



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

Ricardo T. Lemos & Bruno Sansó

To cite this article: Ricardo T. Lemos & Bruno Sansó (2009) A Spatio-Temporal Model for Mean, Anomaly, and Trend Fields of North Atlantic Sea Surface Temperature, Journal of the American Statistical Association, 104:485, 5-18, DOI: [10.1198/jasa.2009.0018](https://doi.org/10.1198/jasa.2009.0018)

To link to this article: <https://doi.org/10.1198/jasa.2009.0018>



Published online: 01 Jan 2012.



Submit your article to this journal [↗](#)



Article views: 312



View related articles [↗](#)



Citing articles: 28 View citing articles [↗](#)

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

Ricardo T. LEMOS and Bruno SANSÓ

We consider the problem of fitting a statistical model to 30 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 analyses. Our model improves upon the current tools used by oceanographers in that it constructs instantaneous temperature fields before 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 nonisotropic and nonstationary space and time dependencies, owing to its use of discrete process convolutions. Particular attention is given to the handling of massive datasets 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; Markov chain Monte Carlo method; Parallel processing.

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 or *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; for example, in numerical ocean modeling, climatologies are used to provide the initial stages of simulations, to define boundary and forcing functions, to keep estimates of properties within realistic bounds, and to evaluate results. As climate change and climate variability research developed, another focal point became the deviations from the mean, dubbed *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, Rayner, Smith, Stokes, and Wang 2002; Boyer et al. 2005) or constant density (Lozier, Owens, and Curry 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. Furthermore, isotropy and stationarity assumptions generally do not hold for ocean properties. Extensions and alternatives to OA include Empirical Orthogonal Functions analyses (Holbrook and Bindoff 2000), variational analyses (Brasseur, Beckers, Brankart, and Schoenauen 1996), and anisotropic loess smoothers (Ridgway, Dunn, and Wilkin 2002), which mitigate some but not all of these issues.

In the last 15 years, several methods were developed to estimate parameters in nonstationary anisotropic models, most of which consider Gaussian processes. The idea of producing a nonlinear transformation of space to achieve isotropy was popularized in the early 1990s, after the work of Sampson and Guttorp (1992). This approach was originally developed using cross-validation and then extended to include likelihood-based methods (Damian, Sampson, and Guttorp 2001; Schmidt and O'Hagan 2003). Another approach used multivariate normal models to estimate general covariance matrices (Brown, Le, and Zidek 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, Swall, and Kern 1999). A related

Ricardo Lemos is a graduate student at Instituto de Oceanografia, Universidade de Lisboa, and Maretec—Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa (E-mail: rtl@net.sapo.pt). Bruno Sansó is Professor, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064 (E-mail: bruno@ams.ucsc.edu). 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. The authors 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.

technique created large classes of nonstationary covariance functions using convolutions (Paciorek and Schervish 2006).

Sophisticated statistical methods are increasingly applied to massive datasets collected *in situ* or from remote sensing (e.g., Wikle, Millif, Nychka, and Berliner 2001). Explicit computations of covariance matrices, which correspond to Gaussian processes observed at millions of locations, are usually avoided, because 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, Wikle, and Royle (2002), Tzeng, Huang, and Cressie (2005), Fuentes (2007), Johannesson, Cressie, and Huang (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 article 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 display smooth contours, comparable 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 come together with measures of uncertainty; e) the method must be easily extended to a larger dataset or geographical domain.

For our approach, we use data from the NODC World Ocean Database 2005 (Boyer et al. 2006), collected with four types of instruments between 1961 and 1990. Screening was performed using quality control criteria similar to those in the NODC web page, 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 s in space S , say $\theta(s)$, as

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

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

This article is organized as follows. In Section 2, we provide the most important aspects of our notation. In Section 3.1, we describe DPCs into greater depth, focusing especially on the anisotropic kernel. In Section 3.2, we present our hierarchical Bayesian model, and show how it is used to produce climatological, anomaly, and trend fields of North Atlantic SST, for the period 1961–1990. In Section 3.3 we specify how the model is fitted, and in Section 4 we present and discuss the most important results. Finally, in Section 5, 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 and matrices 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 $s \equiv (x_s, y_s)$ and $j \equiv (x_j, y_j)$. Variance parameters are designated with τ^2 . All other parameters have time as subscript and spatial location in parentheses e.g., $\beta_t(j)$. 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. MODEL

3.1 Discrete Process Convolutions

To model processes that vary smoothly over S , we define a grid J and assign a latent process to each point $j \in J$. As stated in Equation (1), DPC modeling consists of equating the process of interest, at location s , to a normalized weighted sum of the latent processes. The weight assigned to each latent process depends on s , and its unnormalized value, which we denote with an asterisk (K^*), is in our approach provided by the kernel

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

Here, $\omega \in \mathbb{R}^4$ (note that, to facilitate reading, we omit the spatial dependence of ω in this section), and

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

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

$$\Sigma^{-1} \equiv \begin{pmatrix} \Psi_1 + \Psi_2 \cos 2\omega_4 & \Psi_2 \sin 2\omega_4 \\ \Psi_2 \sin 2\omega_4 & \Psi_1 - \Psi_2 \cos 2\omega_4 \end{pmatrix},$$

$$\Psi = \frac{1}{2} \left(\frac{1}{\omega_2^2} + \frac{1}{\omega_3^2}, \frac{1}{\omega_2^2} - \frac{1}{\omega_3^2} \right).$$

Conditions for the positive definiteness of Σ will be given later. In the next paragraphs, we examine the kernel's flexibility for DPC modeling, and provide Figure 1 as example.

