A Spatio-Temporal Model for Mean, Anomaly and Trend Fields of North Atlantic Sea Surface Temperature

Ricardo T. Lemos and Bruno Sansó *

Abstract

We consider the problem of fitting a statistical model to thirty years of sea surface temperature records collected over a large portion of the Northern Atlantic. The observations were collected sparsely in space and time with different levels of accuracy. The purpose of the model is to produce an atlas of oceanic properties, including climatological mean fields, estimates of historical trends and a spatio-temporal reconstruction of the anomalies, i.e., the transient deviations from the climatological mean. These products are of interest to climate change and climate variability research, numerical modeling and remote sensing analyzes. Our model improves upon the current tools used by oceanographers in that it constructs instantaneous temperature fields prior to averaging them into the climatology, thus giving equal weight to all years in the time frame, regardless of the temporal distribution of data. It also accounts for non-isotropic and non-stationary space and time dependencies, owing to its use of discrete process convolutions. Particular attention is given to the handling of massive data sets such as the one under study. This is achieved by considering compact support kernels that allow an efficient parallelization of the Markov chain Monte Carlo method used in the estimation of the model parameters. Resulting monthly climatologies are compared with those of the World Ocean Atlas 2001, version 2. Different water masses appear better separated in our climatology, and a close link emerges between the kernels' shape and the dominating patterns of ocean currents. The subpolar and the temperate North Atlantic display opposite trends, with the former mainly cooling over the years and the latter mainly warming, especially in the Gulf Stream region. Long-term changes in annual cycles are also detected. As in any hierarchical Bayesian model, parameter estimates come with credibility intervals, which are useful to compare results with other approaches and detect areas where sampling campaigns are needed the most.

KEY WORDS: CLIMATOLOGY, DISCRETE PROCESS CONVOLUTIONS, PARALLEL PROCESSING, MCMC

*Ricardo Lemos is a graduate student at Instituto de Oceanografia, Universidade de Lisboa, and Maretec - Instituto Superior Técnico, Universidade Técnica de Lisboa, Av. Manuel da Maia, 36, 3° esq., 1000-201 Lisboa. E-mail rtl00@yahoo.com, Bruno Sansó is Associate Professor, Department of Applied Mathematics and Statistics, University of California, 1156 High St. MS:SOE2, Santa Cruz, CA-95064, U.S.A. E-mail bruno@ams.ucsc.edu, www.ams.ucsc.edu/∼bruno

1 Introduction

The World Ocean has long been sampled by scientific and military cruises, voluntary observing ships, fixed observatories and buoys. Initially, databases constructed thereof had one major purpose: to depict the mean (*i.e.*, climatological) state of the ocean, so as to understand the properties, distribution and circulation of water masses, and to identify forcing mechanisms. Today, climatological atlases remain a basic tool for this end, but many new uses have emerged, including remote sensor calibration and the spin up, forcing, relaxation and validation of numerical models. As climate change and climate variability research developed, another focal point became the variability around the mean (*i.e.*, the anomalies), on various spatial and temporal scales. Detecting long-term changes in ocean properties, especially temperature, became a subject of intense research (Kushnir, 1994; Casey and Cornillon, 2001; Gouretski and Koltermann, 2007).

Objective analysis (OA) is the most used method to produce ocean climatologies. Boyer et al. (2005) summarize OA as a calculation of mean fields at each grid square, based on a weighted difference between the means at all grid squares within a given radius of influence around a grid point and a first-guess field at the same grid square. For the first-guess field, earlier climatologies or gridded averages of raw data are used. The weighting kernel is problematic to define, because it depends on the covariance structure of the property, which is unknown. Usually, isotropy and stationarity are assumed, along surfaces of either constant depth (Reynolds et al., 2002; Boyer et al., 2005) or constant density (Lozier et al., 1995; Gouretski and Janke, 1999). Via exploratory data analysis, the decorrelation length is calculated and the kernel's shape is fixed accordingly. The number of OA passes varies among applications, and *post hoc* smoothing is often needed to eliminate bull's eyes. In the final product, uncertainty about the kernel's shape, the first-guess field and the adequate number of passes is not included. Further, isotropy and stationarity assumptions generally do not hold for ocean properties. Extensions and alternatives to OA include Empirical Orthogonal Functions analyzes (Holbrook and Bindoff, 2000), variational analyzes (Brasseur et al., 1996) and anisotropic loess smoothers (Ridgway et al., 2002), which mitigate some but not all of these issues.

In the last 15 years, several methods were developed to estimate parameters in non-stationary anisotropic models, most of which considering Gaussian processes. The idea of producing a non-linear transformation of space to achieve isotropy was popularized in the early 1990s, following the work of Sampson and Guttorp (1992). This approach was originally developed using cross-validation and then extended to include likelihood-based methods (Damian et al., 2001; Schmidt and O'Hagan, 2003). Another approach used multivariate normal models to estimate general covariance matrices (Brown et al., 1995), and a third built globally anisotropic processes from convolutions of locally isotropic ones. This was done by considering processes with either spatially varying covariance parameters (Fuentes, 2002) or spatially varying convolving kernels (Higdon et al., 1999). A related technique created large classes of non-stationary covariance functions using convolutions (Paciorek and Schervish, 2006).

Sophisticated statistical methods are increasingly applied to massive data sets obtained e.g. from satellite images. Explicit computations of covariance matrices, which correspond

to Gaussian processes observed at millions of locations, are usually avoided, since decomposing and even storing them may be impossible. Spectral representations or multiresolution methods are preferred, but as they require the data to be on regular grids, data aggregation or statistical imputation of missing values must be considered. Recent examples are Nychka et al. (2002), Tzeng et al. (2005), Fuentes (2007), Johannesson et al. (2007) and Paciorek (2007). Atmospheric scientists have developed a body of literature on covariance structures for Gaussian random fields; the focus is mostly on the properties of classes of correlation functions and the methods to compute them on very large regular grids, with little emphasis on estimation procedures. A good review of the approaches traditionally used in atmospheric data assimilation problems is presented in Xun (2005), and a seminal paper in the area is Gaspari and Cohn (1999).

In this paper we consider the problem of creating climatological, anomaly and long-term linear trend fields of sea surface temperature (SST) in the North Atlantic (14–66°N, 0–100°W). Among other features, this part of the World Ocean combines regions with abundant and scarce sampling, includes eastern and western boundary current systems, and presents strong spatial and temporal variability at the sea surface. It thus provides an adequate testbed for global scale projects, encompassing many ocean properties.

The goals of our application are the following: a) The climatology must be smooth and realistic, when compared to other existing products, namely the 1/4° World Ocean Atlas 2001, version 2 (WOA; Boyer et al., 2005), which is presently the standard climatological product from the National Oceanographic Data Center (NODC); b) anomaly fields must capture medium- to large-scale features and average zero everywhere; c) the trend field must be smooth; d) all fields must be accompanied by measures of uncertainty; e) the method must be easily extended to a larger data set and/or geographical domain.

For our approach, we use data from the NODC World Ocean Database 2005 (WOD05, Boyer et al., 2006), collected with four types of instruments between 1961 and 1990. Screening was performed using quality control criteria similar to those of WOD05, resulting in 1,150,097 valid observations. In the tradition of products used for data based descriptions of the climate, our model is purely empirical and of general purpose. It is based on the representation of a Gaussian process as the convolution of a process with a kernel, as proposed in Higdon (2002). If white noise is used in the convolution, then the covariance of the resulting process is fully determined by the kernel. A discrete approximation of the convolution integral is obtained by sampling the convolved process on a grid. This is the motivating idea to model a random process for any point $\boldsymbol{s}$ in space S, say $\theta(\boldsymbol{s})$, as

$$\theta(\boldsymbol{s}) = \sum_{\boldsymbol{j} \in J} K[\boldsymbol{s} - \boldsymbol{j}, \boldsymbol{\omega}(\boldsymbol{s})] \psi(\boldsymbol{j}) \tag{1}$$

where J is a grid in S with n_J points spaced r_J units apart. $K[\cdot, \boldsymbol{\omega}]$ is a normalized anisotropic kernel that depends on a vector of parameters $\boldsymbol{\omega}$, and $\psi(\boldsymbol{j})$ is a random field with a simple correlation structure. The depence of $\boldsymbol{\omega}$ on the location of the kernel allows for non-stationary correlation structures. We term processes like the one in Equation (1), discrete process convolutions (DPCs). When the dimension of J is small, DPCs provide an effective way of reducing the computational burden required for inference, since only n_J locations need to be considered. Additionally, no imputations or aggregations need to be performed on

irregularly located data.

This paper is organized as follows. Section 2 provides the most important aspects of our notation. Section 3 describes DPCs into greater depth, focusing especially on the anisotropic kernel. Section 4 presents our hierarchical Bayesian model, and shows how it is used to produce climatological, anomaly and trend fields of North Atlantic SST, for the period 1961-1990. In Section 5 we specify how the model is fitted, and in Section 6 we present and discuss the most important results. Finally, in Section 7, we return to the subjects introduced here and consider future modeling endeavors.

