Environmental and Ecological Statistics **5,** 173–190 (1998)

A process-convolution approach to modelling temperatures in the North Atlantic Ocean

DAVID HIGDON

Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708-0251, USA

Received June 1997. Revised February 1998.

This paper develops a process-convolution approach for space-time modelling. With this approach, a dependent process is constructed by convolving a simple, perhaps independent, process. Since the convolution kernel may evolve over space and time, this approach lends itself to specifying models with non-stationary dependence structure. The model is motivated by an application from oceanography: estimation of the mean temperature field in the North Atlantic Ocean as a function of spatial location and time. The large amount of this data poses some difficulties; hence computational considerations weigh heavily in some modelling aspects. A Bayesian approach is taken here which relies on Markov chain Monte Carlo for exploring the posterior distribution.

Keywords: Bayesian inference, moving average, non-stationarity, oceanography, space-time modelling, spatial correlation

1. Introduction

Because of its active deep water formation sites, the North Atlantic plays a key role in evaluating global ocean climate. To this end, atlases of ocean properties such as those in Levitus (1982, 1994) and Lozier *et al.* (1995) have been developed which give mean field properties such as temperature, salinity, pressure and dissolved oxygen. These atlases are compiled using data primarily collected from research vessels over the past 80 years which are archived at the National Oceanographic Data Center (NODC). Using these data, which are predominantly from years after 1970, mean fields have been estimated which show the various property fields as a function of global location; the temporal component of the data is ignored in these constructions.

Such constructions use local averaging of the observations to obtain spatial estimates of the various fields. A statistically based approach to spatial and temporal estimation allows the data to dictate the amount of spatial smoothing that seems appropriate and can also accommodate the temporal component of the data. Not only will this lead to time averaged spatial field estimates that are not biased towards more recent years because of their abundant data, but this will also make it possible to detect changes to the temperature field over time. In addition, statements about uncertainty can also

be supplied from a statistical approach. Another class of models combines physical ocean dynamics with measurements to predict the ocean flow field; see for example Miller (1986), Fukumori and Malanotte-Rizzoli (1995) or Rane *et al.* (1996). Cummings *et al.* (1997) give a description of the US Navy's ocean analysis systems which assimilates data from many sources. This additional information can be used to construct a 'first guess', which is usually incorporated in the mean field of a space or space-time model. Thiébaux (1997) discusses how incorporating such information can affect the resulting estimates. However, such approaches are not considered since the goal here is to obtain a data-based estimate of the temperature field.

In this paper, we focus on the temperatures recorded in an area of the North Atlantic which contains the entrance to the Mediterranean Sea. The data are explored further in the section below. The spatial and temporal components of the data are accounted for using a *process-convolution* approach which accommodates the large amount of data and allows for a dependence structure that evolves over space. This is only an initial attempt to apply a space-time model to this data. There are still a number of modelling issues that require further exploration before this analysis can be considered complete. Details of this model are given in Section 3, methodology for posterior inference is given in Section 4, and initial results based on this model are covered in Section 5. A discussion of aspects of the proposed model and future directions is given in Section 6.

2. Temperature data

This study focuses on North Atlantic temperature data collected between 1908 and 1988 in the region given by 20–50°N latitude and 8W–30W longitude. The primary source of spatial temperature variation is warm, salty water from the Mediterranean Sea spilling out into the Atlantic. This water flows along surfaces of constant potential density, called isopycnals. Potential density is the density a volume of water would have if it were on the ocean surface; hence it does not reflect the additional increase in pressure due solely to depth. Similarly, potential temperature is the temperature a volume of water would have if it were on the ocean surface; hence it does not reflect the gain in temperature attributed to the depth-induced pressure increase. The data, which are given in Fig. 1, show 3987 potential temperatures along an isopycnal which corresponds to depths well below 1000 metres throughout most of the region shown; obviously erroneous observations have been screened out. At this depth, water shows no seasonal variation in temperature and tends to remain on its isopycnal surface; changes in temperature along an isopycnal can be attributed primarily to advection and/or mixing of water from different sources. Hence a temperature map gives insight into how water flows and mixes within the ocean.

3. A spatial process-convolution model

Since the goal here is to obtain a smooth temperature field, effort will focus on determining the large scale properties of the field and changes over time. As is standard for many spatial applications, we begin by modelling the temperature field $y(s,t)$ as a

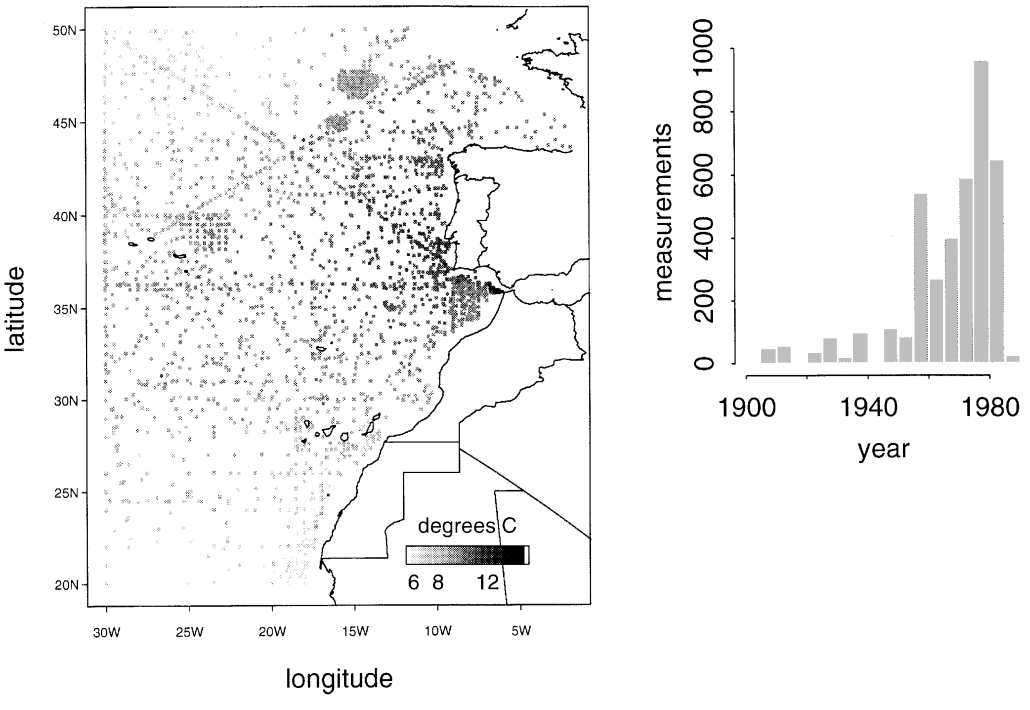

Fig. 1. Temperature measurements taken between 1908 and 1988. The histogram shows the amount of data collected over time.

Gaussian process over space s and time t. Our specification of $y(s,t)$ extends the moving average process (see Thiébaux and Pedder, 1987, Ch. 5) so that the spatial dependence structure changes over space; details are given below. This model with spatially evolving dependence structure is preferred since the smoothness of the process clearly changes over location. For example, the nature of the temperature gradient near the mouth of the Mediterranean differs from that in the region along the 25°N latitude.