If we let $\omega_1 \in (1.5, 5)$, 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 $\Psi_2 = 0$; Σ becomes space-invariant and proportional to the 2×2 identity matrix, and the distance between s and 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

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

which defines an ellipse centered at s . Thus, if $0 < \omega_2 \leq \omega_3$, ω_2 and ω_3 define the sizes of the ellipse's semiminor and semi-major 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 s may have substantially different contributions to the DPC. For example, in Figure 1a, the DPC for a process located at s would result from averaging the latent processes at j and 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 s , at least its nearest

neighbor in J falls within the support. Because 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 process of interest, for any location in S , can be reasonably described by (at most) the four nearest neighboring latent processes located in J . Because the minimum Euclidean distance between 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 $r_J/\sqrt{2} < \omega_2 \leq \omega_3 < r_J$, which guarantees the positive definiteness of Σ .

The preceding choice for the maximum size of the kernel's compact support is an aggressive one, but it comes with two important features: a) in Equation (1), the summation has at most four nonzero terms, easy to map, for all s ; b) only the random processes $\theta(s)$ falling in the circle centered at j with radius r_J are related with $\psi(j)$ (Figure 1a). Thus, once we define a suitable hierarchical model and begin exploring the posterior distribution of its processes (say, θ and ψ) with Markov chain Monte Carlo methods (MCMC; see, e.g., Gamerman and Lopes 2006), summations are rapidly computed and we can make efficient use of a parallel computing architecture. To this end, we divide S into a number of overlapping regions, partition J accordingly, and let a different computing processor sample only the model processes within its region of S and J . Because regions are not independent, each processor must have updated information about model processes in adjacent regions. A scheme of the procedure for the simplest network possible (two processors) is presented in Figure 2 and Table 1.

On the other hand, we could consider the kernel's maximum support, as well as the DPC grid resolution, r_J , as two random quantities, and embed them in the estimation procedure. This would be more consistent, from a purely methodological viewpoint. Alternatively, we could use model selection criteria. Unfortunately, none of these options turn out to be computationally feasible when fitting a hierarchical model to a massive

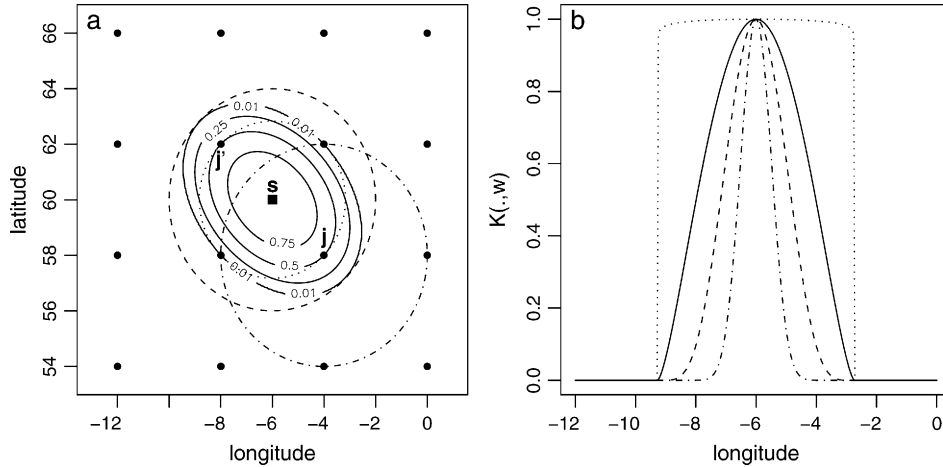


Figure 1. The convolution kernel for $s = (-6, 60)$ and $\omega = (1.5, 2\sqrt{2}, 4, \pi/4)$. (a) Top view. The kernel's contours are depicted by the solid lines. The dashed (dotted) circle delimits the largest (smallest) possible support for any kernel centered at s . The bullets are points of J , and the dashed-dotted circle delimits the area where the DPC weights involving j may be nonzero. (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.

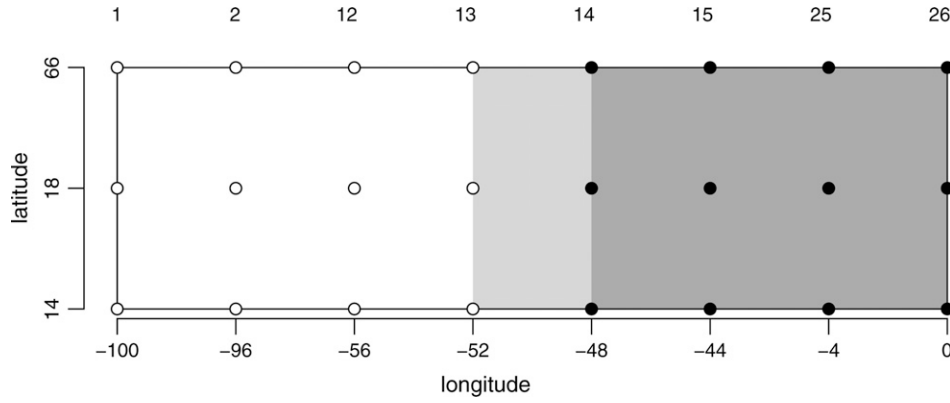


Figure 2. In this example, S is divided into two regions that overlap between 52°W and 48°W . Latent processes in empty (filled) points of J are sampled by processors 1 (2). The columns of J are numbered over S to describe the MCMC algorithm (Table 1).

dataset. As a consequence, for the problem at hand, we equate both the resolution of J and the upper limit for the kernel's semimajor axis to 4° , and we explore other possibilities with smaller domains and test cases. We evaluate our modeling choices based on computation time, convergence, goodness-of-fit assessment, and a qualitative evaluation of the resulting climatological, anomaly, and trend fields. The most important points to consider are: a) the kernel's semiminor axis has $r_J\sqrt{2}/2$ as lower limit; b) the upper limit of the support, relative to the resolution, conditions the size of the DPC summations, which in turn affects the computation time; c) certain configurations of resolution and support may lead to convergence problems, overfitting the data or producing large ripples in the fields; d) decreasing the grid spacing requires more processing power and storage capacity, to the point that computations may be unfeasible; and e) increasing the maximum kernel support renders the parallelization less efficient, because a smaller number of processors can be used and more information must be shared among them.