2 Conventions and notation

Model parameters and derived quantities are written with lower case and upper case greek letters, respectively. Vectors are in boldface and their elements can be singled with subscripts, e.g. Λ_1. The superscripts $^{-1}$ and ^{t} denote inverse and transpose. S is a 2D space and J is a regular grid in S, with n_J points spaced r_J units apart. Points in S and J are respectively denoted by $\boldsymbol{s} \equiv (x_s, y_s)$ and $\boldsymbol{j} \equiv (x_j, y_j)$. Variance parameters are designated with τ^2. All other parameters have time as subscript and spatial location in parentheses e.g., $\beta_t(\boldsymbol{j})$. In some cases, the time component is decomposed into month m and year y. If the parameter does not change over the years, only the month subscript is used.

3 Discrete Process Convolutions

To model processes that vary smoothly over S, we define a grid J and assign a latent process to each point $\boldsymbol{j} \in J$. As stated in Equation (1), DPC modeling consists of equating the process of interest, at location $\boldsymbol{s}$, to a normalized weighted sum of the latent processes. An error term may or may not appear in the equation. The weight assigned to each process depends on $\boldsymbol{s}$, and its unnormalized value, which we denote with an asterisk, is in our approach provided by the kernel

$$K^*[\boldsymbol{s}-\boldsymbol{j}, \boldsymbol{\omega}] \equiv \begin{cases} (1 - ||\boldsymbol{s}-\boldsymbol{j}||^2_{\boldsymbol{\Sigma}})^{\omega_1} & \text{if } ||\boldsymbol{s}-\boldsymbol{j}||_{\boldsymbol{\Sigma}} < 1 \\ 0 & \text{otherwise.} \end{cases}$$

where $\boldsymbol{\omega} \in \mathbb{R}^4$ and

$$||\boldsymbol{s}-\boldsymbol{j}||_{\boldsymbol{\Sigma}} \equiv \sqrt{((x_s - x_j), (y_s - y_j))\, \boldsymbol{\Sigma}^{-1}\, ((x_s - x_j), (y_s - y_j))^t}. \tag{2}$$

The inverse of the symmetric and positive definite matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\omega}) \in \mathbb{R}^{2\times 2}$ is given by

$$\boldsymbol{\Sigma}^{-1} \equiv \begin{pmatrix} \Psi_1 + \Psi_2 \cos 2\omega_4 & \Psi_2 \sin 2\omega_4 \\ \Psi_2 \sin 2\omega_4 & \Psi_1 - \Psi_2 \cos 2\omega_4 \end{pmatrix}, \quad \boldsymbol{\Psi} = \frac{1}{2}\left(\frac{1}{\omega_2^2} + \frac{1}{\omega_3^2}, \frac{1}{\omega_2^2} - \frac{1}{\omega_3^2}\right).$$

Conditions for the positive definiteness of $\boldsymbol{\Sigma}$ will be given further below. In Section 5 we describe the computational advantages derived from using a kernel with compact support. Here, we focus on the kernel's flexibility for DPC modeling; see Figure 1 for an example.

If we provide $\omega_1 \in (1.5, 5)$, the kernel's shape ranges from triangular- to Gaussian-like (Figure 1b). Values outside this range yield jagged convolution surfaces, which are of no interest to us. If we set $\omega_2 = \omega_3$, then $\Psi_2 = 0$; $\boldsymbol{\Sigma}$ becomes space-invariant and proportional to the 2×2 identity matrix, and the distance between $\boldsymbol{s}$ and $\boldsymbol{j}$ becomes proportional to the Euclidean distance. Thus, we obtain an isotropic kernel (termed "Bézier kernel" by Brenning, 2001). To explore anisotropic modeling, we let $\omega_4 = 0$ for a moment. By equating expression (2) to 1, we obtain

$$\frac{(x_j - x_s)^2}{\omega_2^2} + \frac{(y_j - y_s)^2}{\omega_3^2} = 1, \tag{3}$$

which defines an ellipse centered at $\boldsymbol{s}$. Thus, if $0 < \omega_2 \leq \omega_3$, ω_2 and ω_3 define the sizes of the ellipse's semiminor and semimajor axes, respectively. If we now let ω_4 vary in the interval $(-\pi/2, \pi/2)$, the ellipse becomes inclined; the angle between ω_2 and the x-axis corresponds to ω_4. Hence, points of J at the same Euclidean distance to $\boldsymbol{s}$ may have substantially different contributions to the DPC. For example, in Figure 1a, the DPC for a parameter located at $\boldsymbol{s}$ would result from averaging the latent processes at $\boldsymbol{j}$ and $\boldsymbol{j}'$ only.

Small or oblong kernels, centered at some points in S, may have no points of J inside their support. This produces an invalid DPC because the weights amount to zero. We avoid this possibility by ensuring that, for every $\boldsymbol{s}$, at least its nearest neighbor in J falls within the support. Since we are using a regular grid, the maximum Euclidean distance between these two points is $r_J/\sqrt{2}$, as depicted in Figure 1a. Hence, a sufficient condition to obtain a valid DPC is that $\omega_2 > r_J/\sqrt{2}$. On the other hand, we may consider that the behavior of the parameter of interest, for any location in S, can be reasonably described by a weighted average of at most the 4 nearest neighboring processes located in J (this approach also brings computational benefits, discussed in Section 5). Since the minimum Euclidean distance between $\boldsymbol{s}$ and its fifth nearest neighbor in J is r_J, we ensure that this point never belongs to the kernel's support if $\omega_3 < r_J$. In summary, we let $r_J/\sqrt{2} < \omega_2 \leq \omega_3 < r_J$. This guarantees the positive definiteness of $\boldsymbol{\Sigma}$.

4 Model

The hierarchical model presented here has two spatial levels: local and regional, respectively defined in S and J (Figure 2). The temporal resolution is one month. At the local level, we describe an observation of SST collected with instrument $i = 1, \ldots, 4$ (see Table 1), in month m, year y and location $\boldsymbol{s}$, say $x_{i,m,y}(\boldsymbol{s})$, as a sum of true SST, $\theta_{m,y}(\boldsymbol{s})$, and measurement error with variance τ_i^2, thus

$$x_{i,m,y}(\boldsymbol{s}) \sim N\left(\theta_{m,y}(\boldsymbol{s}), \tau_i^2\right).$$

Given the absence of repeated measurements in the data sets, we use the accuracy reports of Boyer et al. (2006) to fix the error variances of measurements made with high resolution conductivity, temperature and depth (CTD) instruments, expendable bathythermographs (XBT), and mechanical bathythermographs (MBT). Assigning an error variance to ocean station data (OSD) is more difficult, because this data set comprises measurements made

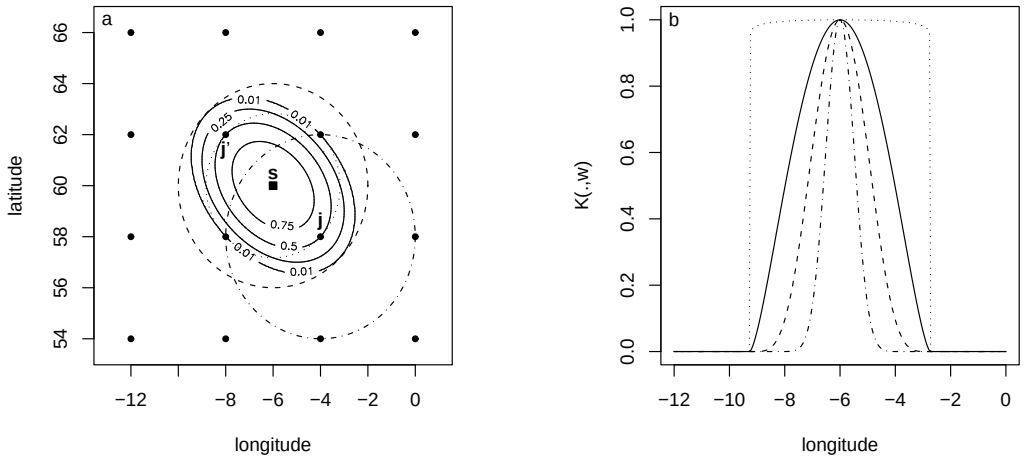

Figure 1: The convolution kernel for $\boldsymbol{s} = (-6, 60)$ and $\boldsymbol{\omega}(\boldsymbol{s}) = (1.5, 2\sqrt{2}, 4, \pi/4)$. a) Top view. The kernel's contours are depicted by the solid lines. The dashed (dotted) circle delimits the largest (smallest) possible support for any kernel centered at $\boldsymbol{s}$. The bullets are points of J, and the dashed-dotted circle delimits the area where the DPC weights involving $\boldsymbol{j}$ may be non-zero. b) Side view (solid line), latitude=60. Kernel shapes for ω_1 equal to 0.01 (dotted line), 5 (dashed line), and 20 (dotted-dashed line) are also shown.

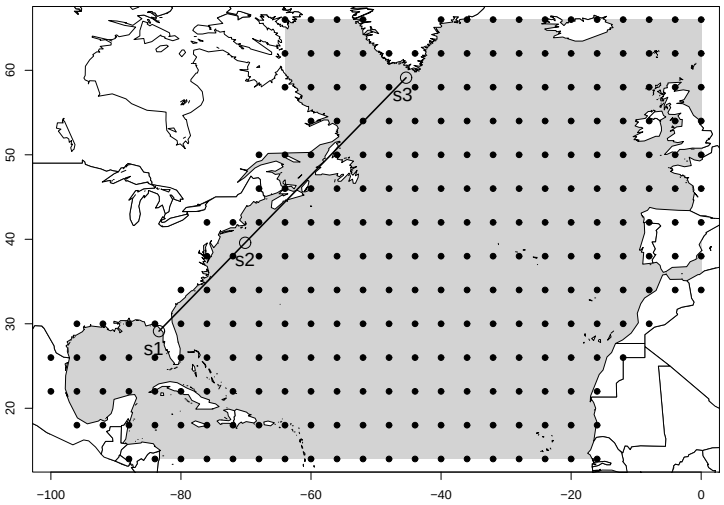

Figure 2: The domain S (gray area) and the grid J (bullets; $r_J = 4°$). The line and open circles denote a transect and three "case study" points.

