ \mathbf{F}_{tk}]$, where the partitioning in $\mathbf{F}_t$ follows the partitioning of $\boldsymbol{\theta}_t$, that is, $\mathbf{F}_{ti}$ is the submatrix which gets multiplied with $\boldsymbol{\theta}_{ti}$. Let $\mathbf{F}'_{t1} = \mathbf{X}_t$, i.e., the first block corresponds to the spatial model (4)-(5). The remaining blocks represent additional components, with state vector $\boldsymbol{\theta}_{ti}$ and design matrix $\mathbf{F}_{ti}$, $i = 2, \ldots, k$. A brief discussion of how the additional parameters $\boldsymbol{\theta}_{t2}, \ldots, \boldsymbol{\theta}_{tk}$ can be used to model trends, seasonal effects, etc. is given in Section 4. For a more detailed discussion, see West and Harrison (1997, chapters 7–9).

In the case of a single component, $k = 1$, the simplest structure for $\mathbf{G}_t$ is diagonal, implying autoregressions for each element of $\boldsymbol{\theta}_t$. A random walk prior, with $\mathbf{G}_t = \mathbf{I}$, is a natural choice when no prior information is available or if no temporal trends are expected. With multiple components, $k > 1$, a block structure can be used, with $\mathbf{G}_t = \text{block-diag}(\mathbf{G}_{t1}, \ldots, \mathbf{G}_{tk})$. This implies a different evolution equation for each model component. Interactions can be included by specifying non-zero elements on the appropriate off-diagonal blocks of $\mathbf{G}_t$.

The evolution variance $\mathbf{W}_t$ can be specified either explicitly or through a discount factor $\alpha \in [0, \infty)$, which defines $\mathbf{W}_t = \alpha \mathbf{P}_t$, where $\mathbf{P}_t = \text{var}(\mathbf{G}_t \boldsymbol{\theta}_{t-1} | \mathbf{Y}_1, \ldots, \mathbf{Y}_{t-1})$. A discount factor of $\alpha = 0$ gives a static model, with the same coefficients for all time periods, while $\alpha \to \infty$ implies coefficients which are independent over time; i.e., no temporal smoothing at all. When $p$ is large, the discount method is often preferred, since it requires specification of only one parameter. This approach can be extended to the case of $k$ components through the use of *block discounting*. In block discounting, one specifies a *vector* of discount factors, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$, with the $i$th factor defining the evolution variance for the $i$th component. That is, we define $\mathbf{W}_{ti} = \alpha_i \mathbf{P}_{ti}$ for the diagonal blocks $\mathbf{W}_{ti}$ and $\mathbf{P}_{ti}$. The off-diagonal blocks of $\mathbf{W}_t$ may be set equal to zero, implying independent temporal innovations for each component.

### 3.2. Computation

In the case when $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$ are known, the full state vector $\boldsymbol{\theta} = (\boldsymbol{\theta}'_0, \ldots, \boldsymbol{\theta}'_T)'$ has a normal posterior distribution that can be obtained in closed form. Recursive equations to obtain the filtering, smoothing and forecast distributions can be found, for example, in West and Harrison (1997, chapter 4). These results can be generalized to the case where $\mathbf{V}$ is invariant over time, i.e., $\mathbf{V}_t = \mathbf{V}$, and we assume a conjugate prior.

A Monte Carlo method known as *forward-filtering-backward-sampling* (FFBS) was developed by Frühwirth-Schnatter (1994) and Carter and Kohn (1994) to generate i.i.d. samples from the posterior (smoothing) distribution $p(\boldsymbol{\theta} | \mathbf{Y}_1, \ldots, \mathbf{Y}_T)$. This technique can be used to numerically evaluate posterior integrals that are analytically intractable. Also, for non-conjugate, conditionally Gaussian models, FFBS can be used as part of a MCMC scheme to resample state parameters conditional on parameters which fix $\mathbf{F}_t$, $\mathbf{G}_t$, $\mathbf{V}_t$ and $\mathbf{W}_t$. This includes, for example, kriging-based models with unknown covariance parameters in $\mathbf{V}_t$.

### 3.3. Model Extensions

In some cases, a model with i.i.d. observation errors might be too restrictive. For more localized spatial processes, a kriging-based approach, which assumes a parametric spatially-dependent covariance function for the errors, $\boldsymbol{\epsilon}_t$, might be preferable. This approach assumes that $V_{ij}$, the $ij$th element of the covariance matrix $\mathbf{V}_t = \mathbf{V}$ is a function of a parameter $\boldsymbol{\lambda}$, and $d_{ij}$, the distance between locations $\mathbf{x}_i$ and $\mathbf{x}_j$,. An empirical Bayes approach, where $\boldsymbol{\lambda}$ is estimated from the data and then held fixed for the analysis, yields a conjugate model with closed forms for all distributions of interest. A more general approach leaves $\boldsymbol{\lambda}$ unknown, and involves specifying a prior distribution, $p(\boldsymbol{\lambda})$. However, the conditional posterior for $\boldsymbol{\lambda}$ does not take the form of any well-known distribution which would allow efficient random variate generation. Thus, numerical methods are necessary to implement posterior inference in such cases, using, for example, MCMC. See Sansó and Guenni (2000) for an example of MCMC in space-time kriging models.

Some spatial processes exhibit highly nonstationary or anisotropic spatial behavior. In such situations, a mixture with multiple types of local surfaces might be appropriate. For example, higher-order polynomials can be used to model regions of rapid change while simpler surfaces might be appropriate in regions of slow change. Spatial anisotropy and nonstationarity could also be modeled by spatially-varying weight functions. Changing kernel widths produce different spatial effects at different locations. For example, tighter Gaussian kernels produce processes that are more localized.