The application of the full hierarchical model with different configurations (grid resolution or kernel shape) is described in 3.3. Here, we draw insight from the simple DPC model

$$x(s) \sim N\left(\sum_j K[s-j, \omega] \psi(j), \tau^2\right).$$

We cast this model in a univariate domain, and try different choices of grid resolution and kernel support. The smoothness parameter ω_1 is set to two, and flat priors are provided to ψ and τ^2 . Figure 3 presents four configurations for the same simulated dataset. We observe that increasing the grid resolution pro-

duces more rugged fits (upper panels), and that coarse grid models (lower panels) are more resilient to producing ripples in regions with no data (gray area). Fits are smoother for kernels with large support (right panels), at the cost that the estimated latent processes are very far from observations. In the real application to which we return later, large spatio-temporal windows with few data are frequent. Thus, an efficient transfer of information between neighboring latent processes is necessary so as to produce realistic convolution surfaces. In models with kernels that have large supports, this transfer is less efficient, because latent processes are highly variable across space even in data rich regions. This, in addition to the computational advantages, prompts us to set our initial preference to the coarse grid model with small support (lower left panel), even though it misses some of the variability present in the observations.

3.2 Hierarchical Model

The model presented in this article has a temporal resolution of 1 month, from January 1961 to December 1990, and J consists of a 4° grid that covers the North Atlantic (Figure 4a). We describe an observation of SST collected with instrument $i = 1, \dots, 4$ (see Table 2), in month m , year y , and location s , say $x_{i,m,y}(s)$, as a sum of true SST, $\theta_{m,y}(s)$, and measurement error with variance τ_i^2 ; thus

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

Measurements are irregularly distributed in space and time, with coastal regions and the mid 1960s being more heavily sampled (Figure 4b and c). Given that SST is never sampled twice in any location, we use the accuracy reports of Boyer

Table 1. MCMC cyclical procedure for the setting displayed in Figure 2.

Step	Processor 1	Processor 2
1	Sample ψ in columns 1–12 of J	Sample ψ in columns 14–25 of J
2	Receive samples of ψ from column 14 of J	Send samples of ψ from column 14 of J
3	Sample ψ in column 13 of J	Sample ψ in column 26 of J
4	Send samples of ψ from column 13 of J	Receive samples of ψ from column 13 of J
5	Sample θ in the area 100°W – 48°W of S	Sample θ in the area 48°W – 0°W of S
6	Send samples of θ in the area 52° – 48°W of S	Receive samples of θ in 52° – 48°W of S

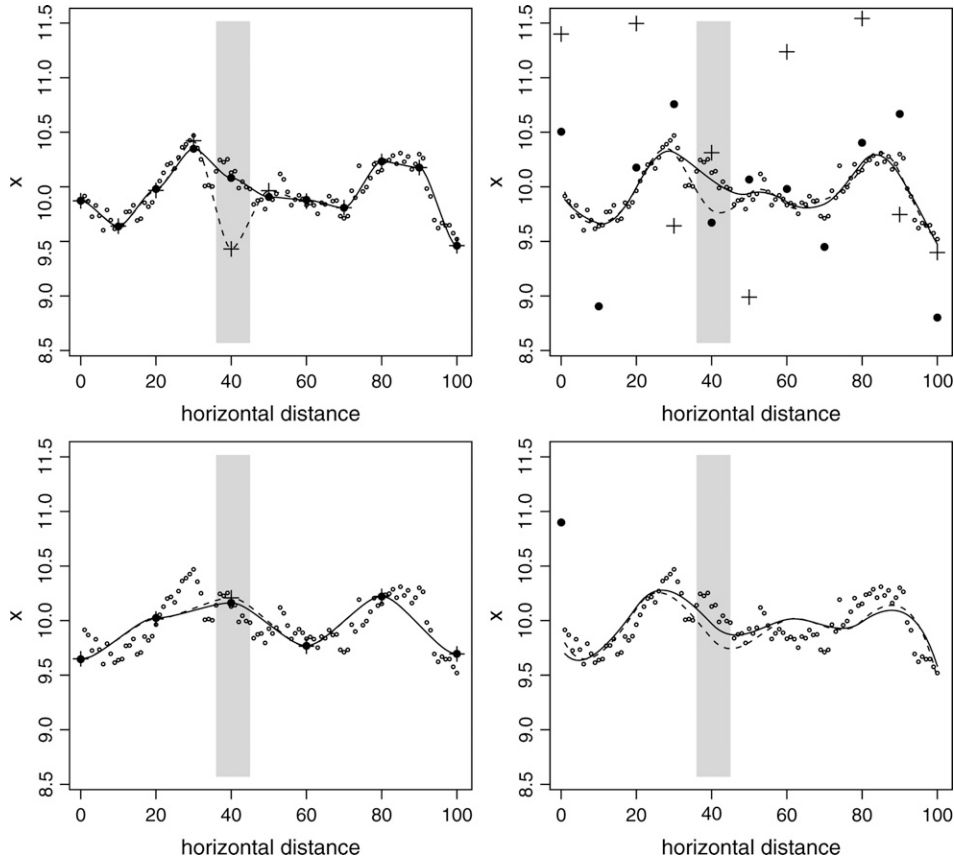


Figure 3. DPC fit of 100 simulated data. Left (right) panels display kernels with support equal to $1 \times (2 \times)$ the grid resolution. In the first (second) row the resolution is 2 (1) grid points per 20 horizontal units. Continuous lines correspond to DPC fits for all data, dotted lines correspond to DPC fits after eliminating data in the gray region. Bullets (crosses) correspond to the posterior means of the latent processes obtained by including (excluding) the data in the gray region; some exceed plot bounds.