Data set name	no. obs.	τ_i^2
OSD	261,172	6.25×10^{-2}
CTD	29,879	6.25×10^{-6}
XBT	419,263	2.5×10^{-3}
MBT	439,783	2.025×10^{-1}

Table 1: NODC data sets and observational variances. See text for acronyms.

with a variety of instruments (*viz.* self-closing sample bottles and calibrated thermometers on cables, lowered from stationary ships). Based on a survey of the general accuracy of these instruments, we set the observational standard error for OSD equal to 0.25.

Under the formulation above, the climatological SST for month m and location $\boldsymbol{s}$, denoted $\Xi_m(\boldsymbol{s})$, is

$$\Xi_m(\boldsymbol{s}) = \frac{1}{30} \sum_{y=1961}^{1990} \theta_{m,y}(\boldsymbol{s}),$$

and the anomaly for month m, year y and location $\boldsymbol{s}$, denoted $\Delta_{m,y}(\boldsymbol{s})$, is

$$\Delta_{m,y}(\boldsymbol{s}) = \theta_{m,y}(\boldsymbol{s}) - \Xi_m(\boldsymbol{s}).$$

To model θ using a DPC we use the following regional scale parameters: α, which describes the average annual SST; the vector $\boldsymbol{\beta}_t \in \mathbb{R}^4$, which captures the annual cycle and transient deviations thereof; and η, which accounts for possible long-term changes in SST. We also characterize the resulting error by means of a DPC of the regional parameter σ. This yields

$$\theta_{m,y}(\boldsymbol{s}) \sim N\left(\sum\nolimits_{\boldsymbol{j}} K[\boldsymbol{s}-\boldsymbol{j}, \boldsymbol{\Lambda}(\boldsymbol{s})] \left(\alpha(\boldsymbol{j}) + \boldsymbol{\beta}_t(\boldsymbol{j})\boldsymbol{w}_t^T + \eta(\boldsymbol{j})(t-180)\right), \sum\nolimits_{\boldsymbol{j}} K[\boldsymbol{s}-\boldsymbol{j}, \boldsymbol{\omega}(\boldsymbol{s})] \exp\left(\sigma(\boldsymbol{j})\right)\right). \quad (4)$$

In this expression, $t = m + 12(y - 1961)$ denotes time in months since December 1960 and

$$\boldsymbol{w}_t = \left(\sin\left(\frac{2\pi t}{12}\right), \cos\left(\frac{2\pi t}{12}\right), \sin\left(\frac{2\pi t}{6}\right), \cos\left(\frac{2\pi t}{6}\right)\right).$$

For future reference, the standard deviation of monthly $\theta_{m,y}(\boldsymbol{s})$ will be denoted as $\Phi(\boldsymbol{s})$. We let the four dimensional kernel vectors $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ change across space, according to

$$\boldsymbol{\Lambda}(\boldsymbol{s}) = \sum\nolimits_{\boldsymbol{j}} K[\boldsymbol{s}-\boldsymbol{j}, \boldsymbol{u}]\boldsymbol{\kappa}(\boldsymbol{j}),$$

$$\boldsymbol{\Omega}(\boldsymbol{s}) = \sum\nolimits_{\boldsymbol{j}} K[\boldsymbol{s}-\boldsymbol{j}, \boldsymbol{u}]\boldsymbol{\rho}(\boldsymbol{j}),$$

to account for location-dependent anisotropy in the mean and variance of SST. In these expressions, the vector $\boldsymbol{u}$ is fixed at $(2, r_J, r_J, 0)$. The parameters in $\boldsymbol{\beta}$ follow independent random walks,

$$\boldsymbol{\beta}_t \sim N\left(\boldsymbol{\beta}_{t-1}, \boldsymbol{W}_t\right) \quad (5)$$

Parameter	Description	var. param.
$\theta_{m,y}(\boldsymbol{s})$	True SST	–
$\boldsymbol{\tau}_i^2$	SST measurement variance (fixed)	–
$\boldsymbol{\kappa}(\boldsymbol{j})$	Kernel shape for mean SST	–
$\alpha(\boldsymbol{j})$	Annual mean SST	τ_α^2
$\boldsymbol{\beta}_t(\boldsymbol{j})$	Seasonal cycle	$\boldsymbol{\tau}_{\beta_1}^2$
$\eta(\boldsymbol{j})$	Long-term linear trend	τ_η^2
$\sigma(\boldsymbol{j})$	SST variance	τ_σ^2
$\nu(\boldsymbol{j})$	Discount factor for the cycle	–
$\boldsymbol{\rho}(\boldsymbol{j})$	Kernel shape for SST variance	–

Table 2: Model parameters.

Quantity	Description	Derived from
$\Xi_m(\boldsymbol{s})$	Monthly climatological SST	$\theta_{m,y}(\boldsymbol{s})$
$\Delta_{m,y}(\boldsymbol{s})$	Monthly SST anomaly	$\theta_{m,y}(\boldsymbol{s})$
$\boldsymbol{\Lambda}(\boldsymbol{s})$	Kernel shape for mean SST	$\boldsymbol{\kappa}(\boldsymbol{j})$
$\boldsymbol{\Omega}(\boldsymbol{s})$	Kernel shape for SST variability	$\boldsymbol{\rho}(\boldsymbol{j})$
$\Phi(\boldsymbol{s})$	Standard deviation of monthly SST	$\sigma(\boldsymbol{j}), \boldsymbol{\rho}(\boldsymbol{j})$

Table 3: Derived quantities.

with the discount factor $\nu(\boldsymbol{j})$ applied to the diagonal evolution variance matrix $\boldsymbol{W}$ (West and Harrison, 1997, Chapter 2). To provide some spatial coherence in the variability of α, we use a Gaussian Markov random field (MRF). Thus,

$$\alpha(\boldsymbol{j}) \sim N\left(\frac{\alpha(\mathrm{N}(\boldsymbol{j})) + \alpha(\mathrm{S}(\boldsymbol{j})) + \alpha(\mathrm{E}(\boldsymbol{j})) + \alpha(\mathrm{W}(\boldsymbol{j}))}{4}, \tau_\alpha^2\right).$$

Here, $\mathrm{N}(\boldsymbol{j})$ denotes the point in J immediately to the north of $\boldsymbol{j}$, and so forth. We couple $\boldsymbol{\beta}_1$, η and σ to analogous MRF structures. Tables 2 and 3 contain a list and short description of all the model's parameters and derived quantities, respectively.

5 Model implementation and assessment

For τ_α^2, $\boldsymbol{\tau}_{\beta_1}^2$ and τ_σ^2, we use inverse Gamma$(1, 0.01)$ prior distributions, as they provide little prior information. To ensure the trend field is smooth, τ_η^2 receives a narrower, but still vague, inverse Gamma$(2, 0.002)$ prior. We use a Uniform $(0.9, 1)$ prior for ν, as this support covers the factor's natural range of variability. The parameters κ_1 and κ_4 receive Uniform priors with support $(1.5, 5)$ and $(-\pi/2, \pi/2)$, respectively; the joint prior for κ_2 and κ_3 is proportional to $I_{r_J/\sqrt{2}<\kappa_3\leq\kappa_2<r_J}$, where I is the indicator variable. We assign analogous priors to $\boldsymbol{\rho}$.

We explore the posterior joint distribution of all the parameters by Markov chain Monte Carlo (MCMC; see, e.g., Gamerman and Lopes, 2006). Closed form full conditional distributions for the Gibbs sampler are provided in the Appendix. We sample from the joint

distribution of $(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{360})$ using a forward filtering, backward sampling algorithm (West and Harrison, 1997, Chap. 15) applied to the conditional multivariate dynamic linear models with observation equation obtained from (4) and evolution equation obtained from (5). We update the samples of $\boldsymbol{\kappa}$, $\boldsymbol{\rho}$, ν and σ with Metropolis-Hastings steps, using truncated normals as jumping distributions and setting their variances upon a pilot run of 3000 iterations.

The model described above was defined with a moderate degree of complexity in terms of spatial and temporal relationships among parameters, while allowing for smooth variations of expectations across all dimensions. Our use of a kernel with compact support is essential to obtain substantial simplifications in discrete process convolutions: all summations involving $\boldsymbol{j}$ are reduced to summations of at most 4 terms, since the kernel weight (K) is non-zero for at most 4 points in J. More importantly, when updating any parameter associated with $\boldsymbol{j}$, only the observations within the circle centered at $\boldsymbol{j}$ and radius r_J can have direct influence on the likelihood (Figure 1a). Thus, in the MCMC method we can make efficient use of a parallel computing architecture. We divide S into a number of overlapping regions, partition J accordingly, and let a different computing process sample only the parameters within its region of S and J. Because regions are not independent, each process must have updated information about parameters in adjacent regions. A scheme of the procedure for the simplest network possible (2 processes) is presented in Figure 3 and Table 4. The results presented in this paper are obtained using 13 processes, so that each works with two columns of J.