Detecting any temporal dependence here is far more difficult than with more common space-time datasets such as in Handcock and Wallis (1994) or Guttorp *et al.* (1994) where data are recorded at a fixed set of locations repeatedly over time. Here no spatial location is ever truly resampled over time, making it very difficult to detect any complicated temporal structure as is done in time series models. Instead we consider simplistic temporal structure based on standard variogram models.

Before getting into modelling details we note that a balance must be struck which, at the same time, allows for a coherent space-time model and a computationally feasible means for fitting the model; this includes inference as well as an estimation of the mean field. The use of kriging methods, which require the construction of a covariance matrix through a variogram model, is not feasible due to the large amount of data. Arguably, a very well equipped and very modern work station could cope with this particular dataset via kriging, however the goal of this research is eventually to model the entire North Atlantic, so that an alternative modelling strategy is necessary. In order to main-

tain the appealing spatial structure of kriging models while avoiding their computational demands, a process-convolution model is developed for this application. Rather than describing a Gaussian process through its mean and covariance structure, a constructive approach is taken which builds a spatially and temporally dependent, continuous process by convolving an independent, discrete process which is defined over a grid.

The temperature field $y(s,t)$ is defined over location s and time t to be the sum of two processes

$$y(s,t) = z(s,t) + \epsilon(s,t) \tag{1}$$

where $z(s,t)$ denotes a smooth Gaussian process, and $\epsilon(s,t)$ is an independent error process. The smooth process $z(s,t)$ is defined to be the convolution of a Gaussian white noise process. As a simple example, suppose x is a 1-dimensional lattice of i.i.d. $N(0,\sigma^2)$ random variables. If this process were then convolved with a Gaussian kernel, or some other kernel, the result would be a smooth, continuous 1-dimensional process as shown in Fig. 2. The idea of convolving a process to produce a smooth surface has been used by a number of researchers in the context of spatial modelling: Thiébaux (1976) and Thiébaux and Pedder (1987, Ch. 5) develop such models in the geophysical context; Matheron (1986) discusses convolving a Poisson process; Wolpert and Ickstadt (1995) develop a rather general approach employing convolutions of gamma and stable processes; and Berry and Ver Hoef (1996) use a process-convolution approach for estimation of the variogram. The approach here differs since the smoothing kernel may vary as a function of spatial location. Another difference is the emphasis on implementational aspects which accommodate the large data set.

We construct the smooth process $z(s,t)$ to model the data in Fig. 1 by taking the convolution of a 3-dimensional lattice process. The convolution kernel which depends on spatial location s is given by $K_s(\Delta s, \Delta t)$. Additionally defining the lattice, or grid process, $x(s,t)$ completely specifies the smooth process $z(s,t)$. The grid process $x = (x_1, \ldots, x_m)$ has space-time coordinates $(\omega_1, \tau_1), \ldots, (\omega_m, \tau_m)$. A cross-section of the grid which shows the spatial locations of x is given by the points in Fig. 4. A

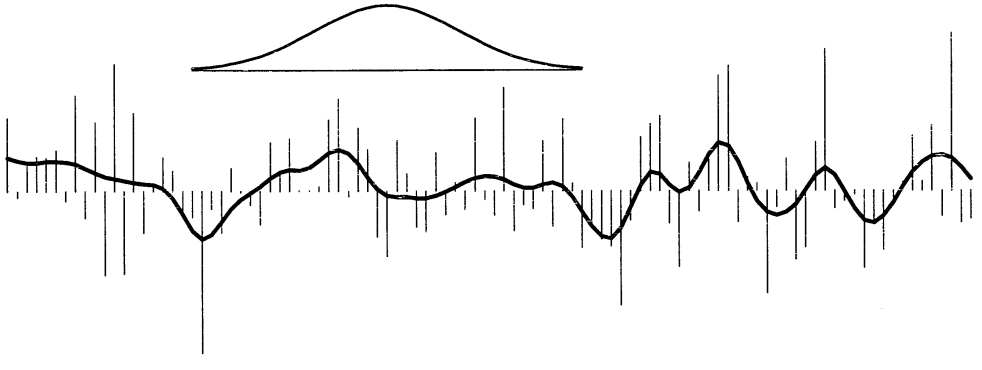

Fig. 2. Taking a convolution of iid Gaussian random variables leads to a smooth, stationary Gaussian field.

priori, each x_j is specified to be independent with a $N(\mu_x, \sigma_x^2)$ distribution. Given x, the smooth field z can be expressed as

$$z(s,t) = \sum_{j=1}^{m} K_s(s - \omega_j, t - \tau_j) \cdot x_j \tag{2}$$

The properties of the convolution kernel determine the smoothness of z. We use a separable kernel so that K_s is represented as a product of a kernel that smooths over space, and one that smooths over time

$$K_s(\Delta s, \Delta t) = C_s(\Delta s) \cdot R(\Delta t)$$

Details on the estimation of the convolution kernel, both its temporal and spatial components, as well as choosing the grid spacing, are given below.

3.1 *Temporal dependence*

Obtaining some idea of the spatial dependence is rather difficult since the data become quite sparse as one goes back in time. This is primarily why previous efforts to construct spatial temperature maps make no effort to account for changes over time. Figure 3 shows a variogram produced by estimating the variance of temperature differences $y(s,t) - y(s',t')$ as a function of the length of time separating the observations $|t - t'|$. In order to limit the effects of the obvious spatial trend in the data, only differences for which $\|s - s'\| < 1$ degree were considered in constructing the variogram. The empirical variogram is rather sensitive to changes in the binning used to estimate the semi-variances and it has been difficult to find a satisfactory measure of the strength of temporal dependence. In most variograms produced, temporal dependence seems to die off somewhere between 5 and 15 years which is consistent with what is considered to be the extent of spatial dependence at this scale. Using a weighted least squares approach (Cressie, 1991), a Gaussian variogram model is fitted to the data; the scale parameter estimate σ_t is 7.0 where σ_t governs the decay of the temporal dependence

$$\mathrm{Cov}(z(s,t), z(s,t')) \propto \exp\left\{ -\frac{1}{2}\left(\frac{t - t'}{\sigma_t}\right)^2 \right\} \qquad \text{for } \; t \neq t'$$

We also consider other values for the scale in checking the sensitivity of the model results. The choice of modelling the temporal dependence using a Guassian variogram model is largely due to computational considerations and is discussed further in the section below. We note here that because of the depth of the measurements, no seasonal trend is present.

3.2 *Spatial dependence*

Currently, the procedure for estimating the local spatial convolution kernel uses a costly maximum likelihood approach. Because of this computational burden, kernels are estimated locally at several locations over the region, and then a simple weighting scheme is used to determine the kernel at intermediate locations. Below we give details on the choice of kernel as well as details on estimation.