et al. (2006) to fix the error variances of measurements made with high resolution conductivity, temperature, and depth instruments, expendable bathythermographs, and mechanical bathythermographs. Assigning an error variance to ocean station data (OSD) is more difficult, because this dataset comprises measurements made 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 preceding, the climatological SST for month m and location s , denoted $\Xi_m(s)$, is

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

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

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

To model θ using a DPC we use the following latent processes: α , which describes the average annual SST; the vector $\beta_t \in \mathbb{R}^4$, 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 latent process σ . This yields

$$\theta_{m,y}(s) \sim N \left(\sum_j K[s-j, \Lambda(s)] (\alpha(j) + \beta_t(j) w_t^T + \eta(j) \times (t - 180)), \sum_j K[s-j, \Omega(s)] \exp(\sigma(j)) \right). \quad (3)$$

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

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

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

$$\begin{aligned} \Lambda(s) &= \sum_j K[s-j, u] \kappa(j), \\ \Omega(s) &= \sum_j K[s-j, u] \rho(j), \end{aligned}$$

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

$$\beta_t \sim N(\beta_{t-1}, W_t), \quad (4)$$

with the discount factor $\nu(j)$ applied to the diagonal evolution variance matrix W (West and Harrison 1997, chap. 2). To

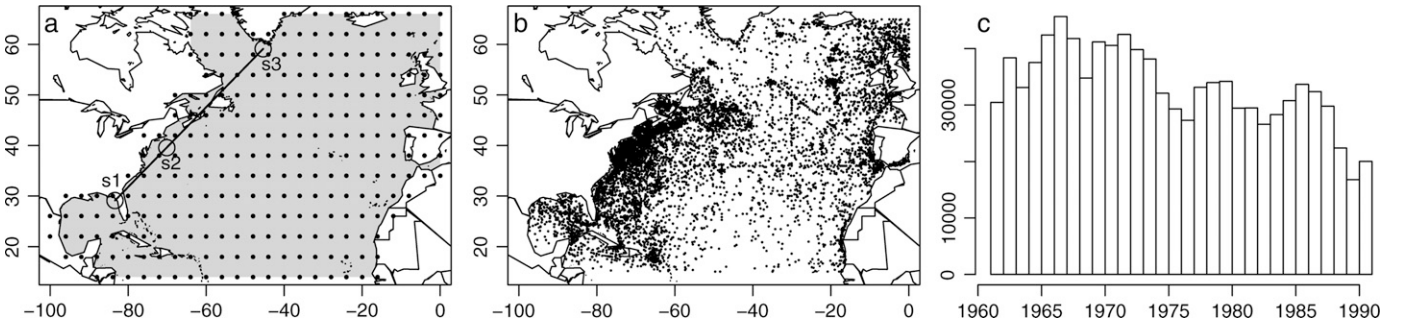


Figure 4. (a) The domain S (gray area), the grid J (bullets; $n_J = 270$, $r_J = 4^\circ$), a transect (line) and three “case study” points (open circles); (b) spatial distribution of a random sample with 1% of the total number of observations used in the model; (c) temporal distribution of the full dataset.

provide some spatial coherence in the variability of α , we use a Gaussian Markov random field (MRF). Thus,

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

Here, $N(j)$ denotes the point in J immediately to the north of j , and so forth. We couple β_1 , η , and σ to analogous MRF structures. Tables 3 and 4 contain a list and short description of all the model’s processes/parameters and derived quantities, respectively.

3.3 Model Implementation and Fit Assessment

For τ_α^2 , $\tau_{\beta_1}^2$, and τ_σ^2 , we use inverse Gamma(1, 0.1) prior distributions, as they provide little prior information. To ensure a smooth trend field, τ_η^2 receives a narrower, but still vague, inverse Gamma(2, 0.002) prior. We use a Uniform (0.9, 1) prior for v , as this support covers the factor’s natural range of variability. The latent processes 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. To ρ we assign priors analogous to κ .

We explore the posterior joint distribution of all model quantities by MCMC. Closed form full conditional distributions for the Gibbs sampler are provided in the Appendix. We sample from the joint distribution of $(\beta_1, \dots, \beta_{360})$ 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 (3) and evolution equation obtained from (4). We update the samples of κ , ρ , v , and σ with Metropolis-Hastings steps, using truncated normals as jumping distributions and setting their variances upon a pilot run of 3,000 iterations.

Table 2. NODC datasets and observational variances. See text for acronyms.