Another extension involving dynamic mixtures allows the weight functions to vary deterministically over time. This can be used as an alternative to static mixtures for processes with significant temporal variation or known changepoints. Here, the weight functions, the number of components, or the basis functions $\mathbf{f}_j$ could change with time. The changes are specified *a priori* through the construction of $\mathbf{F}_t$ and the analytic methods described above remain valid.

How can one select a model for a particular analysis? One possibility is to use cross-validation, where predictions for $\mathbf{Y}_t, \mathbf{Y}_{t+1}, \ldots, \mathbf{Y}_T$ are examined based on a fitted model up to time $t$. Another tool is residual analysis. Spatial correlation in the residuals may indicate that a more flexible spatial structure is needed. This can be solved by the use of additional mixture elements, higher-order local surfaces, smaller weighting kernels, or more complex error structures such as kriging-type models. A third alternative is to consider Bayes' factors. With the marginal likelihood available in closed form as the product of the one-step-ahead forecast densities, Bayes' factors can be calculated using the Kalman filter output.

## 4. Examples

### 4.1. Example 1: Venezuelan Rainfall Data

To illustrate our approach, we analyze Venezuelan rainfall data previously considered in Sansó and Guenni (1999). The data consist of measurements from $n = 80$ fixed monitoring stations throughout the Venezuelan state of Guárico. The original data were collected at 10-day intervals over the period 1968–1983, resulting in a large number of periods with zero rainfall. To deal with the existence of dry periods, Sansó and Guenni (1999) fit a truncated normal model and implemented it using MCMC methods. For our analysis, we avoid the problem of modeling dry periods directly, and instead fit a model to the same data aggregated to the quarterly level.

Fig. 4(a) shows a map of Guárico with the $n = 80$ monitoring stations under consideration. The stations are unequally spaced, with the majority falling in the more-populated northern half and fewer in the southern plains. Fig. 4(b) shows the $T = 63$ quarterly rainfall levels for stations $i = 35, 55$, and 65, whose locations are indicated in Panel (a). As seen in Panel (b), a strong seasonal pattern exists in the state, consisting of a wet period from May–October (Quarters 2 and 3), and a dry period from November–April (Quarters 4 and 1).

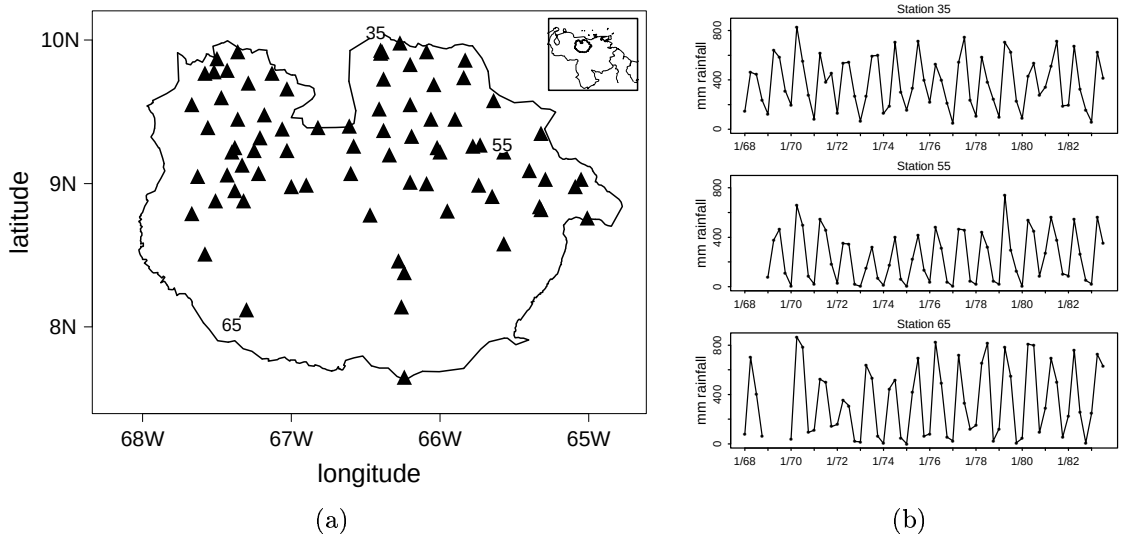


**Fig. 4.** Rainfall monitoring sites in the Venezuelan state of Guárico. Panel (a) gives the locations of the $n = 80$ stations. Panel (b) shows the time series of quarterly rainfall at stations $i$ =35, 55, and 65, which are marked in panel (a). Gaps in the time series indicate missing data.

We begin the analysis by selecting a locally-weighted mixture for spatial modeling. A linear, Gaussian structure is assumed, with $J = 6$ components and Gaussian kernels $\pi_j(\mathbf{x}) \propto |\boldsymbol{\Sigma}_j|^{-1/2}$ $\exp\{-(\mathbf{x} - \boldsymbol{\mu}_j)'\boldsymbol{\Sigma}_j^{-1}(\mathbf{x} - \boldsymbol{\mu}_j)/2\}$. Locations $\boldsymbol{\mu}_j$ of the weighting kernels were chosen by a cluster analysis of the monitoring stations, so that knots were placed in regions with high data density. Placing knots at locations where data points cluster is an ad-hoc approach to allow a priori variability where more data is available.