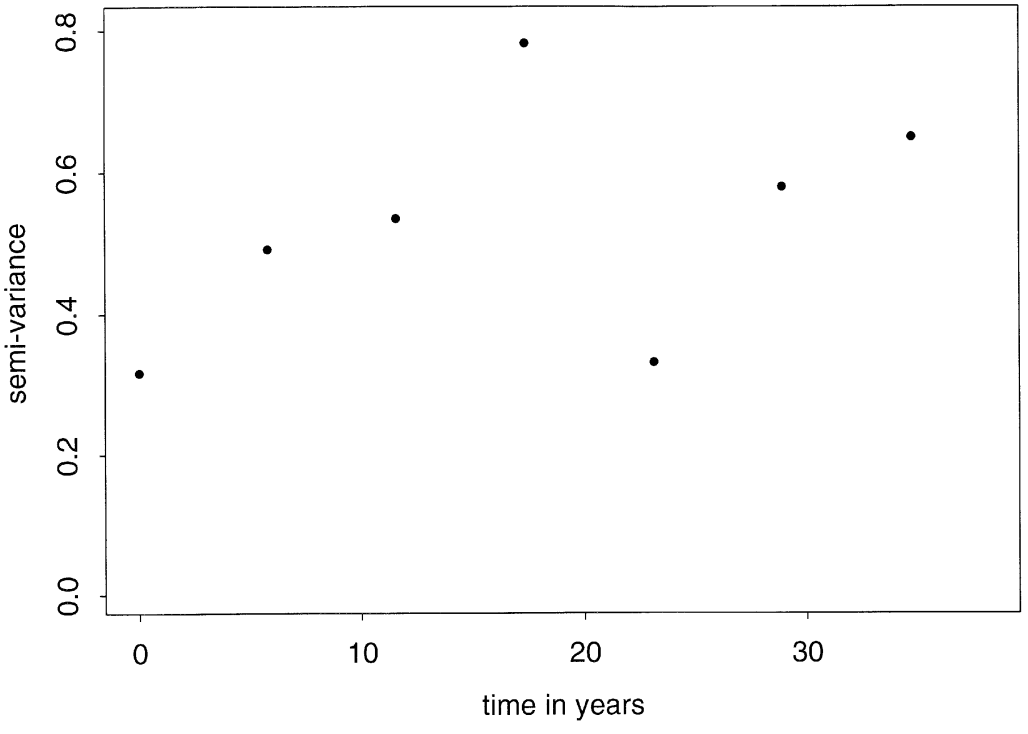

Fig. 3. Empirical variance of temperature differences as a function of time between points. Only points that were within one degree of latitude and longitude were considered in computing variances.

At a given spatial location s we can estimate the local spatial convolution kernel $C_s(\Delta s)$ directly from observations in a neighbourhood of s. We use Gaussian convolution kernels for physical as well as computational reasons. Barnes (1964) gives an argument based on physical ocean dynamics that suggests the temperature field results from a Gaussian convolution. In addition, the Gaussian kernel is straight forward to estimate from the data and it allows for a rather coarse representation of the underlying grid process x without any appreciable bias due to its lack of continuity. A coarse representation of x is important in controlling the computational demands of fitting this model.

A set of *basis* kernels is estimated at eight spatial locations $\ell_1, \ldots, \ell_8$ by first fitting an anisotropic Gaussian variogram model to data within a 4° radius of s. The variogram parameters are then transformed to correspond to parameters of a Gaussian smoothing kernel. This is a simple transformation since a spatial process specified by a Gaussian variogram is equivalent to a process obtained by convolving a white noise process with a Gaussian kernel.

Local anisotropy is captured by allowing a rotation and stretching of the coordinate axes at s. Raw distance $d = (d_1, d_2)$ from s is transformed according to

$$r = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$$

Hence the parameters $(\sigma_1, \sigma_2, \theta)$ control the anisotropy at s. The local variogram model can be written as

$$\frac{1}{2}\mathrm{Var}(y(s,t) - y(s',t')) = \begin{cases} \sigma_\epsilon^2 + \gamma \left[1 - \exp\{-\{12 r^T r - \frac{1}{2}\left(\frac{t-t'}{\sigma_t}\right)^2\}\right] & \text{for } s \neq s' \text{ or } t \neq t' \\ 0 & \text{for } s = s' \text{ and } t = t' \end{cases}$$

The previously estimated σ_t is treated as known, so at any given location, estimation of the anisotropic variogram involves five parameters. However only the parameters $(\sigma_1, \sigma_2, \theta)$ are necessary to determine the Gaussian convolution kernel. Straightforward calculations show that the process defined by the above variogram, without the nugget term σ_ϵ^2, is equivalent to the process obtained by convolving a Gaussian white noise process with a Gaussian density kernel whose covariance matrix Σ is given by

$$\Sigma^{-\frac{1}{2}} = \frac{1}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

Of course the implementation takes a convolution over a grid, rather than a continuous process. However initial calculations show that convolving a rather coarse grid is a good approximation to convolving the resulting continuous process. In this case, spacings no greater than the minimum of $1/\sqrt{2}\sigma_1$ and $1/\sqrt{2}\sigma_2$ give a satisfactory approximation. Figure 4, which shows the spatial grid locations of x, also shows one standard deviation ellipses corresponding to the covariance matrix of each Gaussian convolution kernel. The ellipses are centred over their locations $\ell_1, \ldots, \ell_8$.

The convolution kernel at any given location s is determined by taking a simple weighted average of the eight *basis* kernels $C_{\ell_1}(\Delta s), \ldots, C_{\ell_8}(\Delta s)$. At any s the kernel can be written

$$C_s(\Delta s) = \sum_{k=1}^{8} w_k(s) C_{\ell_k}(\Delta s)$$

where the weights are based on distance according to the formula

$$w_k(s) \propto \exp\left\{ -\frac{1}{2}\|s - \ell_k\|^2 \right\}$$

with the constraint that $\sum_{k=1}^{8} w_k(s) = 1$ for all s. This leads to a weighting that nearly interpolates C_{ℓ_k} when s is near ℓ_k and also gives a smooth transition as the location moves from one basis location to another. Of course this approach of determining C_s is admittedly *ad hoc*. Alternatively one may prefer a weighted least squares approach to estimate the variogram parameters. Whatever the preference, the family of convolution kernels determining C_s should vary slowly over space while conforming to the local data. How to best specify C_s is still an open question and currently under investigation. However, any such approach will need to be computationally efficient and reliable enough to be carried out automatically.