Data Set Name	Number of Observations	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}

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 (1,200 iterations), thin the chain (1/3), check stationarity, and set the adequate sample size (6,000 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 ($\alpha(j) = 30$, for all j), low discount factors ($v(j) = 0.91$), and high variance parameters ($\tau_\alpha^2, \tau_{\beta_1}^2, \tau_\sigma^2 = 10$, $\tau_\eta^2 = 0.01$), whereas the other starts with a cooler ocean ($\alpha(j) = 15$) and average discount factors ($v(j) = 0.95$) and variances ($\tau_\alpha^2, \tau_{\beta_1}^2, \tau_\sigma^2 = 0.1$, $\tau_\eta^2 = 0.001$). Both runs 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 posterior predictive p -values (PPPs; Gelman, Carlin, Stern, and Rubin 2004, chap. 6)

$$p_{i,m,y}^B(s) = \Pr(x_{i,m,y}^{rep}(s) \geq x_{i,m,y}(s) | \text{data}).$$

In this expression, $x_{i,m,y}^{rep}(s)$ is a replicated datum, simulated from the posterior predictive distribution, $x_{i,m,y}(s)$ is an actual SST measurement. If the prediction produced by the model for month m and year y at location s is compatible with the corresponding observation, $p_{i,m,y}^B(s)$ should not be too small or too large. PPPs have been criticized for not being uniformly distributed, a fact that limits their use for formal model assessment. Unfortunately, corrections like the one proposed in Hjort, Dahl, and Steinbakk (2006) can depend strongly on the prior distributions and are too computationally demanding to be used in this application. The alternative proposed in Johnson (2007), based on pivotal quantities, is not easily implemented for data with space and time dependencies.

To explore the influence of different grid resolutions and kernel shapes on the resulting climatological fields, we perform two tests on small domains, plus a basin-wide test. In the basin-wide test, we replace our flexible kernels with Bézier kernels with smoothness parameter 2, i.e., we equate $\Omega(s)$ and $\Lambda(s)$ to $(2, r_J, r_J, 0)$, for all s . Hence, we exclude the latent processes

Table 3. Model processes and variance parameters.

Process	Description	Variance Parameters
$\theta_{m,y}(s)$	True SST	—
—	SST measurement variance (fixed)	τ_i^2
$\kappa(j)$	Kernel shape for mean SST	—
$\alpha(j)$	Annual mean SST	τ_α^2
$\beta_i(j)$	Seasonal cycle	$\tau_{\beta_i}^2$
$\eta(j)$	Long-term linear trend	τ_η^2
$\sigma(j)$	SST variance	τ_σ^2
$\nu(j)$	Discount factor for the cycle	—
$\rho(j)$	Kernel shape for SST variance	—

κ and ρ from the model. This yields a considerable reduction in computation time and brings the model, dubbed “low resolution, Bézier kernels model”, closer to WOA, which also uses isotropic kernels. We consider an even simpler model, with no short or long term trends:

$$\theta_{m,y}(s) \sim N(\Xi_m^*(s), 1),$$

where $\Xi_m^*(s)$ are the WOA climatologies. We call this the “time-stationary model.” To perform a quantitative comparison between the fit provided by the three models, we use the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, and van der Linde 2002) and the posterior predictive loss criterion (PPLC; Gelfand and Ghosh 1998).

In the small domain tests, we analyze into greater depth the findings described in the 1D toy example described in 3.1, *viz.* that the fit provided by larger kernels depends more on data density, and that increasing the resolution yields more rugged surfaces—which can be an asset or a disadvantage. Hence, we select two regions with the same latitudinal bounds, 20°N–32°N, but with very different data densities. Region one (R1), located off Florida, covers longitudes 76°W to 88°W and has an abundance of observations (71,181). Region two (R2), located in the center of the North Atlantic gyre, goes from 40°W to 52°W and has fewer observations (5,749). For each region separately, we fit our model—dubbed “low resolution, small kernels model”—and three others: the first is the low resolution, Bézier kernels model; the second, called “low resolution, large kernels model”, has the same grid resolution as ours (4°) and flexible kernels as well, but the upper bound for the kernel’s semimajor axis is twice the grid resolution (8°); the third, called “high resolution, large kernels model”, uses a 2° grid and also lets the kernel’s semimajor axis reach twice the grid resolution (4°). To compare models we use DIC, PPLC, and we compare the resulting climatologies visually.

Table 4. Derived quantities.

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

4. RESULTS

4.1 Climatological, Anomaly, and Trend Fields

Figure 5 compares North Atlantic climatologies produced with our method and WOA. In terms of the means for January and July (Figures 5a through d), the two products display the same large scale patterns, but ours presents more sinuous contours, especially in the northwest Atlantic. Along the transect drawn in Figure 4, other differences become clear (Figure 5e): our climatology presents stronger 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 5f): whereas in s_3 there is a fairly good agreement between the two products, s_1 and s_2 present disparities. In s_2 the WOA cycle is systematically cooler than ours, whereas in 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 6a). The extremes of these trends are found in the Northwest Atlantic: between 1961 and 1990, SST warmed as much as 2.6°C off the mouth of the Chesapeake Bay, and cooled as much as –5°C north of Newfoundland. The 95% posterior intervals for these trends are (2.4, 2.8)°C and (–7.0, –3.3)°C, respectively, which reflect the difference in the amount of sampling (Figure 4b).

In 2° × 2° squares centered at s_1 , s_2 , and s_3 , the proportion of the total data collected after 1975 (the midpoint of the time frame analyzed) was respectively 26%, 34%, and 43%. In light of this information and the trends presented in Figure 6a, we may attribute the results in Figure 5d to sampling error. In other words, we postulate that undersampling during the cold (warm) period in the vicinity of s_1 (s_2) may have introduced a warm (cold) bias on WOA. Greater equilibrium in the temporal distribution of data around s_3 justifies the proximity between the two climatologies.