For convergence diagnostics, we use the methods developed by Heidelberger and Welch (1983), Gelman and Rubin (1992), Geweke (1992), Raftery and Lewis (1992a,b) and Brooks and Gelman (1998), which are available in the package Bayesian Output Analysis Program (`BOA`) (Smith, 2005) within `R` (R Development Core Team, 2005). With `BOA` we define the length of the burn-in stage (1200 iterations), thin the chain (1/3), check stationarity and set the adequate sample size (6000 iterations from the thinned chain) to achieve the precision required in the estimation of 95% posterior intervals. To check that the initial conditions are irrelevant to the posterior distributions, we compare two runs where one starts with a warm ocean ($\alpha(\boldsymbol{j}) = 30$, for all $\boldsymbol{j}$), low discount factors and high variance parameters, while the other starts with a cooler ocean ($\alpha(\boldsymbol{j}) = 15$) and average discount factors and variances. Both start with no seasonality or trends.

Finally, we assess lack of fit of the data with respect to the posterior predictive distribution by searching for noticeable instrument dependent, spatial or temporal patterns in Bayesian p-values (Gelman et al., 2004, Chap. 6)

$$p^B_{i,m,y}(\boldsymbol{s}) = \Pr\left(x^{rep}_{i,m,y}(\boldsymbol{s}) \geq x_{i,m,y}(\boldsymbol{s}) | \boldsymbol{\Theta}\right).$$

In this expression, $x^{rep}_{i,m,y}(\boldsymbol{s})$ is a replicated datum, simulated from the posterior predictive distribution, $x_{i,m,y}(\boldsymbol{s})$ is an actual SST measurement, and $\boldsymbol{\Theta}$ denotes the collection of all parameters in the model.

6 Results

Figure 4 compares North Atlantic climatologies produced with our method and WOA. In terms of the means for January and July (Figures 4a through d), the two products display

Step	Process 1	Process 2
1	Sample parameters in columns 1-12 of J	Sample parameters in columns 14-25 of J
2	Receive samples of parameters from column 14 of J	Send samples of parameters from column 14 of J
3	Sample parameters in column 13 of J	Sample parameters in column 26 of J
4	Send samples of parameters from column 13 of J	Receive samples of parameters from column 13 of J
5	Sample parameters in the area $100° - 48°$W of S	Sample parameters in the area $48° - 0°$W of S
6	Send samples of parameters in the area $52° - 48°$W of S	Receive samples of parameters in $52° - 48°$W of S

Table 4: MCMC cyclical procedure for the setting displayed in Figure 3.

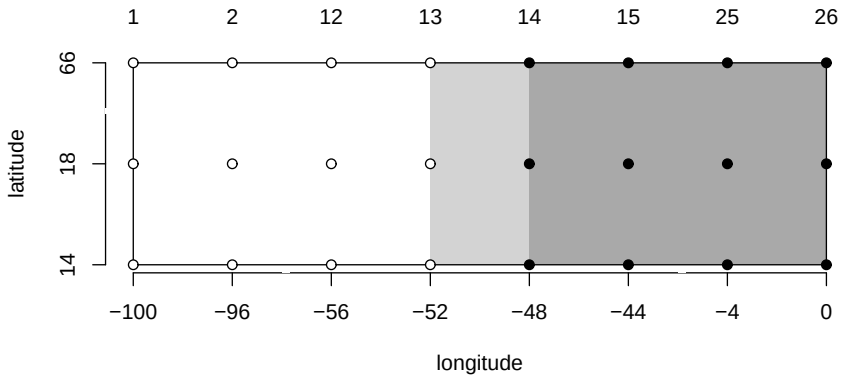

Figure 3: In this example, S is divided into two regions that overlap between 52°W and 48°W. Empty (filled) points of J are updated by computing process 1 (2). The columns of J are numbered over S to describe the MCMC algorithm (Table 4).

the same large scale patterns, but our presents more sinuous contours, especially in the northwest Atlantic. Along the transect drawn in Figure 2, other differences become clear (Figure 4e): our climatology presents sharper SST gradients between water masses separated by land, *viz.* the Florida peninsula and Nova Scotia (which cause the gaps around 30°N and 45°N), and differentiates regions with constant temperature (e.g., 37–40°N, 41–43°N) from others with sharp gradients (e.g., 40–41°N). Throughout the study area, WOA generally presents smoother gradients and less defined plateaus. With respect to the seasonal cycle, the three points in the transect show distinctive cases (Figure 4f): while in $\boldsymbol{s}_3$ there is a fairly good agreement between the two products, $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ present marked disparities. In $\boldsymbol{s}_2$ the WOA cycle is systematically cooler than ours, while in $\boldsymbol{s}_1$ the opposite occurs, but uncertainty allows for compatibility of results in some months.

The inclusion of a long-term linear trend in the model reveals a bipolar pattern, with the midlatitudes mostly warming and the subpolar regions mostly cooling (Figure 5a). The extremes of these trends are found in the Northwest: between 1961 and 1990, SST warmed

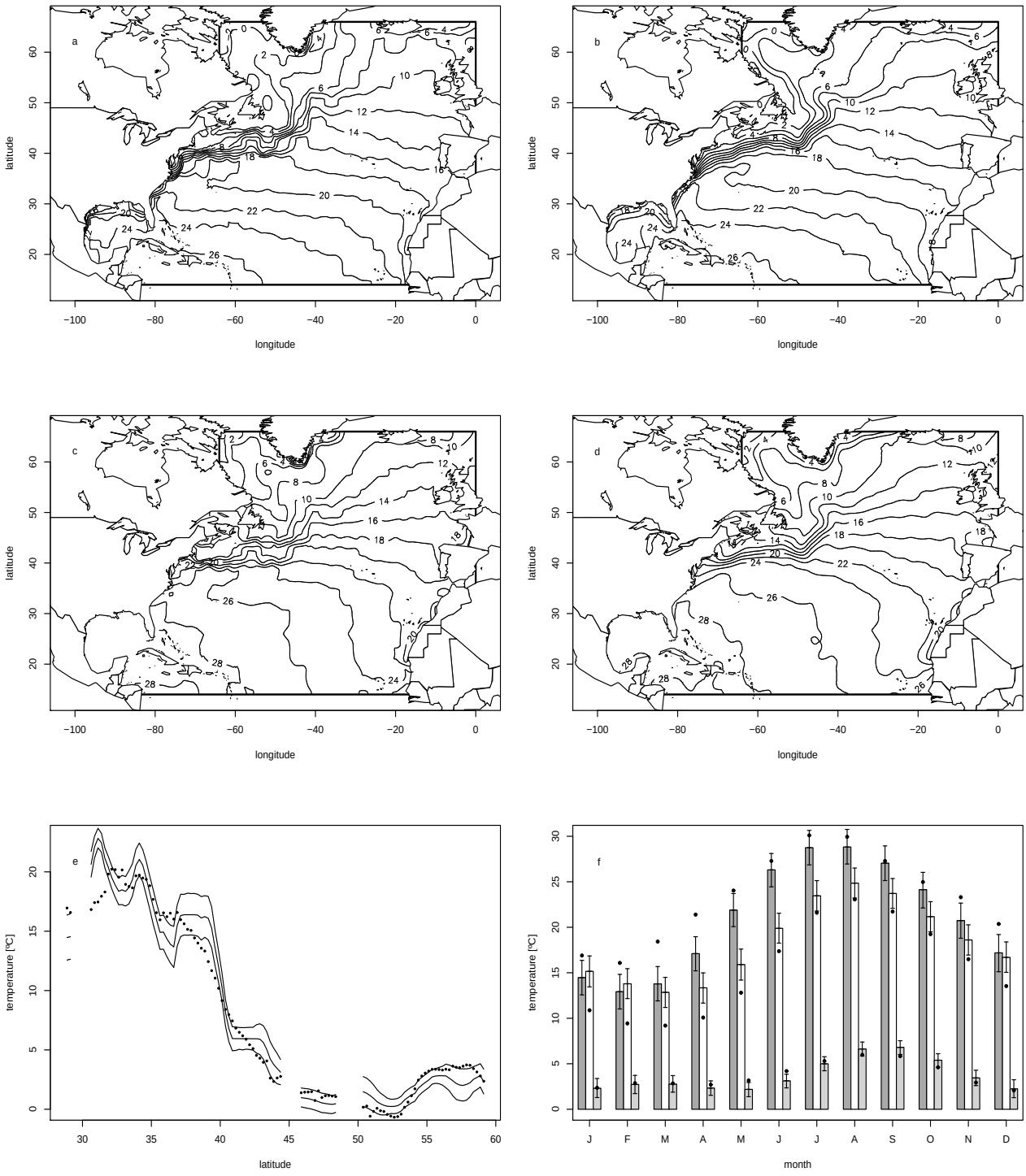

Figure 4: SST climatologies. a) January mean for the North Atlantic, results from our model. b) WOA January mean. c) July mean, results from our model. d) WOA July mean. e) January climatology along the transect depicted in Figure 2. Lines denote mean and 95% posterior intervals from our model, bullets denote WOA. f) Seasonal cycle in the points depicted in Figure 2. Rectangles (dark gray for $\boldsymbol{s}_1$, white for $\boldsymbol{s}_2$ and light gray for $\boldsymbol{s}_3$) and whiskers denote means and 95% posterior intervals from our model, bullets denote WOA.