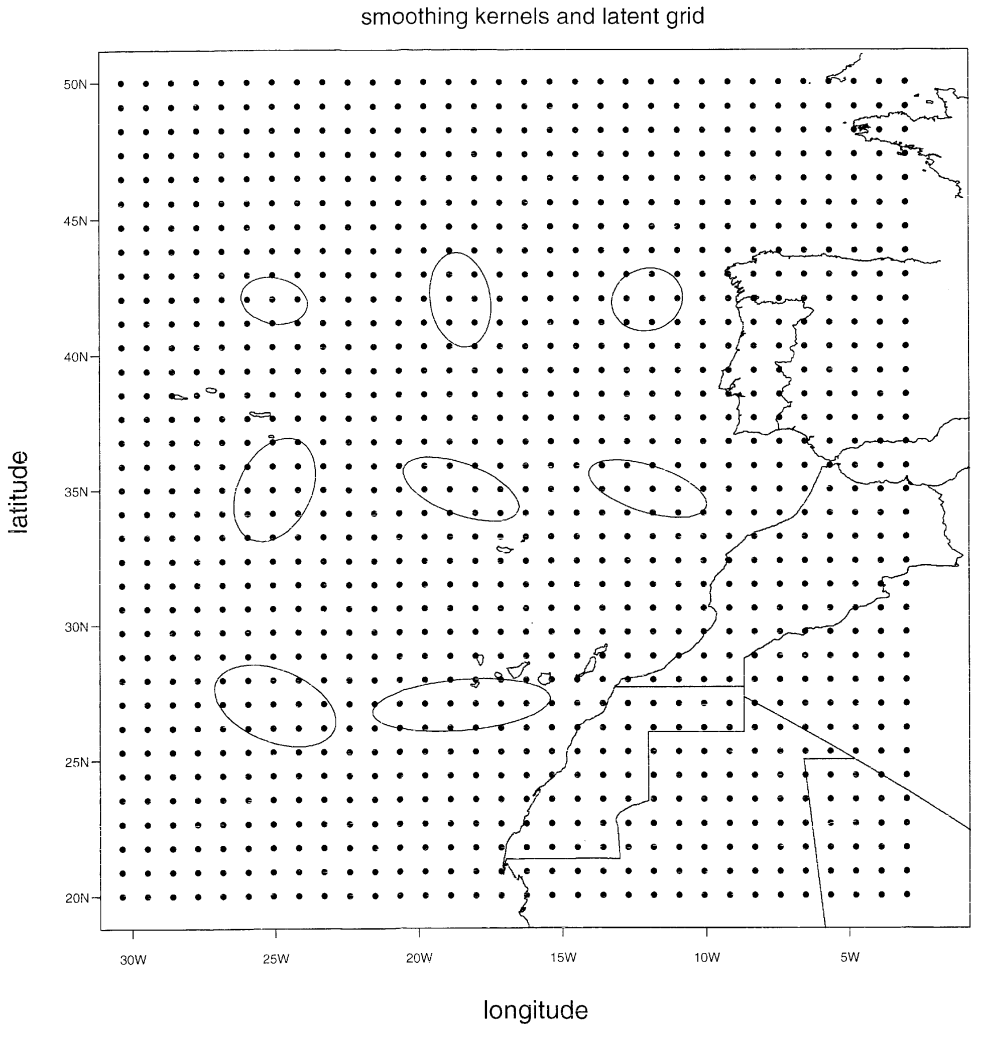

Fig. 4. Spatial locations of the underlying grid process x are given by the dots. One standard deviation ellipses corresponding to the covariance matrix of the Gaussian smoothing kernels are also shown.

4. The posterior distribution

Let the vector y denote the n temperatures $y_1, \ldots, y_n$ recorded at space-time locations $(s_1, t_1), \ldots, (s_n, t_n)$. We can define the likelihood as a function of the error variance σ_ϵ^2 and the grid process $x = (x_1, \ldots, x_m)$ which has space-time coordinates $(\omega_1, \tau_1), \ldots, (\omega_m, \tau_m)$

$$L(y|x, \sigma_\epsilon^2) \propto \sigma_\epsilon^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2}\sigma_\epsilon^{-2}(y - Kx)^T(y - Kx) \right\}$$

where K is a $n \times m$ matrix with (i, j) element

$$k_{ij} = C_{s_i}(s_i - \omega_j) \cdot R(t_i - \tau_j)$$

The model formulation is completed by specifying prior distributions for σ_ϵ^2, x, μ_x and σ_x^2. The priors are given below

$$\begin{aligned}
\pi(\sigma_\epsilon^2) &\propto \sigma_\epsilon^{-2(a-1)} \exp\{-b/\sigma_\epsilon^2\},\ \sigma_\epsilon^2 > 0 \\
\pi(x|\mu_x, \sigma_x^2) &\propto \exp\left\{-\frac{1}{2}\sigma_x^{-2}(x-\mu_x)^T(x-\mu_x)\right\} \\
\pi(\mu_x) &\propto 1 \\
\pi(\sigma_x^2) &\propto \sigma_x^{-2(c-1)} \exp\{-d/\sigma_x^2\},\ \sigma_x^2 > 0
\end{aligned}$$

The priors for σ_ϵ^2 and σ_x^2 are rather diffuse inverse gamma distributions. We use $a = c = 1$ and $b = d = 0.005$ as the constants that specify the prior. Since the data are so plentiful, the resulting inference is quite insensitive to the choice of a, b, c and d. As detailed earlier, each element x_j of the grid process is i.i.d. $N(\mu_x, \sigma_x^2)$; a vague, uniform prior is specified for μ_x. The smoothing kernels which depend on spatial location are treated as fixed so that the estimation will be computationally tractable. This means that uncertainty regarding the smoothing kernels is not accounted for in the analysis. Note that conditioning on the smoothing kernels is not equivalent to conditioning on the estimated variogram. In this analysis, uncertainty about the parameters σ_ϵ^2 and σ_x^2 is accounted for in the posterior distribution.

The above formulation leads to the posterior distribution of the parameters

$$\pi(x, \sigma_\epsilon^2, \mu_x, \sigma_x^2|y) \propto L(y|x, \sigma_\epsilon^2) \times \pi(x|\mu_x, \sigma_x^2) \times \pi(\sigma_\epsilon^2) \times \pi(\mu_x) \times \pi(\sigma_x^2) \tag{3}$$

Computational approaches to locating a point estimate for the parameters as well as exploring the posterior rely heavily on the full conditional distributions of (3). The full conditional distribution of a given parameter is its distribution conditional on all other parameters, including the data. Below are the resulting full conditionals

$$\begin{aligned}
x_j|\cdots &\sim N\left(\frac{\frac{1}{\sigma_\epsilon^2}r^{jT}k^j + \frac{1}{\sigma_x^2}\mu_x}{\frac{1}{\sigma_\epsilon^2}k^{jT}k^j + \frac{1}{\sigma_x^2}}, \frac{1}{\frac{1}{\sigma_\epsilon^2}k^{jT}k^j + \frac{1}{\sigma_x^2}}\right) \\
1/\sigma_\epsilon^2|\cdots &\sim \Gamma\left(a + \frac{n}{2}, b + \frac{1}{2}(y-Kx)^T(y-Kx)\right) \\
\mu_x|\cdots &\sim N\left(\frac{1}{m}\sum_{j=1}^m x_j, \frac{\sigma_x^2}{n}\right) \\
1/\sigma_x^2|\cdots &\sim \Gamma\left(c + \frac{m}{2}, d + \frac{1}{2}(x-\mu_x)^T(x-\mu_x)\right)
\end{aligned}$$

where k^j is jth column of the matrix K and r^j is a $n \times 1$ vector given by

$$r^j = y - \sum_{j' \neq j} k^{j'} x_{j'}$$

Model fitting is done in two stages:

1. Estimation of initial point estimates for σ_ϵ^2, x, μ_x, and σ_x^2.
2. Exploration of the posterior distribution $\pi(x, \sigma_\epsilon^2, \mu_x, \sigma_x^2|y)$ via Markov chain Monte Carlo (MCMC).

Both steps require either maximization of or simulation from the full conditional distributions above. If inference is not required, then stopping at step 1 may be sufficient. However, a primary goal of this analysis is to quantify uncertainty about changes in flows over this region.