As stated in Section 3.2, β describes the annual cycle and nonlinear trends. In the neighborhood of s_2 , seasonality is strong and β_1 —the component associated with $\sin(2\pi t/12)$ —dominates the signal (Figures 5f and 6b). 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 later), and thus it is not surprising that transient anomalies display spatial patterns different from long-term trends. Figure 6c, for example, provides a snapshot of mean SST anomalies in July 1979. Here we observe that most anomalies are close to zero, indicating that the North Atlantic is mostly close to its climatological state. The Gulf Stream and the North Atlantic Current carry water warmer than average, as indicated by positive anomalies, whereas negative anomalies at the boundaries of those currents indicate that the ocean is cooler than average. Returning to Figure 6b, the component β_1 stands out for presenting a trend, which dampens the amplitude of the 12 months harmonics, making winters

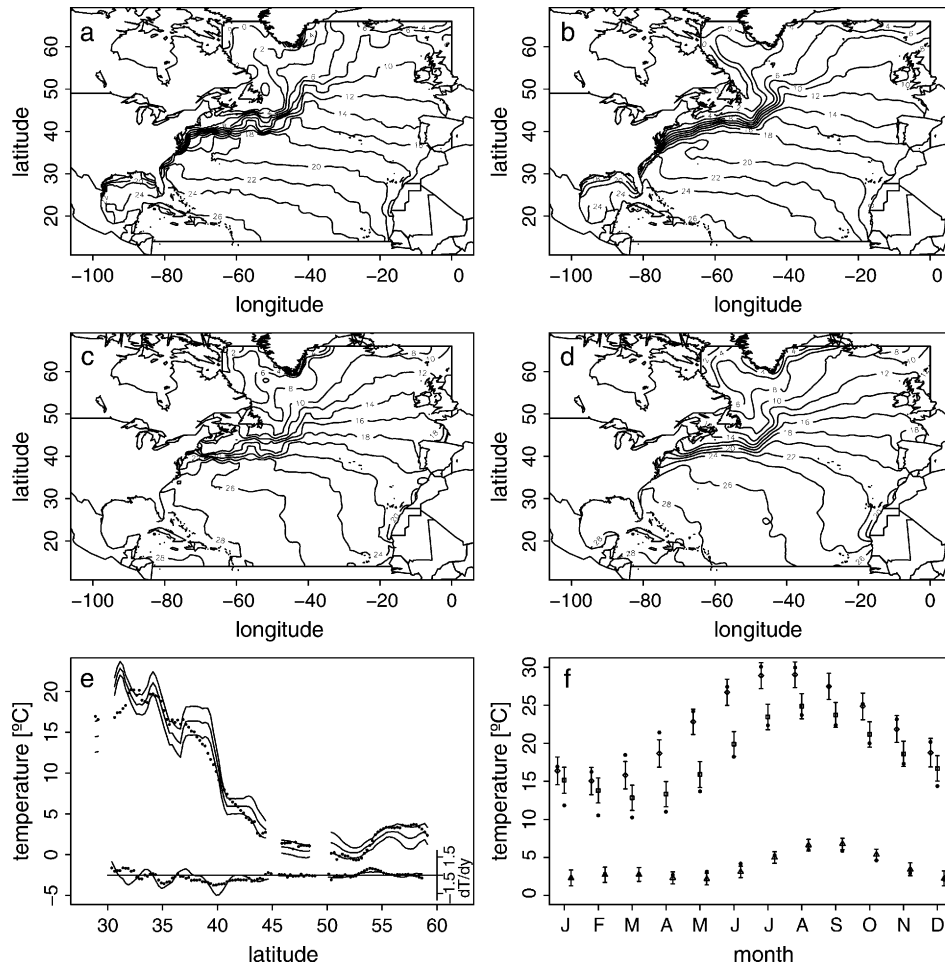


Figure 5. SST climatologies. (a) January mean for the North Atlantic, results from our model. (b) WOA January mean. (c) July mean, results from our model. (d) WOA July mean. (e) January climatology along the transect depicted in Figure 4. Lines denote mean and 95% posterior intervals from our model, bullets denote WOA. (f) Seasonal cycle in the points depicted in Figure 4. Posterior means: \diamond for s_1 , \square for s_2 , and \triangle for s_3 . Whiskers denote 95% posterior intervals from our model, bullets denote WOA estimates.

warmer and summers cooler. When combined with the positive linear trend (Figure 6a), this intensifies winter warming and weakens summer SST change, as Figure 6d shows.

4.2 Kernels, Discount Factors, and Variance Parameters

The posterior distributions of κ and ρ are alike, and thus Figure 7a shows information relative only to the former. With 20 random draws of κ from the stationary MCMC, we construct 20 ellipses for each point $j \in J$, by equating $K^*[s - j, \kappa(j)] = 0.5$. We use these ellipses to investigate the orientation and size of the convolution kernel for monthly mean SST, 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 Λ , defined as $\varepsilon = \sqrt{1 - \Lambda_2^2 / \Lambda_1^2}$. 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 7b we draw annual mean sea surface currents data from Mariano Global Surface Velocity Analysis (MGSVA; Mariano and Brown 1992; Mariano, Ryan, Smithers, and Perkins 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 7a 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 (NADC) and enveloping currents, *viz.* the Labrador Current (LC), the East Greenland Current (EGC), and the West Greenland Current (WGC), are also discernible, and so is the North Atlantic Drift Current. Furthermore, the kernels' semimajor axes are often parallel to the direction of flow. This seems a sensible result, because points in the same streamline should have weaker SST differences than points in different streamlines. We find the same features in ρ , which reinforces the link between SST variability and the dominating patterns of currents.