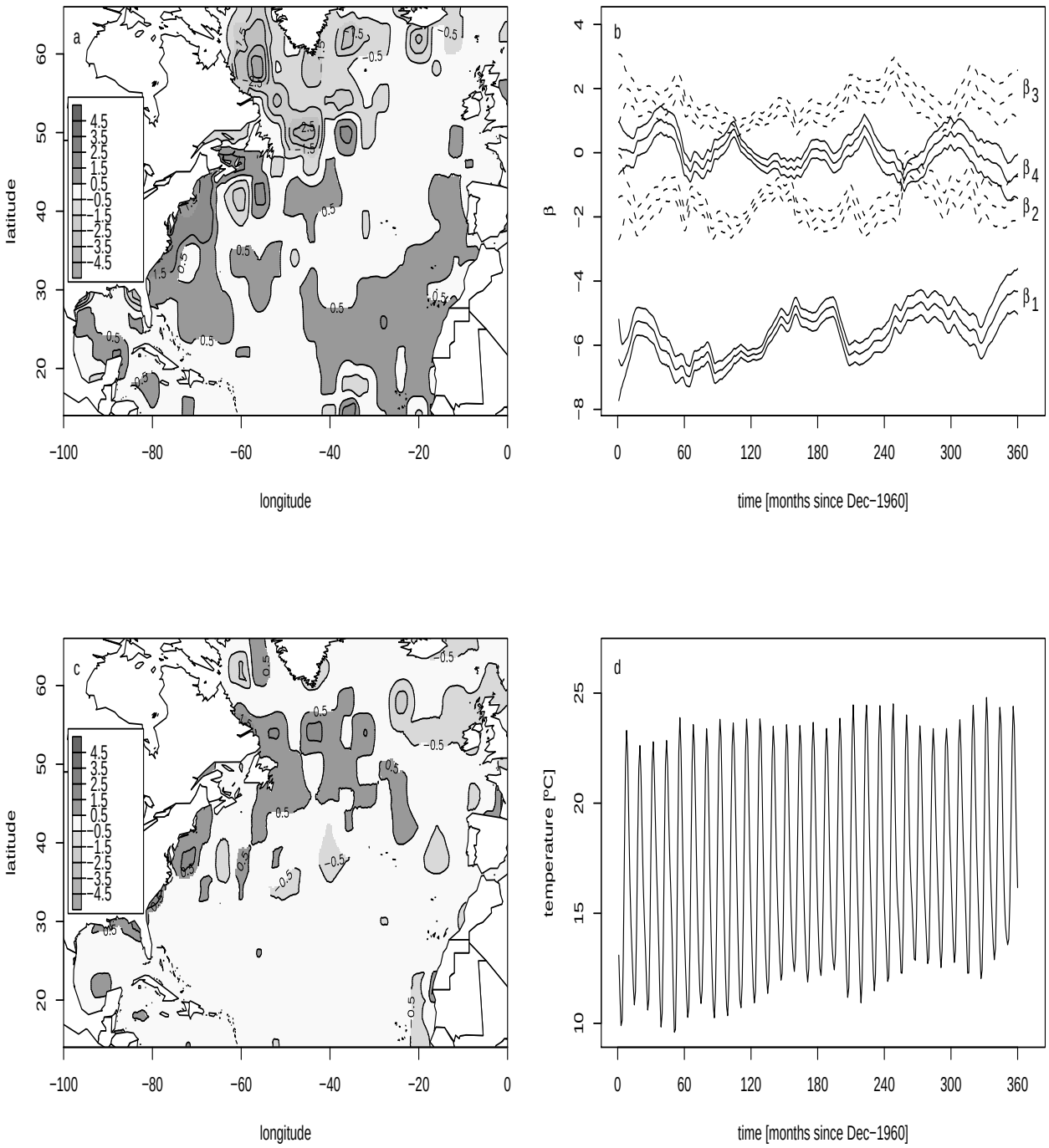

Figure 5: Linear and non-linear SST trends. a) Posterior mean long-term linear trend (°C/30yr). b) Posterior mean and 95% intervals of $\boldsymbol{\beta}_t(\boldsymbol{j})$, $\boldsymbol{j}$ = (72°W, 38°N). c) Posterior mean anomalies in July 1979 (°C). d) Posterior mean for monthly SST at $\boldsymbol{s}_2$ (°C).

up to 2.6°C off the US coast and cooled up to -5°C in the Labrador Sea. The difference in uncertainty, however, is large, mostly because of poorer sampling in the subpoles. Thus, the nuclei of cooling north and east of Newfoundland have $(-7.0, -3.3)$°C and $(-4.5, -3.4)$°C as 95% posterior intervals for the overall temperature change, respectively. The water mass off the mouth of Chesapeake Bay has $(2.4, 2.8)$°C as 95% posterior interval for the trend.

In $2° \times 2°$ squares centered at $\boldsymbol{s}_1$, $\boldsymbol{s}_2$ and $\boldsymbol{s}_3$, the proportion of the total data collected after 1975 (the midpoint of the time frame analysed) was respectively 26%, 34% and 43%. In light of this information and the trends presented in Figure 5a, we may attribute the results in Figure 4d to sampling error. In other words, we postulate that undersampling during the cold (warm) period in the vicinity of $\boldsymbol{s}_1$ ($\boldsymbol{s}_2$) may have introduced a warm (cold) bias on WOA. Greater equilibrium in the temporal distribution of data around $\boldsymbol{s}_3$ justifies the proximity between the two climatologies.

As stated in Section 4, $\boldsymbol{\beta}$ describes the annual cycle and non-linear trends. In the neighborhood of $\boldsymbol{s}_2$, seasonality is strong and β_1 dominates the signal (Figures 4f and 5b). Over time, all components experience wide fluctuations, owing to the low discount factor that affects the random walk. This feature is common to the whole basin (see below), and thus it is not surprising that transient anomalies display spatial patterns different from long-term trends. Figure 5c, for example, provides a snapshot of mean SST anomalies in July 1979. Overall, the North Atlantic is close to its climatological state, but the Gulf Stream and the North Atlantic Current carry water warmer than average, while their sides rest cooler than average. Returning to Figure 5b, the component β_1 stands out for presenting a trend, which dampens the amplitude of the 12 months harmonics, making winters warmer and summers cooler. When combined with the positive linear trend (Figure 5a), this intensifies winter warming and weakens summer SST change, as Figure 5d shows.

The posterior distributions of $\boldsymbol{\kappa}$ and $\boldsymbol{\rho}$ are alike, and thus Figure 6a shows information relative to the former, only. With 20 random draws of $\boldsymbol{\kappa}$ from the stationary MCMC, we construct 20 ellipses for each point $j \in J$, by equating $K^*[\boldsymbol{s} - \boldsymbol{j}, \boldsymbol{\kappa}(\boldsymbol{j})] = 0.5$. We use these ellipses to investigate the orientation and size of the convolution kernel for monthly SST mean, as well as the uncertainty related to the kernel's shape, which is revealed by the lack of superposition of the ellipses. Underneath the ellipses, we depict the posterior mean eccentricity of the kernels associated with $\boldsymbol{\Lambda}$, defined as $\epsilon = \sqrt{1 - \Lambda_2^2/\Lambda_3^2}$. As the plot reveals, large circular kernels are more common in the mid-Atlantic than elsewhere. Elongated kernels generally coincide with coastal zones, but also occur in the open ocean. Land areas hold the few small kernels that exist. As expected, uncertainty about the kernel's shape is more pronounced on land and along the margins of the study area.

In Figure 6b we draw annual mean sea surface currents data from Mariano Global Surface Velocity Analysis (MGSVA; Mariano and Brown, 1992; Mariano et al., 1995). To facilitate the identification of strong and weak currents, we depict the mean speed underneath the vectors. The resemblance between the shaded plot obtained and that of Figure 6a is noteworthy: eccentric kernels generally occur in regions with strong currents, *viz.* the North Equatorial Current (NEC), the Florida Current (FC), the Gulf Stream (GS), the North Atlantic Current (NAC), and the Canary Current (CC); in contrast, the center of the North Atlantic Gyre (NAG), where currents are weak, is dominated by large isotropic kernels. The North Atlantic Subpolar Gyre (NASG) and enveloping currents, *viz.* the Labrador Current (LC), the East