We note that for this particular dataset, one could marginalize out the x process and deal with the induced $n \times n$ covariance matrix governing y analytically. However, in the future, we envisage a fully Bayesian analysis which can deal with larger datasets and will treat the parameters governing the smoothing kernels K_s as unknowns. So even though an analytical approach may be practical with this dataset on a nicely equipped modern workstation, we use an MCMC approach since it will easily adapt to future extensions of this formulation, and it will still give equivalent results (up to Monte Carlo standard errors).

4.1 *Obtaining point estimates*

The parameters are initialized by estimating σ_ϵ^2 and σ_x^2 from the data and then estimating x and μ_x by alternately maximizing $\pi(x|\cdots)$ and $\pi(\mu_x|\cdots)$. The parameter σ_ϵ^2 can be estimated by looking at the semi-variogram for data points that are very close together in space and time. If we let $i \sim i'$ denote the set of all nearby data locations such that $\|s_i - s_{i'}\| < 1$ degree and $|t_i - t_{i'}| < 2$ years, then σ_ϵ^2 may be estimated by

$$\widehat{\sigma_\epsilon^2} = \frac{1}{2}\frac{1}{N_{i\sim i'}}\sum_{i\sim i'}(y_i - y_{i'})^2$$

where $N_{i\sim i'}$ is the number of pairs in $i \sim i'$.

By taking the variance of both sides of (1) one obtains the equation

$$\mathrm{Var}(y) = \sigma_x^2 K K^T + \sigma_\epsilon^2 I_n$$

Because of the large amount of data, determination of $\widehat{\sigma_x^2}$ directly via matrix computations is too demanding a task; a simple compromise is to estimate σ_x^2 with the equation

$$\widehat{\sigma_y^2} = \widehat{\sigma_x^2}\,\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} k_{ij}^2 + \widehat{\sigma_\epsilon^2}$$

where $\widehat{\sigma_y^2}$ is the empirical variance of the temperature data. Note that because the convolution of x uses a spatially varying mixture of kernels, $\sum_j k_{ij}^2$ does vary, though $\sum_j k_{ij}$ remains constant.

Conditional on the estimated values for σ_ϵ^2 and σ_x^2, the remaining two parameters are then taken to maximize $\pi(x, \mu_x|\widehat{\sigma_\epsilon^2}, \widehat{\sigma_x^2}, y)$. This is carried out using the iterated conditional modes (ICM) scheme (Besag, 1986), which repeatedly replaces each parameter value by the value that maximizes its full conditional. After initializing μ_x, ICM is implemented using the following steps

1. For $j = 1, \ldots, m$, set

$$x_j = \frac{\frac{1}{\sigma_\epsilon^2} r^{jT} k^j + \frac{1}{\sigma_x^2} \mu_x}{\frac{1}{\sigma_\epsilon^2} k^{jT} k^j + \frac{1}{\sigma_x^2}}$$

2. Set

$$\mu_x = \frac{1}{m} x_j$$

3. Repeat steps 1 and 2 until the values have converged.

Since the conditional distribution of x and μ_x is log-concave, the ICM algorithm is certain to converge to the global maximum. In step 1 above, when x_j is updated, the newly updated values of $x_1, \ldots, x_{j-1}$ are used on the right-hand side. For this data set, the solution $(\hat{x}, \hat{\mu}_x)$ converges after about 30 complete iterations.

4.2 *Posterior inference via MCMC*

Once the initial estimates of the parameters have been obtained, inference can proceed by generating samples from the posterior distribution via MCMC. The sample is produced by constructing a partial realization of a Markov chain whose stationary distribution is given by (3). Detailed descriptions of the methodology can be found in the Royal Statistical Society collection (Smith and Roberts, 1993; Besag and Green, 1993; Gilks *et al.*, 1993); Geyer (1992), Tierney (1994) and Besag *et al.* (1995), including the accompanying discussions.

In this application we initialize the chain at the point estimate obtained above. The partial realization is then generated using a Gibbs sampler (Gelfand and Smith, 1990), which replaces each parameter in turn by a draw from its full conditional distribution. One cycle of the Gibbs sampler consists of a single update to each of the parameters. Since the initial estimate is not necessarily a draw from the posterior, the chain is allowed to 'burn in' until the stationary distribution is reached. Diagnostics from Johnson (1996) indicate the chain reaches the stationary distribution after about 150

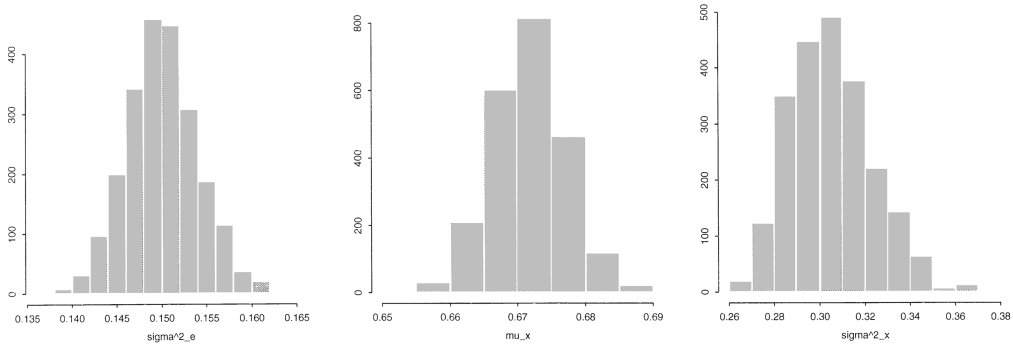

Fig. 5. Histograms of the MCMC realizations for σ_ϵ^2, μ_x, and σ_x^2.