Recall that $\nu \approx 1$ indicates that β_t is close to constant in time, and $\nu \approx 0.9$ indicates large variability in β_t between consecutive

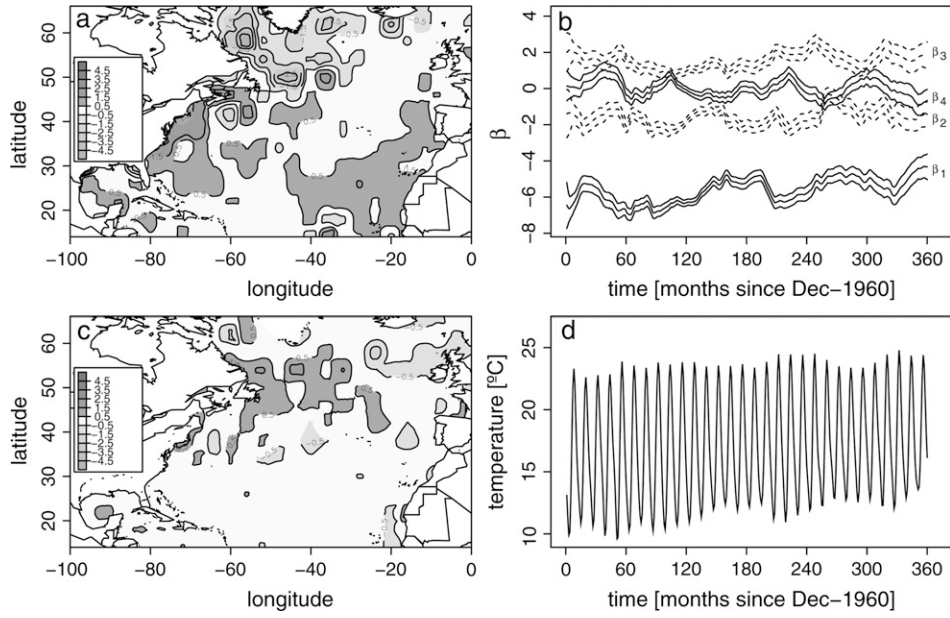


Figure 6. Linear and nonlinear SST trends. (a) Posterior mean long-term linear trend ($^{\circ}\text{C}/30\text{ yr}$). (b) Posterior mean and 95% intervals of $\beta_i(j)$, $j = (72^{\circ}\text{W}, 38^{\circ}\text{N})$. (c) Posterior mean anomalies in July 1979 ($^{\circ}\text{C}$). (d) Posterior mean for monthly SST at s_2 ($^{\circ}\text{C}$).

times. Although we provided the same prior distribution to all $v(j)$, 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 v substantiates our expectation (Figure 7c). With a larger dataset, we believe the spatial heterogeneity would be more pronounced.

The posterior mean of $\Phi(s)$ clearly displays the mark of the Gulf Stream and, to a lesser extent, other strong currents (Figure 7d). Much of the variability in these regions is lost to error because, even with anisotropic kernels, the surfaces

produced for $\theta_{m,y}(s)$ from latent processes located on a 4° grid cannot resolve sharp meanders (i.e., wave-like features evident in ocean currents) and eddies (i.e., rotating parcels of fluid with diameters up to 400 km). Moreover, upwelling events (i.e., movement of cold water from the bottom to the surface) and other transient features cannot be described adequately with a temporal resolution of 1 month. These losses of information are common when constructing basin scale climatologies, because the heterogeneous spatio-temporal distribution of data only allows finer resolutions up to a point, as our experiments with 3° and 2° grids proved.

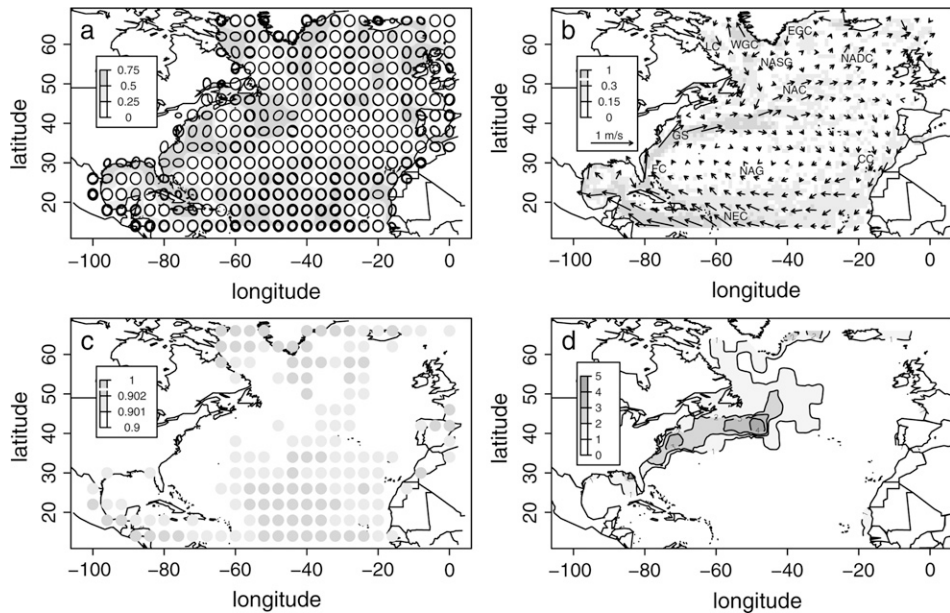


Figure 7. (a) Mean eccentricity of the kernels associated with $\Lambda(s)$ (shaded map), and 20 realizations of the kernel at every point of J . (b) Direction (vectors drawn at points of J) and speed (vector length and shaded map) of annual mean surface currents, according to MGSVA (Mariano and Brown, 1992; Mariano et al. 1995). See text for acronyms. (c) Posterior mean of $v(j)$. (d) Posterior mean of $\Phi(s)$.