The kernels were taken to be spherical, with standard deviation 0.5 on a standardized coordinate scale. For the regression parameters, $\boldsymbol{\beta}_t$, we chose a random walk evolution, $\mathbf{G}_{t1} = \mathbf{I}$, reflecting our prior uncertainty about the underlying spatial process. We added a seasonal component to the model, corresponding to an annual cycle with period 4. Using the Fourier representation described in West and Harrison (1997, chapter 8), the second block of the state vector is comprised of two seasonal parameters and the corresponding block of the evolution matrix is given by

$$\mathbf{G}_{t2} = \begin{pmatrix} \cos(\pi/2) & \sin(\pi/2) \\ -\sin(\pi/2) & \cos(\pi/2) \end{pmatrix}.$$

We assume *a priori* that the spatial and cyclical processes are independent, giving a block diagonal

evolution matrix $\mathbf{G}_t$

$$\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{t1} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_{t2} \end{pmatrix}.$$

The observation errors are assumed to be Gaussian, with mean zero and covariance $\mathbf{V}_t = \mathbf{V} = \sigma^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix of order 80. Leaving $\sigma^2$ unknown, we select an inverse-gamma prior: $\sigma^2 \sim IG(1/2, 1/2)$. The evolution variance is specified through a discount factor, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, with $\alpha_1 = 0.10$ for the $3J = 18$ spatial parameters, and $\alpha_2 = 0.05$ for the 2 seasonal parameters. The lower value is used for the seasonal block since we expect the seasonal effect to be more stable over time. To complete the model specification, a diffuse prior was chosen for the initial state vector: $\boldsymbol{\theta}_0 \sim N(\mathbf{0}, 100^2\mathbf{I})$, where $\mathbf{I}$ is the $20 \times 20$ identity matrix.

Fig. 5 shows the estimates of the Quarter 2 mean precipitation field for the years 1968, 1972, 1976, and 1980. Here, the field $\mathbf{z}$ is defined at time $t$ and location $\mathbf{x}$ as $z_t(\mathbf{x}) = \mathbf{F}_t(\mathbf{x})\boldsymbol{\theta}_t$, and maps are based on the full posterior distribution $p(\boldsymbol{\theta}_t|\mathbf{Y}_1, \ldots, \mathbf{Y}_{63})$. The plots indicate that the mean rainfall decreased from 1968 to 1972 in the eastern and southern parts of the state, and then increased throughout the state from 1976 to 1980. We find that estimated rainfall was highest in the southeast in general. As in any regression, the usual caveat about extrapolation beyond the range of the data applies.

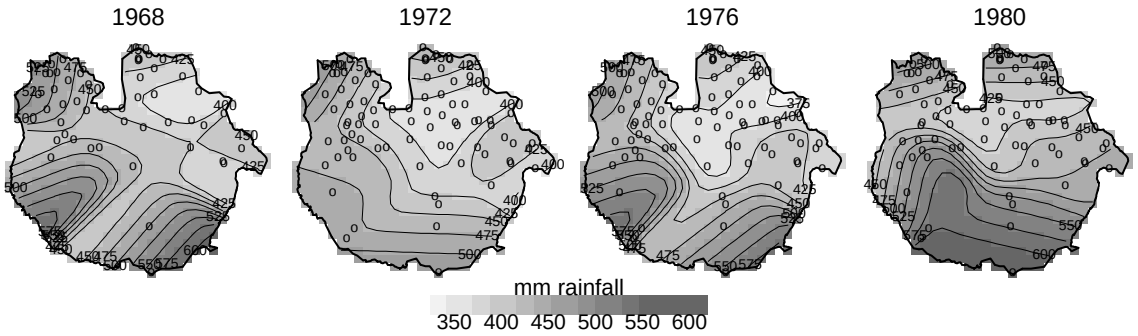


**Fig. 5.** Posterior maps of mean rainfall in Guárico, Quarter 2: 1968, 1972, 1976, 1980. Contours correspond to posterior means of the precipitation field $\mathbf{z}_t$, based on the full data set.

Fig. 6 shows the corresponding uncertainty maps for mean rainfall in Quarter 2. By incorporating this information, we can greatly improve upon inference provided by the mean field alone. As expected, posterior uncertainty is highest in areas with sparse data and lowest near the clusters of stations. The standard deviations are about eight times greater in the southern corners than within the "butterfly region" in the north.

To gauge the interaction between the spatial and temporal processes, we compared the Quarter 2 rainfall maps above with those from Quarter 4. Maps for Quarter 4 (not shown) showed spatial contours similar in shape to those seen in Fig. 5, but with an overall shift due to the seasonal effect. The fact that the spatial pattern changes little from Quarter 2 to 4 indicates that spatial-seasonal interaction is minimal.

In addition to maps of estimated rainfall, we can also perform kriging, or prediction of rainfall at intermediate locations. The distribution of interest here is $p(Y_t^*|Y_1, \ldots, Y_T)$ for any $t$, where $Y_t^*$

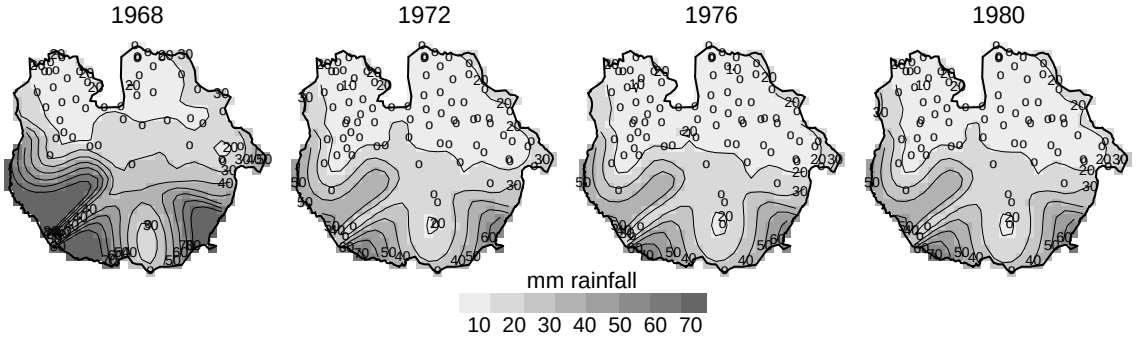


**Fig. 6.** Posterior uncertainty maps in Guárico, Quarter 2: 1968, 1972, 1976, 1980. Contours correspond to posterior standard deviations of the precipitation field, $\mathbf{z}_t$, based on the full data set.