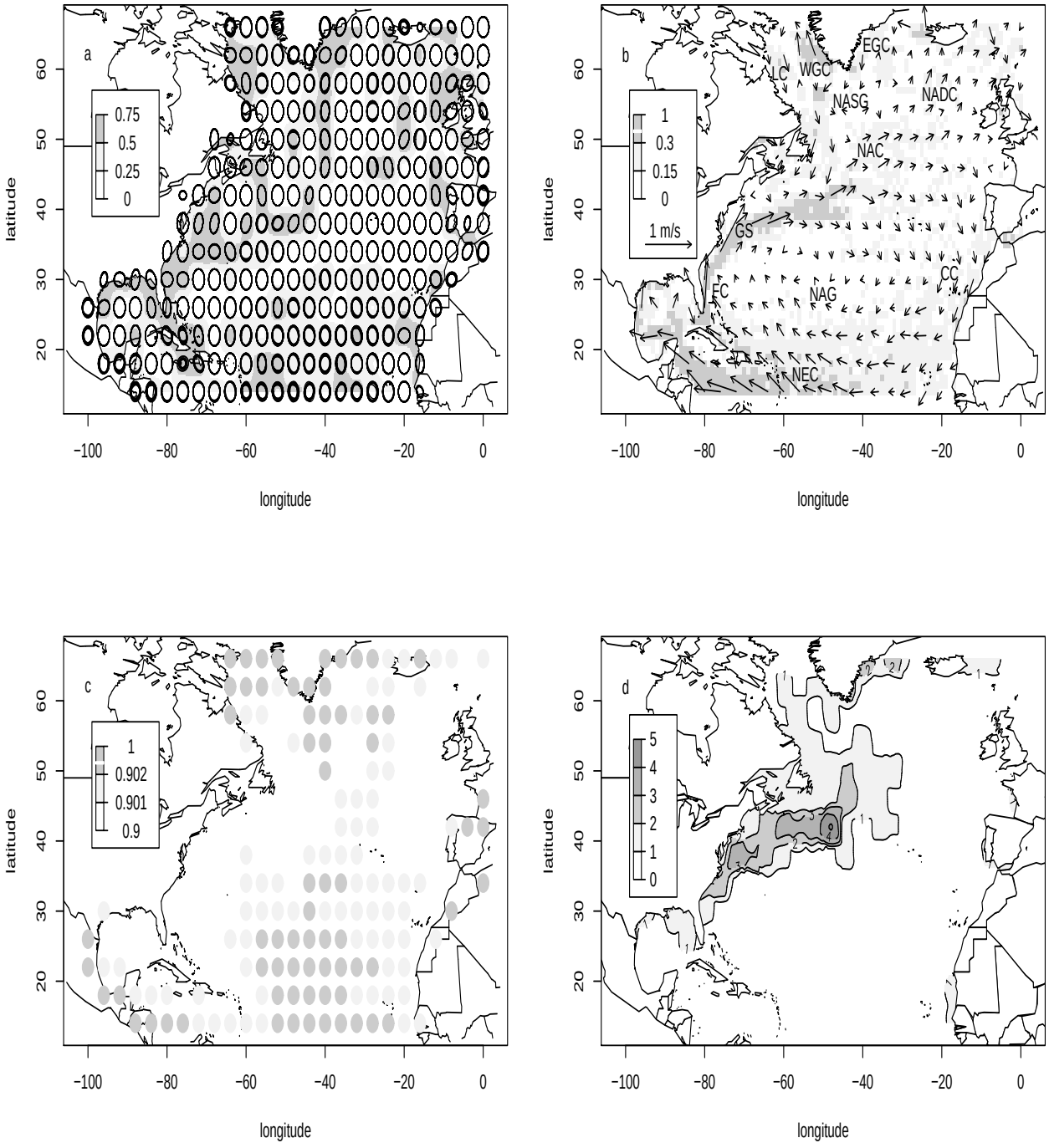

Figure 6: a) Mean eccentricity of the kernels associated with $\mathbf{\Lambda}(\boldsymbol{s})$ (shaded map), and 20 realizations of the kernel at every point of J. b) Direction (vectors drawn at points of J) and speed (vector length and shaded map) of annual mean surface currents, according to MGSVA (Mariano and Brown, 1992; Mariano et al., 1995). See text for acronyms. c) Posterior mean of $\nu(\boldsymbol{j})$. d) Posterior mean of $\Phi(\boldsymbol{s})$.

Parameter	Post. mean	95% posterior interval
τ^2_α	3.782	$(3.154, 4.501)$
$\tau^2_{\beta_{1,1}}$	1.070	$(0.804, 1.401)$
$\tau^2_{\beta_{1,2}}$	0.735	$(0.528, 1.000)$
$\tau^2_{\beta_{1,3}}$	0.347	$(0.236, 0.498)$
$\tau^2_{\beta_{1,4}}$	0.197	$(0.127, 0.289)$
τ^2_η	2.58×10^{-5}	$(2.15 \times 10^{-5}, 3.09 \times 10^{-5})$
τ^2_σ	0.529	$(0.440, 0.633)$

Table 5: Posterior means and 95% intervals for the variance parameters.

Greenland Current (EGC) and the West Greenland Current (WGC), are also discernible, and so is the North Atlantic Drift Current (NADC). Furthermore, the kernels' semimajor axes are often parallel to the direction of flow. This seems a sensible result, since points in the same streamline should have weaker SST differences than points in different streamlines. We find the same features in $\boldsymbol{\rho}$, which reinforces the link between SST variability and the dominating patterns of currents.

Recall that $\nu \approx 1$ indicates that $\boldsymbol{\beta}_t$ is close to constant in time, and $\nu \approx 0.9$ indicates large variability in $\boldsymbol{\beta}_t$ between consecutive times. Although we provided the same prior distribution to all $\nu(\boldsymbol{j})$, we expected coastal zones and regions with strong currents to present lower discount factors, due to their richer temporal dynamics. To some extent, the posterior distribution of ν substantiates our expectation (Figure 6c). With a larger data set, we believe the spatial heterogeneity would be more pronounced.

The posterior mean of $\Phi(\boldsymbol{s})$ clearly displays the mark of the Gulf Stream and, to a lesser extent, other strong currents (Figure 6d). Much of the variability in these regions is lost to error because, even with anisotropic kernels, the surfaces produced for $\theta_{m,y}(\boldsymbol{s})$ from parameters located on a 4° grid cannot resolve sharp meanders and eddies. Moreover, upwelling events and other transient features cannot be described adequately with a temporal resolution of one month. These losses of information are common, when constructing basin scale climatologies, because the heterogenous spatio-temporal distribution of data only allows finer resolutions up to a point, as our experiments with 3° and 2° grids proved.

Unlike the prior distributions of the MRF variance parameters, where most are identical, the posteriors display clear differences (Table 5), implying the model was able to learn from the data. The harmonics of the seasonal cycle present two noticeable features: the 12-months cycle is more spatially variable than the 6-months, and the sine component is more spatially variable than its cosine counterpart.

Figure 7 displays the spatial distribution of the 704 posterior p-values that fall outside the range $(0.005, 0.995)$. Clusters occur mainly in regions with strong currents (cf. Figure 6b) and moderate sampling, which we do not find surprising given the coarseness of the model, as discussed above. One way of coping with this shortcoming would be to provide different spatial and temporal resolutions to the model, according to location. On the other hand, this low proportion of extreme p-values (0.06%) does not indicate severe lack of fit, and the fact that several of them belong to the same cruise suggests that errors may have occurred in

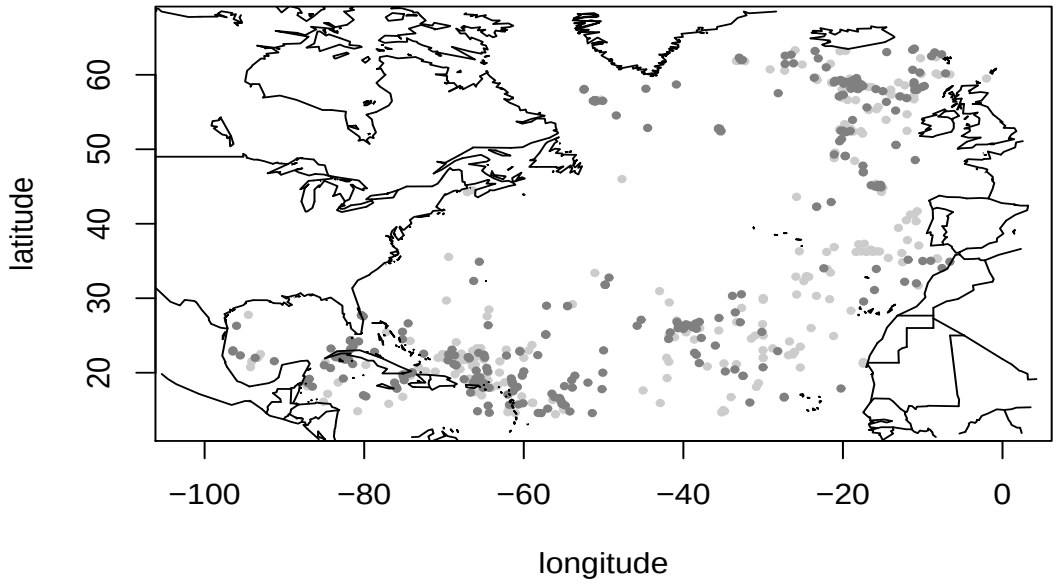

Figure 7: Spatial distribution of posterior predictive p-values smaller (greater) than 0.005 (0.995), in light (dark) gray.

data collection. We conclude that increasing the sampling effort around the North Atlantic Gyre would assist in determining the quality of fit and allowing a multiresolution model. Plots of extreme p-values for different seasons, decades and instruments do not indicate any other localized lack of fit.

Owing to their massive size, samples of $\theta(\boldsymbol{s})$ and $\boldsymbol{\beta}_t(\boldsymbol{j})$ cannot be stored for convergence analysis. For the remaining parameters, our choices of burn-in, length of the chain and thinning seem acceptable. More than 95% of the parameters pass at least three out of four `BOA` tests. Detailed tables and trace plots are provided in `http://www.ams.ucsc.edu/~bruno/climatology/`.