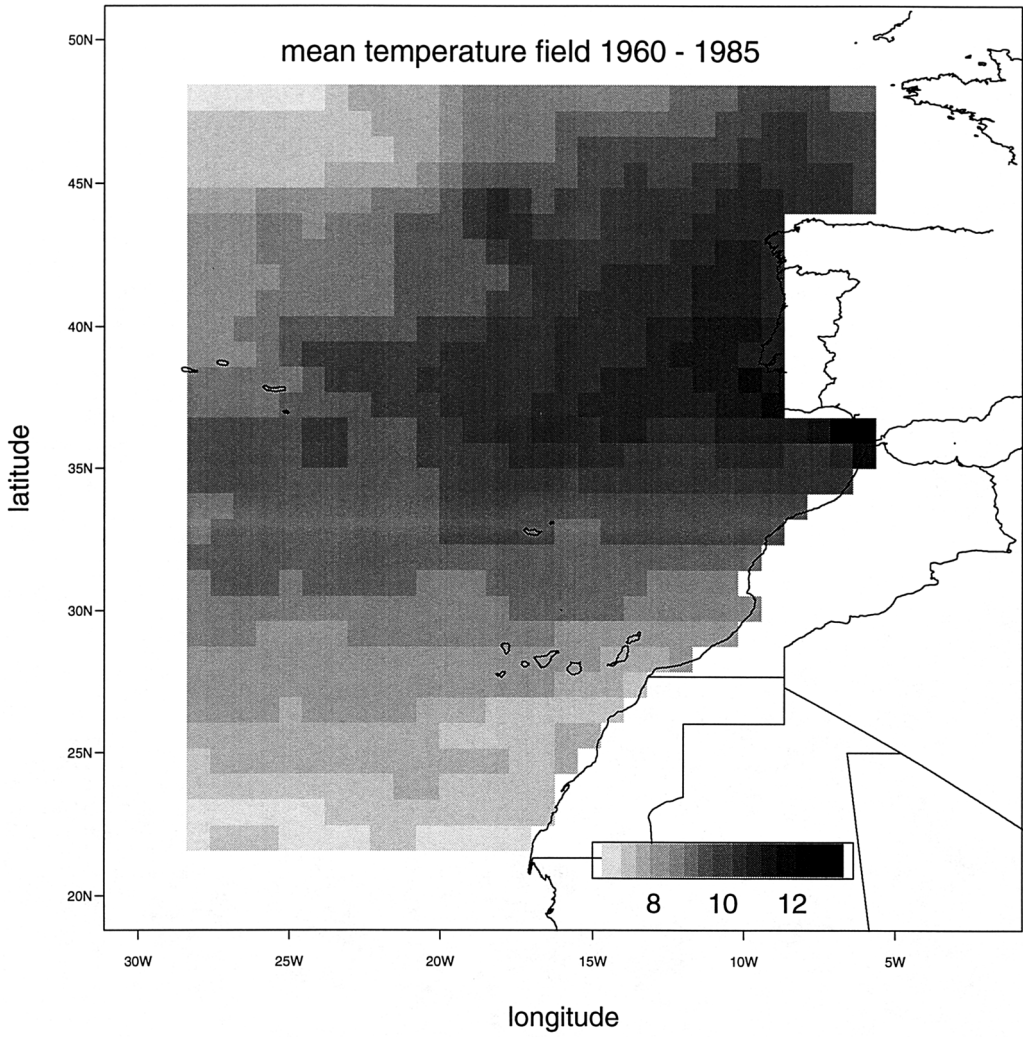

Fig. 6. Estimated mean temperature field for 1960–1985.

iterations. After this initial burn in phase, the chain is then run for 2000 iterations. From these realizations, posterior distributions, credible intervals and posterior probabilities may be computed for a variety of functionals as demonstrated in the following section.

5. Results

From the MCMC realizations, one can obtain estimates of the marginal posterior distributions for the model parameters. Those of σ_ϵ^2, μ_x and σ_x^2 are shown in Fig. 5. However what is of more interest are the realizations of the fitted temperature

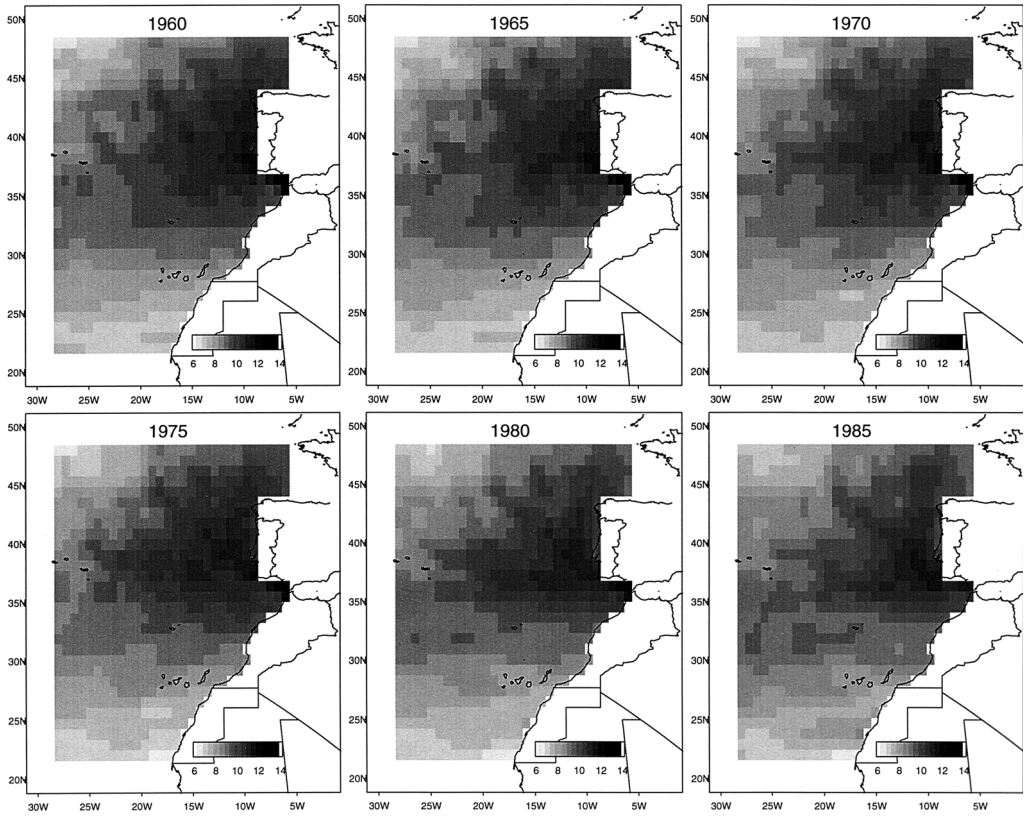

Fig. 7. Estimated mean fields for various years.

surface $z(t, s)$ which can be constructed from the x realizations using Equation (2). Although there is no restriction on the resolution of z, a grid with spacing slightly smaller than 1 degree longitude by 1 degree latitude was used to display z for each year. An alternative MCMC formulation might marginalize over x and simulate z given the remaining parameters using methodology described in Omre *et al.* (1993).

Since the data are quite scarce prior to 1960, we restrict attention to parameters associated with post-1960 data. Figure 6 gives the time averaged mean field estimate for the time period between 1960 and 1985. Unlike previous estimates, this one weighs each year equally so is not biased towards years for which proportionally more data have been collected. This field is not very different from those produced by Lozier *et al.* (1995) which do not account for uneven sampling over time. The dominant feature of this region is the Mediterranean 'tongue' of warm water which flows out of the Straits of Gibralter, hugging the southwestern Iberian coastline and travelling around the southwest cape of Portugal, Cabo de São Vicente, before spreading out both northerly and westerly.

Besides estimating the time averaged temperature field with the posterior realizations we can assess potential temperature changes over time. Figure 7 shows the estimates of the temperature fields for 1960, 1965, ..., 1985; Fig. 8 shows the difference