denotes the rainfall levels at new locations $\mathbf{x}^* = (\mathbf{x}_1^*, \ldots, \mathbf{x}_{n^*}^*)$. To illustrate the use of kriging with model (5)-(6), we consider a covariance matrix $\mathbf{V}_t$ in (5) given by

$$V_{tij} \quad = \quad \sigma^2 \exp\{-d_{tij}/\lambda\}, \tag{8}$$

where $V_{tij}$ represents the $ij$th element of $\mathbf{V}_t$ and $d_{tij} = ||\mathbf{x}_{ti} - \mathbf{x}_{tj}||$ is the Euclidean distance between $\mathbf{x}_{ti}$ and $\mathbf{x}_{tj}$. After considering a finely-spaced grid of values for the range parameter $\lambda$, we selected $\lambda = 1$, which maximized the marginal likelihood. Fixing $\lambda$ allows us to obtain the kriging distribution analytically at all time periods.

Figure 7 shows a comparison between the posterior mean rainfall (Quarter 3, 1972) from a static kriging model and the GSSM described above. The static model assumes a constant mean function and an exponential variogram (8), and is estimated using MCMC. See Handcock and Stein (1993) and Sansó and Guenni (2000) for examples of Bayesian kriging. The two surfaces are very similar, but the spatio-temporal model provides additional smoothing based on neighboring time periods.

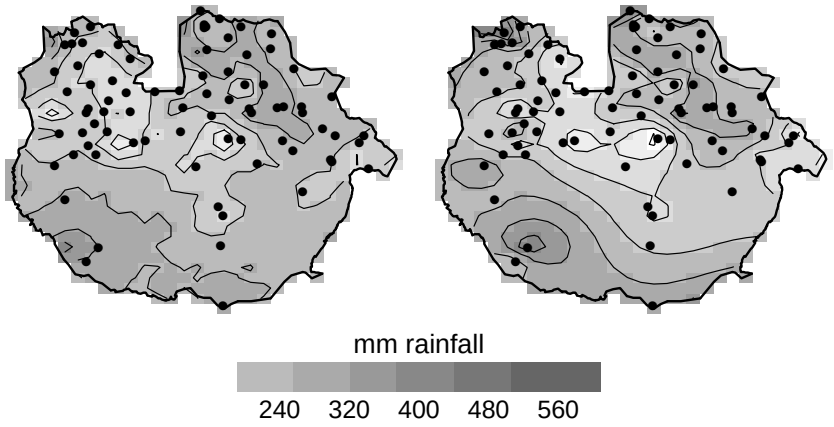


**Fig. 7.** Comparison of predictive mean field for Quarter 3, 1972. Left panel: static kriging model with exponential variogram. Right panel: space-time model with exponential variogram.

One advantage of the analytic approach is that inference can be implemented "on-line". That

is, as new observations arrive, estimates of current parameters and predictions of future observations can be updated very quickly. This can be important when data are observed at short time intervals, or when the underlying system is highly dynamic.

To assess on-line estimation and predictive performance, we split the data into two parts: a training sample consisting of data up to time 40; and a validation set with the remaining observations. The training sample was used to compute the on-line filtering distributions $p(\boldsymbol{\theta}_t|\mathbf{Y}_1,\ldots,\mathbf{Y}_t),\ t = 1,\ldots,40$. After time 40, we computed the forecast distributions for the remaining observations, $p(\mathbf{Y}_t|\mathbf{Y}_1,\ldots,\mathbf{Y}_{40}),\ t = 41,\ldots,63$.