7 Final remarks

Climatologies are valuable data-based products for the ocean scientific community, since they summarize what is known from the observational record. In order to produce smooth gridded maps, and because some regions and periods have been poorly sampled, an underlying model is required to distribute information across space and time in a simple, seamless way. In this paper we introduce such a model for North Atlantic SST. Its construction allows a straightforward application to the World Ocean, other layers and properties, in the same manner as WOA, *i.e.*, with each layer and property being analysed separately. Apart from the climatology, the model also produces fields of transient anomalies and long-term trends, which are of interest for climate variability research and can be compared to other methods

(e.g., Levitus et al., 1994; Grey et al., 2000; Casey and Cornillon, 2001; Polyakov et al., 2005).

Several aspects differ between the objective analysis of WOA and our method. In contrast with WOA, which aggregates data without considering the year of observation, our approach constructs "instantaneous" fields prior to averaging them into monthly climatologies, similarly to Higdon (1998). In this manner, it incorporates our knowledge that temperature fluctuates over time, and mitigates biasing towards years with more data or distorting seasonal harmonics. *Post hoc* corrections and smoothing of the annual cycle, as performed in WOA, are not needed. Due to land barriers and the flow of water masses with different properties, non-stationarity and location dependent anisotropy are other features that should be accounted for when constructing a SST climatology (Ridgway et al., 2002). Our use of Gaussian processes located on a grid, convolved by a kernel whose shape evolves across space, addresses this issue in line with Higdon (1998) and Higdon et al. (1999), while WOA employs one set of isotropic kernels worldwide. Despite the increased computational burden associated with the estimation of these parameters, the close link found between the kernels' shape and climatological sea surface currents is a captivating result. The climatological gradients presented along a transect indicate a better distinction of water masses, both adjoining and separated by land. Under a MCMC setting for posterior sampling, the number of iterations in the procedure becomes subject to convergence analyses. On the other hand, the initial conditions become of no importance, as our experiment with two cold and warm initial fields demonstrated. These features differ from WOA, where the number of iterations (3) is decided beforehand and the initial field appears relevant to the final product.

Apart from including the parameters that control the kernels' shape in the model ($\boldsymbol{\kappa}$ and $\boldsymbol{\rho}$), we follow other suggestions made by Higdon (1998) to improve spatiotemporal models for ocean temperature, *viz.*: specifying some point processes as MRFs, with simple neighbor dependencies on the grid (α, β, η and σ); using a kernel with compact support; and carefully designing the MCMC algorithm, to reduce computation time. This latter feature proves invaluable for coping with such a large data set: with each parallel process working on a 8° longitudinal strip of the North Atlantic, the MCMC algorithm was able to run in just 3 days (22,200 iterations). So long as each process does not work with a wider strip, increasing the domain should have no noticeable impact on the duration of the run. To ensure the kernels vary smoothly over space, we also allow the parameters to evolve with location, according to isotropic DPCs. Unlike Higdon et al. (1999), however, we do not use any hyperparameter to model these kernels, to accelerate convergence of the others in the MCMC. The Bayesian framework in which this model is cast allows us to assign estimates of uncertainty to all quantities. Owing to this feature, comparing the model's results to other climatologies such as WOA, the output of numerical models, or fields derived from remote sensing, becomes straightforward.

The North Atlantic sea surface presents strong dynamics on various temporal scales. Hence, our model benefits from the inclusion of sinusoidal components, linear trends and discount factors as parameters, which are novelties relative to previous approaches, focused on the deep ocean (Higdon, 1998) or smaller regions (Lemos and Sansó, 2006; Sahu and Challenor, 2007). The diversity of results in this basin demonstrate the potential for application of the model to wider domains, several depths and more comprehensive data sets,

including other ocean properties. The joint modeling of temperature and salinity, however, requires a careful consideration of the variation of density with depth, to avoid producing unrealistic, unstable water columns (Boyer et al., 2005). Salinity and nutrient climatologies are also impaired by the serious shortage of data, and a more elaborate model, including for example the equations of flow, may be required to fill in gaps. Such approach would part it from traditional climatological analyses, where fluid dynamics are not used, and make it more similar to ocean data assimilation models.

8 Acknowledgements

This work was performed under the scope of project Portcoast (POCTI /CLI /58348 /2004). R. Lemos acknowledges grant SFRH /BD /17929 /2004 from Fundação para a Ciência e a Tecnologia. B. Sansó was partially supported by the National Science Foundation grant NSF-Geomath 0417753. We thank Scott Brandt, Luis Fernandes, Paulo Leitão, Tiago Pereira, Guillaume Rifflet and Sage Weil for their help in implementing the software on computer clusters, and Sun Microsystems, Inc. for a grant of 10,000 compute hours on the Sun Grid Compute Utility.

9 Appendix - Full conditional distributions

- $(\theta_{m,y}(\boldsymbol{s})|\ldots) \sim N\left(D_\theta d_\theta, D_\theta\right)$
  $D_\theta^{-1} = \frac{1}{\sum_j K[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Omega}(\boldsymbol{s})]\exp(\sigma(\boldsymbol{j}))} + \frac{1}{\tau_i^2}$
  $d_\theta = \frac{\sum_j K[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Lambda}(\boldsymbol{s})]\left(\alpha(\boldsymbol{j})+\boldsymbol{\beta}_t(\boldsymbol{j})\boldsymbol{w}_t^T+\eta(\boldsymbol{j})(t-180)\right)}{\sum_j K[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Omega}(\boldsymbol{s})]\exp(\sigma(\boldsymbol{j}))} + \frac{x_{i,m,y}(\boldsymbol{s})}{\tau_i^2}$,

- $(\alpha(\boldsymbol{j})|\ldots) \sim N\left(D_\alpha d_\alpha, D_\alpha\right)$
  $D_\alpha^{-1} = \frac{1}{\tau_\alpha^2} + \sum_{\boldsymbol{s}} \frac{K^2[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Lambda}(\boldsymbol{s})]}{\sum_{\boldsymbol{j}'\in J} K[\boldsymbol{s}-\boldsymbol{j}',\boldsymbol{\Omega}(\boldsymbol{s})]\exp(\sigma(\boldsymbol{j}'))}$
  $d_\alpha = \frac{\alpha(\mathrm{N}(\boldsymbol{j}))+\alpha(\mathrm{S}(\boldsymbol{j}))+\alpha(\mathrm{E}(\boldsymbol{j}))+\alpha(\mathrm{W}(\boldsymbol{j}))}{4\tau_\alpha^2} + \sum_{\boldsymbol{s}} \frac{K[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Lambda}(\boldsymbol{s})](\theta_{m,y}(\boldsymbol{s})-C_\alpha)}{\sum_{\boldsymbol{j}'\in J} K[\boldsymbol{s}-\boldsymbol{j}',\boldsymbol{\Omega}(\boldsymbol{s})]\exp(\sigma(\boldsymbol{j}'))}$,
  $C_\alpha = K[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Lambda}(\boldsymbol{s})]\left(\boldsymbol{\beta}_t(\boldsymbol{j})\boldsymbol{w}_t^T + \eta(\boldsymbol{j})(t-180)\right) +$
  $+ \sum_{\boldsymbol{j}'\in J\setminus\{\boldsymbol{j}\}} K[\boldsymbol{s}-\boldsymbol{j}',\boldsymbol{\Lambda}(\boldsymbol{s})]\left(\alpha(\boldsymbol{j}') + \boldsymbol{\beta}_t(\boldsymbol{j}')\boldsymbol{w}_t^T + \eta(\boldsymbol{j}')(t-180)\right)$

- $(\eta(\boldsymbol{j})|\ldots) \sim N\left(D_\eta d_\eta, D_\eta\right)$
  $D_\eta^{-1} = \frac{1}{\tau_\eta^2} + \sum_{\boldsymbol{s}} \frac{K^2[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Lambda}(\boldsymbol{s})]}{\sum_{\boldsymbol{j}'\in J} K[\boldsymbol{s}-\boldsymbol{j}',\boldsymbol{\Omega}(\boldsymbol{s})]\exp(\sigma(\boldsymbol{j}'))}$
  $d_\eta = \frac{\eta(\mathrm{N}(\boldsymbol{j}))+\eta(\mathrm{S}(\boldsymbol{j}))+\eta(\mathrm{E}(\boldsymbol{j}))+\eta(\mathrm{W}(\boldsymbol{j}))}{4\tau_\eta^2} + \sum_{\boldsymbol{s}} \frac{K[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Lambda}(\boldsymbol{s})](\theta_{m,y}(\boldsymbol{s})-C_\eta)}{\sum_{\boldsymbol{j}'\in J} K[\boldsymbol{s}-\boldsymbol{j}',\boldsymbol{\Omega}(\boldsymbol{s})]\exp(\sigma(\boldsymbol{j}'))}$,
  $C_\eta = K[\boldsymbol{s}-\boldsymbol{j},\boldsymbol{\Lambda}(\boldsymbol{s})]\left(\alpha(\boldsymbol{j}) + \boldsymbol{\beta}_t(\boldsymbol{j})\boldsymbol{w}_t^T\right) +$
  $+ \sum_{\boldsymbol{j}'\in J\setminus\{\boldsymbol{j}\}} K[\boldsymbol{s}-\boldsymbol{j}',\boldsymbol{\Lambda}(\boldsymbol{s})]\left(\alpha(\boldsymbol{j}') + \boldsymbol{\beta}_t(\boldsymbol{j}')\boldsymbol{w}_t^T + \eta(\boldsymbol{j}')(t-180)\right)$