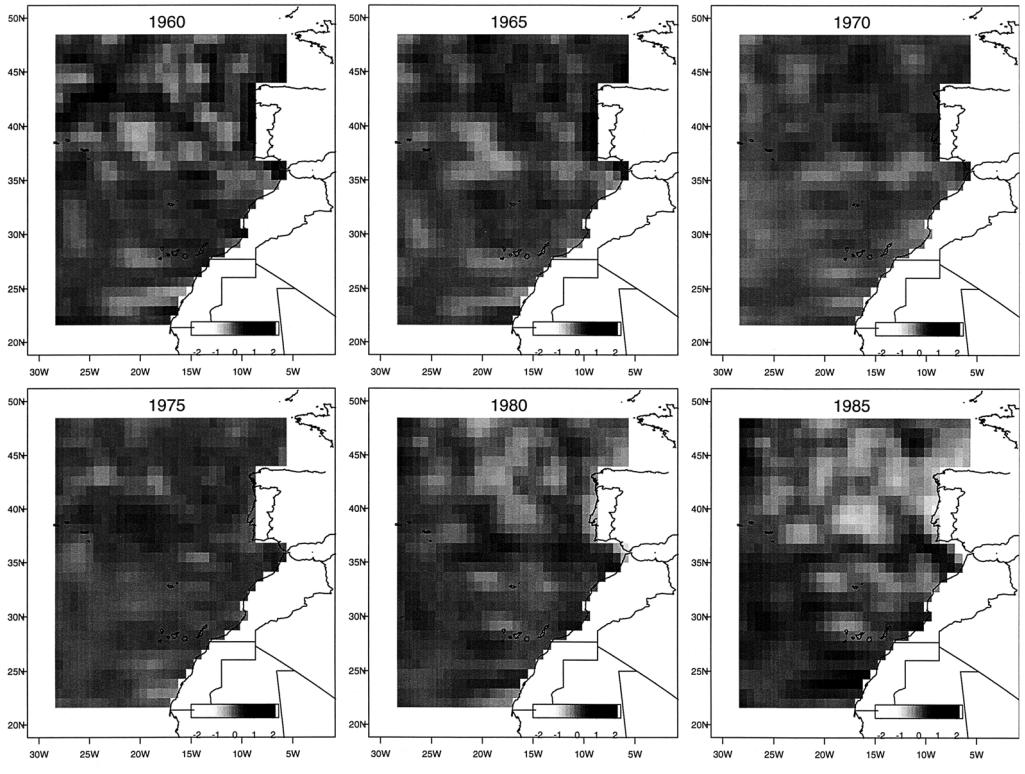

Fig. 8. Estimated deviations from the mean temperature field for 1960–1985.

between these yearly mean field estimates and the overall mean given in Fig. 6. One noticeable feature of these plots is the apparent southern drift of the Mediterranean tongue of warm water: the temperatures along the Portugal coast are cooler in 1980 and 1985; also warm water flowing outward from the Mediterranean seems to travel farther west along the 37°N latitude line in 1980 and 1985. In order to asses the statistical significance of these perceived differences the pointwise posterior probabilities that $z(s,t)$ is greater than the 1960–1985 average are given in Fig. 9. The darker regions correspond to probabilities between 1.0 and 0.95, and the lighter regions correspond to probabilities between 0 and 0.05. The shade is slightly darker if the estimated probability is exactly 1.0 or 0. Ideally, one should compute simultaneous credible surfaces as described in Besag *et al.* (1995). However, accurate estimation of these surfaces requires many more posterior realizations than is computationally feasible at present.

6. Discussion

The aim of this paper is to demonstrate how one can use the process-convolution approach to develop non-stationary space-time models for datasets that are too large for a straightforward kriging based approach. This methodology makes it possible to

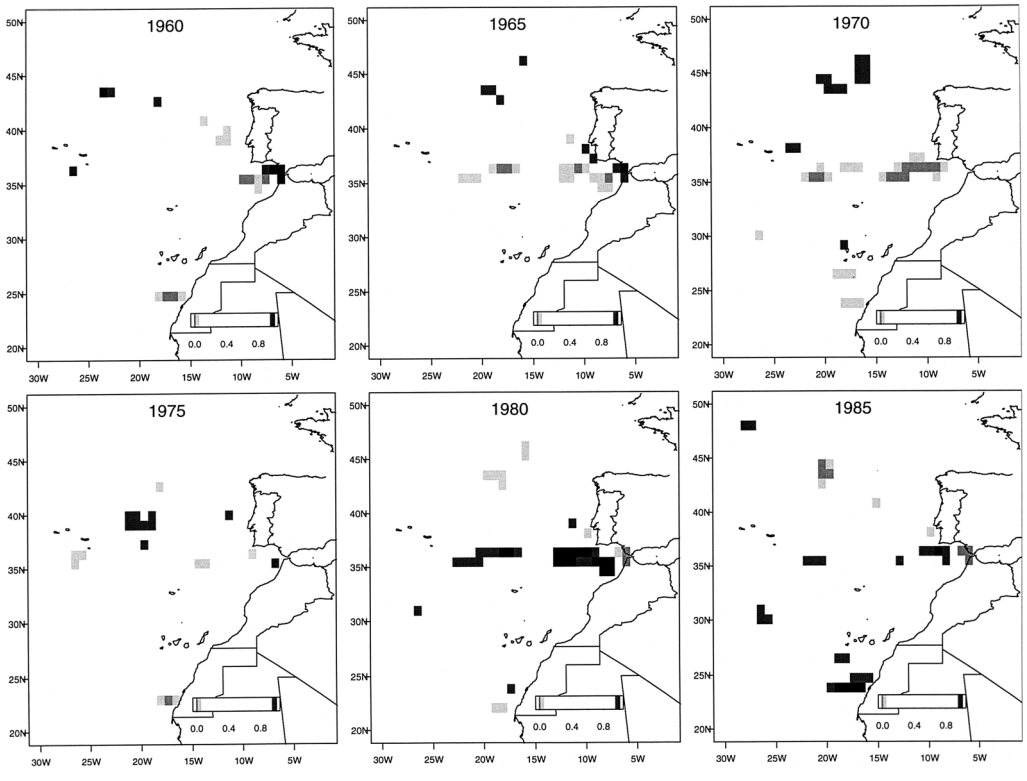

Fig. 9. Estimated pointwise posterior probabilities that each yearly mean field is greater than the 1960–1985 time averaged mean. Only probabilities less than 0.05 (light) and greater than 0.95 (dark) are shown. The slightly darker colouring marks locations for which the estimated probabilities are exactly 0 or 1.

estimate the mean temperature field using a coherent space-time model. In contrast to previous the approaches of Levitus (1982) and Lozier *et al.* (1995), this modelling effort allows:

- estimation of the mean field $z(s,t)$ over both space and time;
- data driven estimation of the amount of spatial and temporal averaging used to estimate the mean temperature field;
- quantification of uncertainty about the estimates through the posterior distribution.

Although this is a step towards a more statistically based approach for estimating the mean temperature field, there is still the need for additional exploration of various modelling issues before one can seriously regard the analysis presented here as more than preliminary.

Of the modelling issues that require additional exploration, local estimation of the convolution kernel K_s as well as its form are likely to have substantial impact on computation and properties of the spatial field $z(s,t)$ respectively. The current maximum likelihood approach for estimating the spatially dependent parameters of K_s is quite

demanding computationally. Therefore the kernel can only be estimated at a small number of locations and interpolated for remaining locations. It would be preferable to have a method that allows quick, reliable estimation of K_s at any s. Specifying K_s to have the form of a Gaussian density dictates the smoothness of $z(s,t)$. Although there is some reason to expect such a smooth field, it would be preferable to allow the data to determine the form of the kernels, which could also depend on location and/or time. In addition, the inference presented here is conditional on the estimated covariance parameters. This could potentially have an impact on the resulting posterior distribution. It is worth noting, however, that any choice of K_s that allows less smooth realizations of $z(s,t)$ will generally require a finer grid for the latent x process. Since computational considerations require as sparse a grid as possible, it may be infeasible to model much 'roughness' in $z(s,t)$.