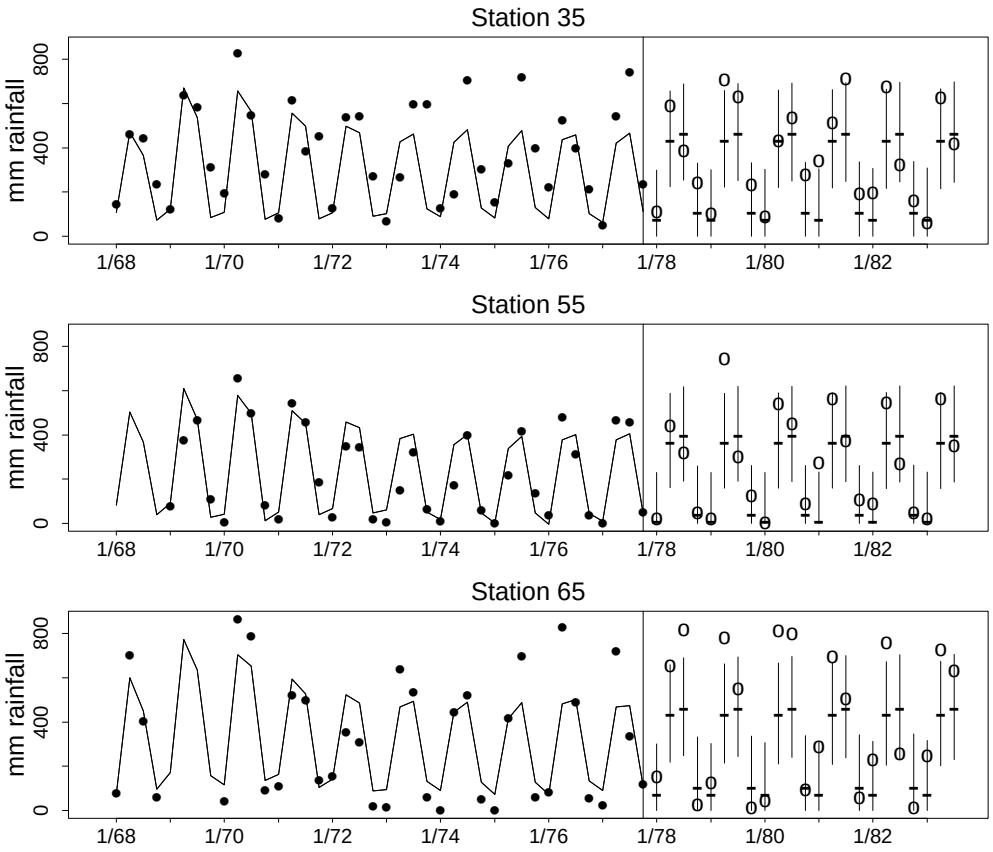


**Fig. 8.** Filtered means and forecast intervals for stations $i = 35, 55$, and 65. The vertical line at Quarter 4/77 ($t = 40$) represents the end of the training sample. Lines before $t = 40$ represent the filtered means; closed circles are the training data. Line segments and hash marks after $t = 40$ represent 95% forecast intervals and forecast means, respectively; open circles are the validation data.

Fig. 8 shows the filtered means and forecast intervals for stations $i = 35, 55$, and 65, along with the observed data at the corresponding stations. The filtered means are defined as $E(z_{ti}|\mathbf{Y}_1,\ldots,\mathbf{Y}_t)$, $t = 1,\ldots,40$, where $z_{ti} = \mathbf{F}_t(\mathbf{x}_i)\boldsymbol{\theta}_t$ is the mean rainfall at time $t$ and site $i$. The model performs well in the filtering stage, capturing both cyclical behavior and spatial variability across the 80 stations. Predictive performance is quite reasonable, as can be seen by comparing the 95% forecast intervals (vertical line segments) with the validation data (open circles). Performance is very good for the

northeastern stations 35 and 55, but the model underpredicts slightly for station 65.

### *4.2. Example 2: North Atlantic Water Temperatures*

We consider a set of Atlantic ocean temperatures analyzed previously by Lavine and Lozier (1999) and Higdon (1999). The data are unevenly distributed in space and time, collected over the period 1905–1988 by research vessels traversing the waters. The original data consist of measurements from various depths ranging from 500–5000 meters; however, we limit the analysis to data from a constant potential density, which corresponds to depths of around 1000 meters. Fig. 9(a) shows the locations and temperatures of the 3987 data points considered here. The majority of points fall along the coastlines of the Iberian peninsula and the shores of Northwest Africa. Due to the flow of warm fresh water from the sea into the ocean, temperatures are warmest near the mouth of the Mediterranean Sea and the coast of Portugal, where most of the measurements were taken. Fig. 9(b) shows a histogram of the sampling times, illustrating the uneven temporal distribution of the data. From Fig. 9(b), we see that over 90% of the data were collected in the last 35 years.

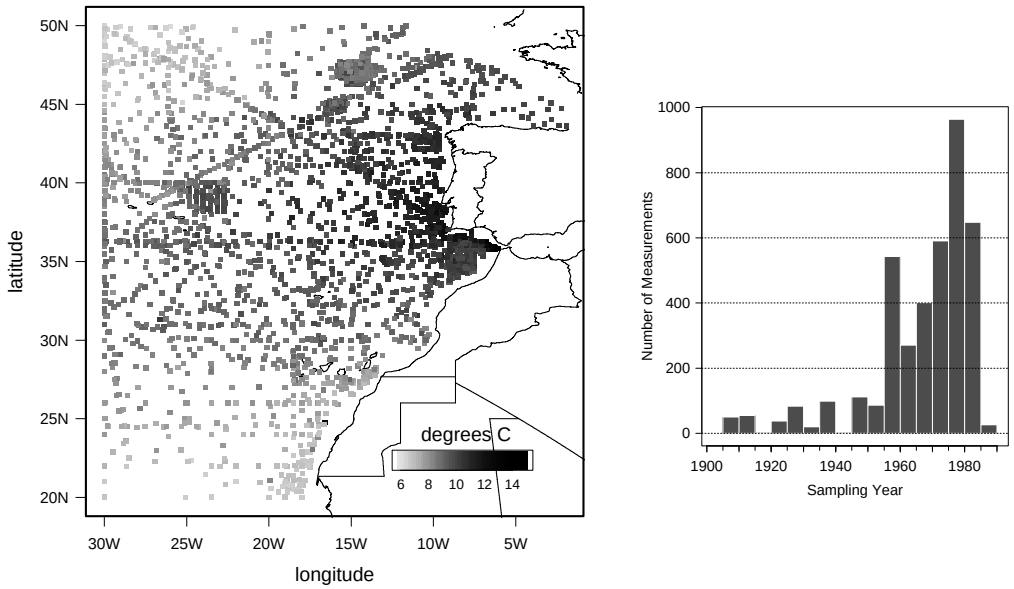


**Fig. 9.** North Atlantic ocean data. Panel (a) shows locations and water temperature for the 3987 observations under consideration. Note that the observations are unequally spaced, often appearing in clusters or along straight lines which correspond to trajectories of the research vessels. Panel (b) shows sampling times for the 3987 observations. The data are also unequally distributed in time, with around 3400 observations collected between 1955 and 1985.