- $\left(\tau_\alpha^2|\ldots\right) \sim IG\left(\frac{1}{2}\left(n_0+n_J\right), \frac{1}{2}\left(n_0S_0 + \sum_{\boldsymbol{j}}\left(\alpha(\boldsymbol{j}) - \frac{\alpha(\mathrm{N}(\boldsymbol{j}))+\alpha(\mathrm{S}(\boldsymbol{j}))+\alpha(\mathrm{E}(\boldsymbol{j}))+\alpha(\mathrm{W}(\boldsymbol{j}))}{4}\right)^2\right)\right)$

- Let τ_k^2 and β_k designate the k-th component of the vectors $\boldsymbol{\tau}_{\beta_1}^2$ and $\boldsymbol{\beta}_1$, $k = 1,\ldots,4$. Then
  $\left(\tau_k^2|\ldots\right) \sim IG\left(\frac{1}{2}\left(n_0+n_J\right), \frac{1}{2}\left(n_0S_0 + \sum_{\boldsymbol{j}}\left(\beta_k(\boldsymbol{j}) - \frac{\beta_k(\mathrm{N}(\boldsymbol{j}))+\beta_k(\mathrm{S}(\boldsymbol{j}))+\beta_k(\mathrm{E}(\boldsymbol{j}))+\beta_k(\mathrm{W}(\boldsymbol{j}))}{4}\right)^2\right)\right)$

- $(\tau_\eta^2|\ldots) \sim IG\left(\frac{1}{2}(n_0+n_J), \frac{1}{2}\left(n_0 S_0 + \sum_{\boldsymbol{j}}\left(\eta(\boldsymbol{j}) - \frac{\eta(\mathrm{N}(\boldsymbol{j}))+\eta(\mathrm{S}(\boldsymbol{j}))+\eta(\mathrm{E}(\boldsymbol{j}))+\eta(\mathrm{W}(\boldsymbol{j}))}{4}\right)^2\right)\right)$

References

Boyer, T., Levitus, S., Garcia, H., Locarnini, R. A., Stephens, C., and Antonov, J. (2005). Objective analyses of annual, seasonal, and monthly temperature and salinity for the World Ocean on a 0.25 grid. *Int. J. Climatol.*, 25:931–945.

Boyer, T. P., Antonov, J. I., Garcia, H. E., Johnson, D. R., Locarnini, R. A., Mishonov, A. V., Pitcher, M. T., Baranova, O. K., and Smolyar, I. V. (2006). World Ocean Database 2005. Technical report, NOAA Atlas NESDIS 60, U.S. Government Printing Office, Washington D.C., 190 pp., DVDs.

Brasseur, P., Beckers, J. M., Brankart, J. M., and Schoenauen, R. (1996). Seasonal temperature and salinity fields in the Mediterranean Sea: climatological analyses of a historical data set. *Deep-Sea Res. Pt. I*, 43(2):159–192.

Brenning, A. (2001). Geostatistics without stationarity assumptions within geographical information systems. *Freiberg Online Geoscience*, 6:1–108.

Brooks, S. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.*, 7:434–55.

Brown, P. J., Le, N. D., and Zidek, J. V. (1995). Multivariate Spatial Interpolation and Exposure to Air Pollutants. *Can. J. Stat.*, 22:489–509.

Casey, K. S. and Cornillon, P. (2001). Global and regional sea surface temperature trends. *J. Clim.*, 14:3801–3818.

Damian, D., Sampson, P. D., and Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structure. *Environmetrics*, 12:161–176.

Fuentes, M. (2002). Periodogram and other spectral methods for nonstationary spatial processes. *Biometrika*, 89:197–210.

Fuentes, M. (2007). Approximate Likelihood for Large Irregularly Spaced Spatial Data. *J. Am. Stat. Assoc.*, 102:321–331.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. Chapman and Hall, London, UK, second edition.

Gaspari, G. and Cohn, S. E. (1999). Construction of Correlation Functions in Two and Three Dimensions. *Q. J. Roy. Meteor. Soc.*, 125:723–757.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis, Second Edition*. Chapman & Hall/CRC.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stati. Sci.*, 7:457–72.

Geweke, J. (1992). Evaluating the accuracy of sampling-based i approaches to calculating posterior moments. In Bernardo, J. M., Berger, J. O., Dawid, P., Smith, A. F. M., and West, M., editors, *Bayesian Statistics 7*. Clarendon Press, Oxford, UK.

Gouretski, V. V. and Janke, K. (1999). A description and quality assessment of the historical hydrographic data for the South Pacific Ocean. *J. Atmos. Ocean. Tech.*, 16(11):1791–1815.

Gouretski, V. V. and Koltermann, K. P. (2007). How much is the ocean really warming? *Geophys. Res. Lett.*, 34, L01610, doi:10.1029/2006GL027834.

Grey, S. M., Haines, K., and Troccoli, A. (2000). A Study of Temperature Changes in the Upper North Atlantic: 1950-94. *J. Climate*, 13(15):2697–2711.

Heidelberger, P. and Welch, P. (1983). Simulation run length control in the presence of an initial transient. *Oper. Res.*, 31:1109–44.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the North Atlantic Ocean. *Environmental and Ecological Statistics*, 5(2):173–190.

Higdon, D. M. (2002). Space and space-time modeling using process convolutions. In Anderson, C., Barnett, V., Chatwin, P. C., and El-Shaarawi, A. H., editors, *Quantitative Methods for Current Environmental Issues*, pages 37–56, London. Springer Verlag.

Higdon, D. M., Swall, J., and Kern, J. (1999). Non-Stationary Spatial Modeling. In *Proceedings of the Sixth Valencia International Meeting*, pages 761–768.

Holbrook, N. J. and Bindoff, N. L. (2000). A statistically efficient mapping technique for four-dimensional ocean temperature data. *J. Atmos. Ocean. Tech.*, 17:831–846.

Johannesson, G., Cressie, N., and Huang, H.-C. (2007). Dynamic multi-resolution spatial models. *Environ. Ecol. Stat.*, to appear.

Kushnir, Y. (1994). Interdecadal variations in North Atlantic sea surface temperature and associated atmospheric conditions. *J. Clim.*, 7:141–157.

Lemos, R. T. and Sansó, B. (2006). Spatio-temporal variability of ocean temperature in the Portugal Current System. *J. Geophys. Res. Oceans*, 111 (C4)(C04010).

Levitus, S., Antonov, J. I., and Boyer, T. P. (1994). Interannual Variability of Temperature at a Depth of 125 Meters in the North Atlantic Ocean. *Science*, 266.(5182):96–99.

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

Mariano, A. J. and Brown, O. B. (1992). Efficient objective analysis of dynamically heterogeneous and nonstationary fields via the parameter matrix. *Deep-Sea Res.*, 39(7/8):1255–1271.

Mariano, A. J., Ryan, E. H., Smithers, S., and Perkins, B. (1995). The Mariano Global Surface Velocity Analysis. Technical report, U.S. Coast Guard Technical Report CG-D-34-95.

Nychka, D., Wikle, C., and Royle, J. A. (2002). Multiresolution models for nonstationary spatial covariance functions. *Stat. Model.*, 2:315–332.

Paciorek, C. J. (2007). Bayesian smoothing for irregularly-spaced data using Fourier basis functions. *J. Stat. Soft.*, submitted.

Paciorek, C. J. and Schervish, M. J. (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17:483–506.

Polyakov, I. V., Bhatt, U. S., Simmons, H. L., Walsh, D., Walsh, J. E., and Zhang, X. (2005). Multidecadal Variability of North Atlantic Temperature and Salinity during the Twentieth Century. *J. Climate*, 18:4562–4581.

R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Raftery, A. E. and Lewis, S. M. (1992a). Comment: One long run with diagnostics: Implementation strategies for markov chain monte carlo. *Stat. Sci.*, 7:493–7.

Raftery, A. E. and Lewis, S. M. (1992b). How many iterations in the Gibbs sampler? In Bernardo, J. M., Berger, J. O., Dawid, P., and Smith, A. F. M., editors, *Bayesian Statistics 4*, pages 765–776. Oxford University Press.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *J. Climate*, 15:1609–1625.

Ridgway, K. R., Dunn, J. R., and Wilkin, J. L. (2002). Ocean interpolation by four-dimensional weigthed least squares - application to the waters around Australasia. *J. Ocean. Atmos. Tech.*, 19:1357–1375.

Sahu, S. K. and Challenor, P. (2007). A space-time model for joint modeling of ocean temperature and salinity levels as measured by Argo floats. *Environmetrics*, 10.1002/env.895.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *J. Am. Stat. Assoc.*, 87:108–119.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structures via spatial deformations. *J. Roy. Stat. Soc., Ser. B*, 65:743–758.

Smith, B. J. (2005). *Bayesian Output Analysis program (BOA) for MCMC*. R package version 1.1.5-2.

Tzeng, S., Huang, H.-C., and Cressie, N. (2005). A fast, optimal spatial-prediction method for massive datasets. *J. Am. Stat. Assoc.*, 100:1343–1357.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models.* Springer Verlag, New York, second edition.

Xun, Q. (2005). Representations of Inverse Covariances by Differential Operators. *Adv. Atmos. Sci.*, 22:181–198.