In conclusion, some future directions for this process-convolution approach are listed below. Issues stem from both modelling and computing considerations.

- More flexible models could be used for $K_s(\Delta s, \Delta t)$. Rather than rely on the equivalence between the Gaussian variogram and a Gaussian convolution kernel, alternative approaches may be taken to estimate K_s. A non-parametric approach or one that represents K_s as a mixture of a small number of basis kernels may prove to be useful in estimating the convolution kernel K_s. It may also be preferable to allow K_s to vary with t as well as s.
- Restricting the support of K_s, along with careful algorithmic design could lead to a substantial reduction in computation time necessary to produce the MCMC output. The current implementation takes about 5 days to complete the 2000 MCMC cycles.
- The choice of 'prior' for x could easily be extended beyond a simple independence process. Perhaps the simplest alternative would be to specify x as a Markov random field with simple nearest neighbour dependencies on the lattice. This specification would allow for a long range trend and a non-stationary mean in the resulting z process while adding very little computational burden.

The above ideas are currently under consideration in ongoing research.

Acknowledgements

I thank Professor Susan Lozier for her expertise in oceanography and her many helpful suggestions. I also thank H. Jean Thiébaux and anonymous referees for their many helpful comments.

References

Barnes, S.L. (1964) A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, **3**, 396–409.

Berry, R.P. and Ver Hoef, J. (1996) Blackbox kriging: spatial prediction without specifying variogram models. *Journal of Agricultural, Biological, and Ecological Statistics*, **1**, 297–322.

Besag, J. (1986) On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society (Series B)*, **48**, 259–302.

Besag, J. and Green, P.J. (1993) Spatial statistics and Bayesian computation (with discussion). *Journal of the Royal Statistical Society (Series B)*, **16**, 395–407.

Besag, J., Green, P.J., Higdon, D.M. and Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3–66.

Brown, P.J., Le, N.D. and Zidek, J.V. (1994) Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics*, **22**, 489–509.

Cane, M.A., Clement, A.C., Kaplan, A., Kushnir, Y., Pozdnyakov, D., Seager, R., Zebiak, S.E. and Murtugudde, R. (1997) Twentieth-century sea surface temperature trends. *Science*, **275**, 957–60.

Cressie, N. A.C. (1991) *Statistics for Spatial Data*. Wiley-Interscience.

Cummings, J.A., Szczechowski, C. and Carnes, M. (1997) Global and regional ocean thermal analysis systems. *Marine Technical Society Journal*, **31**, 63–75.

Fukumori, I. and Malanotte-Rizzoli, P. (1995) An approximate Kalman filter for ocean data assimilation: an example with an idealized Gulf Stream model. *Journal of Geophysical Research*, **100**, 6777–93.

Gelfand, A.E. and Smith, A.F.M. (1990) Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 389–409.

Geyer, C.J. (1992) Practical Markov chain Monte Carlo (with discussion). *Statistical Science*, **7**, 473–511.

Gilks, W.R., Clayton, D.G., Spiegelhalter, D.J., Best, N.G., McNeil, A.J., Sharples, L.D. and Kirby, A.J. (1993) Modelling complexity: Applications of Gibbs sampling in medicine (disc: P53-102). *Journal of the Royal Statistical Society (Series B)*, **55**, 39–52.

Guttorp, P., Meiring, W. and Sampson, P.D. (1994) A space-time analysis of ground-level ozone data. *Environmetrics*, **5**, 241–54.

Handcock, M.S. and Wallis, J.R. (1994) An approach to statistical spatial-temporal modelling of meteorological fields (with discussion). *Journal of the American Statistical Association*, **89**, 368–90.

Isaaks, E.H. and Srivastava, R.M. (1990) *An Introduction to Applied Geostatistics*, Oxford University Press, New York.

Johnson, V.E. (1996) A coupling-regeneration scheme for diagnosing convergence in markov chain monte carlo algorithms. *Technical Report 96-05*. Institute of Statistics and Decision Sciences, Duke University, North Carolina.

Journel, A.G. and Huijbregts, C.J. (1979) *Mining Geostatistics*. Academic Press, San Diego.

Levitus, S. (1982) *Climatological atlas of the world ocean, Professional Paper 13*, NOAA, US Government Printing Office, Washington, DC.

Levitus, S. (1994) *World ocean atlas 1994 CD-ROM sets. Informal Report 13*. National Oceanographic Data Center, Washington, D.C.

Lozier, M.S., Owens, W.B. and Curry, R.G. (1995) The climatology of the North Atlantic. *Progress in Oceanography*, **36**, 1–44.

Matérn, B. (1986) *Spatial Variation (Second Edition)*. Springer-Verlag, New York.

Matheron, G. (1963) Principles of geostatistics. *Economic Geology*, **58**, 1246–66.

Meiring, W., Monsetiez, P., Sampson, P.D. and Guttorp, P. (1996) Developments in the modelling of nonstationary spatial covariance structure from space-time monitoring data. In E.Y. Baafi and N. Schofield (eds), *Proceedings of the 5th International Geostatistics Congress*, Kluwer.

Miller, R.N. (1986) Toward the application of the Kalman filter to regional open ocean modelling. *Journal of Physical Oceanography*, **16**, 72–86.

Le, N.D. and Sun, J.V.Z. (1997) Bayesian multivariate spatial interpolation with data missing by design. *Journal of the Royal Statistical Society (Series B)*, **59**, 501–10.

Omre, H., Sølna, K. and Tjelmeland, H. (1993) Simulation of random functions on large lattices. In A. Soares (ed.), *Geostatistics Troia '92*, pp. 179–199.

Parrilla, G., Lavin, 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.

Ripley, B.D. (1981) *Spatial Statistics*, John Wiley & Sons, New York.

Sampson, P.D. and Guttorp, P. (1992) Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, **87**, 108–19.

Sampson, P.D., Guttorp, P. and Meiring, W. (1994) Spatio-temporal analysis of regional ozone data for operational evaluation of an air quality model. *Proceedings of the American Statistical Association*.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

Smith, A.F.M. and Roberts, G.O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society (Series B)*, **55**, 3–23.

Thiébaux, H.J. (1976) Anisotropic correlation functions for objective analysis. *Monthly Weather Review*, **104**, 994–1002.

Thiébaux, H.J. (1997) The power of the duality in spatial-temporal estimation. *Journal of Climatology*, **10**, 567–573.

Thiébaux, H.J. and Pedder, M.A. (1987) *Spatial Objective Analysis: with applications in atmospheric science*. Academic Press, San Diego.

Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, **21**, 1701–1762.

Waller, L.A., Carlin, B.P., Xia, H. and Gelfand, A. (1996) Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, **92**, 607–17.

Wolpert, R.L. and Ickstadt, K. (1995) *Gamma/Poisson random field models for spatial statistics. Discussion Paper 95-43*. Institute of Statistics and Decision Sciences, Duke University, North Carolina.

Biographical sketch

Dave Higdon is an assistant professor for the Institute of Statistics and Decision Sciences at Duke University. His interests are in spatial and environmental statistics, Bayesian image analysis, computing and statistical consulting. Currently, he is the Director of the Statistical Consulting Center at Duke University.