For the following analysis, we use a mixture consisting of $J = 17$ linear surfaces and spherical Gaussian weighting kernels with a standard deviation of $3^o$. We place the majority of the kernels near the Iberian coast and the mouth of the Mediterranean, where the temperatures are expected to change most rapidly. The observation errors are assumed to be i.i.d. with an unknown variance $\sigma^2$.

We use model (5) without additional latent components beyond the spatial model (3), as seasonal effects are negligible at these ocean depths. We specify a random walk for the evolution matrix, reflecting our lack of knowledge about the state process. A discount factor of $\alpha = 0.15$ is assigned for the $p = 3J = 51$-dimensional state vector, and we choose the priors $\boldsymbol{\theta}_0 \sim N(\mathbf{0}, 10^2\mathbf{I})$ and $\sigma^2 \sim IG(1/2, 1/2)$.

Fig. 10 shows the annual mean temperature field every fifth year from 1965 to 1980. These plots can be compared directly to Fig. 7 in Higdon (1999). Although the temporal aspect of the models proposed here and in Higdon (1999) are entirely different, the resulting analyses are very similar. The mean structure in Fig. 10 appears to be fairly stable over time, although contours near the mouth elongate slightly as time increases. Whether this is a real phenomena or merely an artifact of the model is not clear at this point.

One way to answer this and other questions regarding model selection is to consider model sensitivity. For our problem, one interesting question is how the choice of the mixture affects results. To explore this question, we fit three different mixture models, each with linear surfaces and Gaussian weight functions, and compared the estimated 1980 posterior mean field for each.

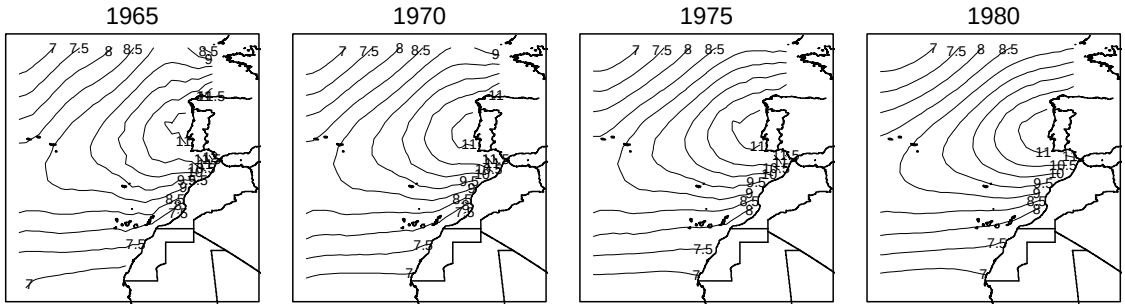


**Fig. 10.** Posterior maps of mean ocean temperatures: 1965, 1970, 1975, 1980. Means are based on the full data set.

Fig. 11 shows kernel locations and the resulting mean fields for the three models: Model A, with $J = 17$ unequally-spaced kernels and a $1^o$ standard deviation; Model B, with the same $J = 17$ knot locations but larger standard deviations of $3^o$; and Model C, with $J = 9$ equally-spaced knots and even larger kernel sizes, $5^o$. Model A, (shown in row 1) illustrates how smaller kernels can lead to insufficient smoothing. The mean field for Model A appears to be unnaturally jagged, giving almost a piecewise linear surface. Model C (row 3) clearly oversmooths, missing important details near the Mediterranean mouth and the coast of Portugal. Model B (row 2) seems to produce the most realistic surfaces, indicating that the intermediate kernel size is appropriate.

In a 1994 letter to Nature Magazine, Parrilla et al. (1994) reported a $1^o$C/century rise in North Atlantic ocean temperatures at depths of 1000 meters. This claim was based on an analysis of data collected along the North Atlantic transect of $24^o$N during the years 1957, 1981 and 1992. A similar analysis was performed in Lavine and Lozier (1999) using temperatures collected over the $20^o$N-$30^o$N transect during the years 1904-1988. Lavine and Lozier (1999) refuted the claim of a warming trend and speculated that the conclusions of Parrilla et al. (1994) were the result of the strong influence of 1957, an anomalously cold year. We ran a similar analysis using the same data as Lavine and

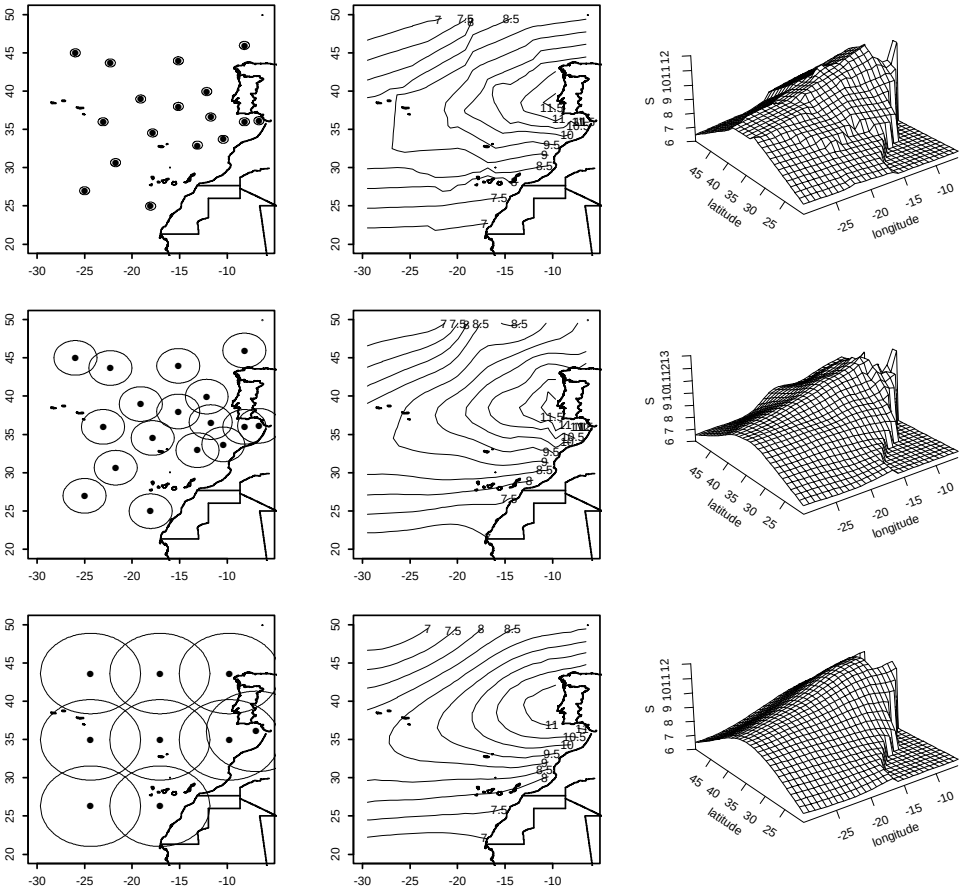


**Fig. 11.** Comparison of 1980 posterior mean field for three models. First column: kernel locations (20% HPD contours for the Gaussian weighting kernels); Second column: mean field contours; Third column: mean field surfaces. Top row: Model A; Middle row: Model B; Bottom row: Model C.

Lozier (1999), and the results are presented below.

We used a mixture model with $J = 10$ spherical Gaussian kernels equally spaced along the line $25^o$N, with linear component surfaces. To study the possibility of temporal trends, an annual mean level component was added to the state vector, making $p = 31$. Defining $\bar{z}_t$ as the average temperature at time $t$ in the region bounded by $20^o$N, $30^o$N, and the coasts of Africa and Florida, we use FFBS to generate 1000 i.i.d. samples from the posterior distribution of $\Delta = \bar{z}_{1981} - \bar{z}_{1957}$. Fig. 12 shows a histogram of samples from $p(\Delta|\mathbf{Y}_{1905}, \ldots, \mathbf{Y}_{1988})$. The distribution is centered at -0.04$^o$C, with standard deviation approximately of 0.75$^o$C, which has a similar location but about twice the spread of the histogram shown in Lavine and Lozier (1999, p. 262). Given that $0^o$C is well within any reasonable highest posterior density (HPD) interval, we concur with the results of Lavine and Lozier (1999), and conclude that no warming occurred in that region of the North Atlantic during the period 1957–1981.

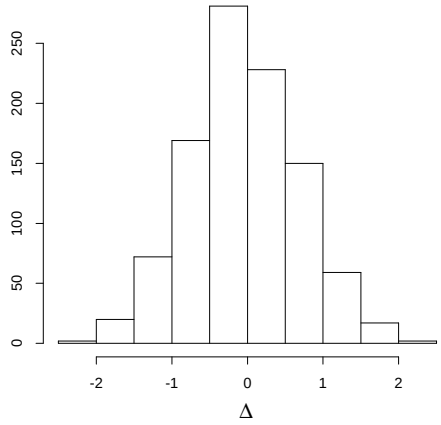


**Fig. 12.** Histogram of the difference in mean temperature from 1981 to 1957. Note that zero is well within any reasonable HPD interval for the distribution.

## 5. Summary

The proposed spatio-temporal models provide a flexible framework for modeling temporal trends, seasonality, and exogenous predictor variables. They allow for nonstationary spatial processes and non-separable space-time correlation functions. The models are valid for any dataset that is discrete in time and continuous in space, which includes data observed on irregular grids or at locations that change over time. This is a major advantage over most existing models which require either fixed observation stations or lattice-based data.

The proposed approach is computationally efficient, allowing for on-line implementation for moderately-sized datasets (say, $n_t < 500$). As new data are collected, the analytic filter provides updated estimates of the mean and uncertainty field as well as predictions at any desired location. This is critical in problems like ozone monitoring, where fast algorithms are necessary to issue warnings in a timely manner. Also, this type of implementation is infeasible for methods involving iterative simulation (e.g., MCMC) or numerical optimization (e.g., the EM algorithm).

Off-line inference can also be implemented efficiently through the recursive smoothing algorithm, which involves inversion of $T$ $(n \times n)$-matrices and $T$ $(p \times p)$-matrices. EM algorithms and other standard non-Bayesian methods for space-time data rely on repeated inversion of such matrices at each iteration. The computational costs for these methods can be enormous. Most current Bayesian approaches to nonstationary space-time modeling are equally costly. For example, the temperature fields shown in Fig. 10 were obtained in nine CPU minutes on a 433 MHz DEC Personal Workstation, whereas similar calculations in Higdon (1999) require five days of CPU time. Interested readers can obtain the C source code used to implement the examples by sending email to the authors.

Besides the models presented in Section 4, a number of extensions are possible. For example, kriging-based approaches can be incorporated as in Example 1, as well as random weight functions and surface types. Such extensions can be implemented using empirical Bayes approaches or standard MCMC simulation. The proposed models, including the extension discussed above, can also be applied to multivariate data, where multiple observations are observed at each point in space and

time.

## 6. Acknowledgements

We wish to thank David Higdon, Lurdes Inoue, Michael Lavine, Doug Nychka, Michael Stein, the Editor, the Associate Editor, and two anonymous referees for their helpful comments that improved the manuscript significantly. This research was partially supported by Grants 97-09696 and xx-xxxx from the National Science Foundation and Grant G97-000592 from Consejo Nacional de Investigaciones Cientificas y Tecnologicas de Venezuela.

## References

Box, G., Jenkins, G. and Reinsel, G. (1994) *Time Series Analysis: Forecasting and Control.* Englewood Cliffs: Prentice Hall, 3rd edn.

Carroll, R., Chen, R., George, E., Li, T., Newton, H., Schmiediche, H. and Wang, N. (1997) Ozone exposure and population density in Harris County, Texas. *J. Am. Statist. Ass.*, **92**, 392–415.

Carter, C. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81**, 541–553.

Cressie, N. (1993) *Statistics for Spatial Data.* New York: Wiley-Interscience, revised edn.

Dellaportas, P., Denison, D. and Mallick, B. (1998) Space-time modelling without distance. *Tech. rep.*, Department of Statistics, Texas A&M University.

Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *J. Time Ser. Anal.*, **15**, 183–202.

Fuentes, M. (2000) High frequency kriging for nonstationary environmental processes. *Tech. rep.*, Department of Statistics, North Carolina State University.

Gelfand, A., Ghosh, S., Knight, J. and Sirmans, C. (1998) Spatio-temporal modeling of residential sales data. *J. Bus. Econom. Statist.*, **16**, 312–321.

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

Handcock, M. and Stein, M. (1993) A Bayesian analysis of kriging. *Technometrics*, **35**, 403–410.

Handcock, M. and Wallis, J. (1994) An approach to statistical spatial-temporal modeling of meteorological fields. *J. Am. Statist. Ass.*, **89**, 368–390.

Haslett, J. and Raftery, A. (1989) Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Appl. Statist.*, **38**, 1–50.

Higdon, D. (1999) A process-convolution approach to modeling temperatures in the north Atlantic Ocean. *Env. Ecol. Statist.*, **5**, 173–190.

Higdon, D., Swall, J. and Kern, J. (1999) Non-stationary spatial modeling. In *Bayesian Statistics 6* (eds. J. Bernardo, J. Berger, A. Dawid and A. Smith), 761–768. Oxford: Oxford University Press.

Huang, H.-C. and Cressie, N. (1996) Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Comput. Statist. Data Anal.*, **22**, 159–175.

Kern, J. (2000) *Bayesian Process-Convolution Approaches to Specifying Spatial Dependence Structure*. Ph.D. thesis, Duke University.

Lavine, M. and Lozier, S. (1999) A Markov random field spatio-temporal analysis of ocean temperature. *Env. Ecol. Statist.*, **6**, 249–273.

Mardia, D., Goodall, C., Redfern, E. and Alonso, F. (1998) The kriged kalman filter. *Test*, **7**, 217–285.

Parrilla, G., Lavín, A., Bryden, H., García, M. and Millard, R. (1994) Rising temperatures in the subtropical North Atlantic Ocean over the past 35 years. *Nature*, **369**, 48–51.

Pfeifer, P. and Deutsch, S. (1980a) Independence and sphericity tests for the residuals of space-time ARMA models. *Comm. Statist. B - Simulation Comput.*, **9**, 533–549.

— (1980b) Stationarity and invertibility regions for low order STARMA models. *Comm. Statist. B - Simulation Comput.*, **9**, 551–562.

Sampson, P. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Statist. Ass.*, **87**, 108–119.

Sansó, B. and Guenni, L. (1999) Venezuelan rainfall data analysed using a Bayesian space-time model. *Appl. Statist.*, **48**, 345–362.

— (2000) A non-stationary multisite model for rainfall. to appear in *J. Am. Statist. Ass.*

Smith, R. and Robinson, P. (1997) A Bayesian approach to the modelling of spatial-temporal precipitation data. In *Case Studies in Bayesian Statistics III* (eds. C. Gatsonis, J. Hodges, R. Kass, R. McCulloch, P. Rossi and N. Singpurwalla), Lecture Notes in Statistics, 237–269. New York: Springer Verlag.

Stoffer, D. (1986) Estimation and identification of space-time ARMAX models in the presence of missing data. *J. Am. Statist. Ass.*, **81**, 762–772.

Tonellato, S. (1997) Bayesian dynamic linear models for spatial time series. *Tech. Rep. Rapporto di ricerca 5/1997*, Dipartimento di Statistica – Università Ca' Foscari di Venezia, Venice, Italy.

— (1998) Spatial prediction with space-time models. *Tech. Rep. Rapporto di ricerca 2/1998*, Dipartimento di Statistica – Università Ca' Foscari di Venezia, Venice, Italy.

Waller, L., Carlin, B., Xia, H. and Gelfand, A. (1997) Hierarchical spatio-temporal mapping of disease rates. *J. Am. Statist. Ass.*, **92**, 607–617.

West, M. and Harrison, P. (1997) *Bayesian Forecasting and Dynamic Models*. New York: Springer, 2nd edn.

Wikle, C., Berliner, M. and Cressie, N. (1999) Hierarchical Bayesian space-time models. *Env. Ecol. Statist.*, **5**, 117–154.

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