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 month cycle is more spatially variable than the 6 month, and the sine component is more spatially variable than its cosine counterpart.

4.3 Convergence and Model Assessment

Owing to their massive size, only some samples of $\theta(s)$ and $\beta_i(j)$ were stored for convergence analysis. For those processes, as well as for all other model quantities, our choices of burn-in, length of the chain, and thinning seem acceptable. Our selection of initial values proves not to influence the outcome, as the two runs (“cold” versus “warm start”) described in 3.3 yield practically the same results. More than 95% of the quantities pass at least three out of four BOA tests. Detailed tables and trace plots are provided in <http://www.ams.ucsc.edu/~bruno/climatology/>.

Figure 8 displays the spatial distribution of the 704 PPPs that fall outside the range (0.005, 0.995). Clusters occur mainly in regions with strong currents (cf. Figure 7b) and moderate sampling, which we do not find surprising given the coarseness of the model, as discussed previously. Plots of extreme PPPs for different seasons, decades, and instruments do not indicate any other localized lack of fit. Several of these PPPs belong to the same cruise, suggesting that more stringent quality control criteria could have been used to remove erroneous observations. Another conclusion is that a multiresolution model might be an interesting approach to constructing SST climatologies. The issue is thus how to define an adequate grid for the problem at hand, and what type of weighting kernel to use.

To shed some light into this matter, we compare our model with others, both easier and harder to fit. At the basin-wide scale, our model performs better than the low resolution, Bézier kernels model and far better than the time-stationary model, according to DIC and PPLC (Table 6). We include this third model not only to provide a baseline for the comparison of the other two but also to confirm that the consideration of interannual variability is an essential component of SST modeling, even when the main interest is in producing climatologies. The computation time, on the other hand, is unfavorable to our model, because 75% of total computation time is spent on the latent processes that control the kernels’ shape. If we look at localized fits (Figure 9, first two rows), we see that in regions with weak currents, such as the North Atlantic Gyre, the two climatologies

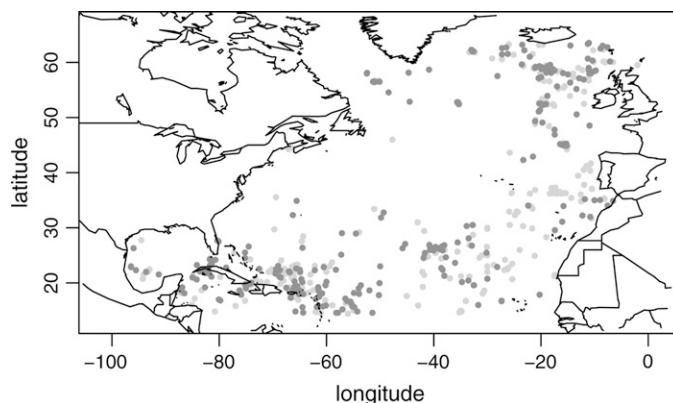


Figure 8. Spatial distribution of posterior predictive p -values smaller (greater) than 0.005 (0.995), in light (dark) gray.

are virtually identical, as expected. The major differences emerge around Florida, where the Bézier kernels model essentially produces an oblong bubble of cold water, whereas our (low resolution, small kernels) model allows isotherms to follow the bathymetry more closely. Further tests with isotropic kernels show that these have difficulty in capturing gradients that are oblique to the grid axes, such as the one west of Florida. Hence, rotating the grid by 45° would mitigate fit problems in some areas but create others. In other words, the isotropic kernel model is more sensitive to the grid, and the resulting convolution surfaces reflect that through less realistic contours.

The difference in tractability becomes even greater when models that allow large-support kernels are considered, because in these the DPC involve more terms and the parallelization procedure is less efficient. Nonetheless, a substantial payoff emerges in data rich regions, with the high resolution grid model producing the best fit (left columns in Figure 9 and Table 7). On the other hand, these models behave poorly in data sparse regions: for example, DIC and PPLC give indication of some overfitting in the North Atlantic Gyre—the 2° grid model now ranks last according to DIC, and third according to PPLC. More importantly, the resulting climatologies present bull’s-eyes and jagged contours (Figure 9, right column panels). In conclusion, our choices of grid resolution and maximum kernel support seem to produce an overall adequate fit, and the added computational effort to pursue anisotropy proves worthwhile.

5. FINAL REMARKS

Climatologies are valuable data based products for the ocean scientific community, because they summarize what is known from the observational record. 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

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

Parameter	Posterior mean	95% Posterior Interval
τ_α^2	3.782	(3.154, 4.501)
$\tau_{\beta_{1,1}}^2$	1.070	(0.804, 1.401)
$\tau_{\beta_{1,2}}^2$	0.735	(0.528, 1.000)
$\tau_{\beta_{1,3}}^2$	0.347	(0.236, 0.498)
$\tau_{\beta_{1,4}}^2$	0.197	(0.127, 0.289)
τ_η^2	2.58×10^{-5}	$(2.15 \times 10^{-5}, 3.09 \times 10^{-5})$
τ_σ^2	0.529	(0.440, 0.633)

Table 6. Model comparison criteria for the North Atlantic domain.

	DIC	PPLC
Low resolution, small kernels model	1,960,543	262,427
Low resolution, Bézier kernels model	1,965,571	263,933
Time-stationary model	1.7×10^{10}	406,635

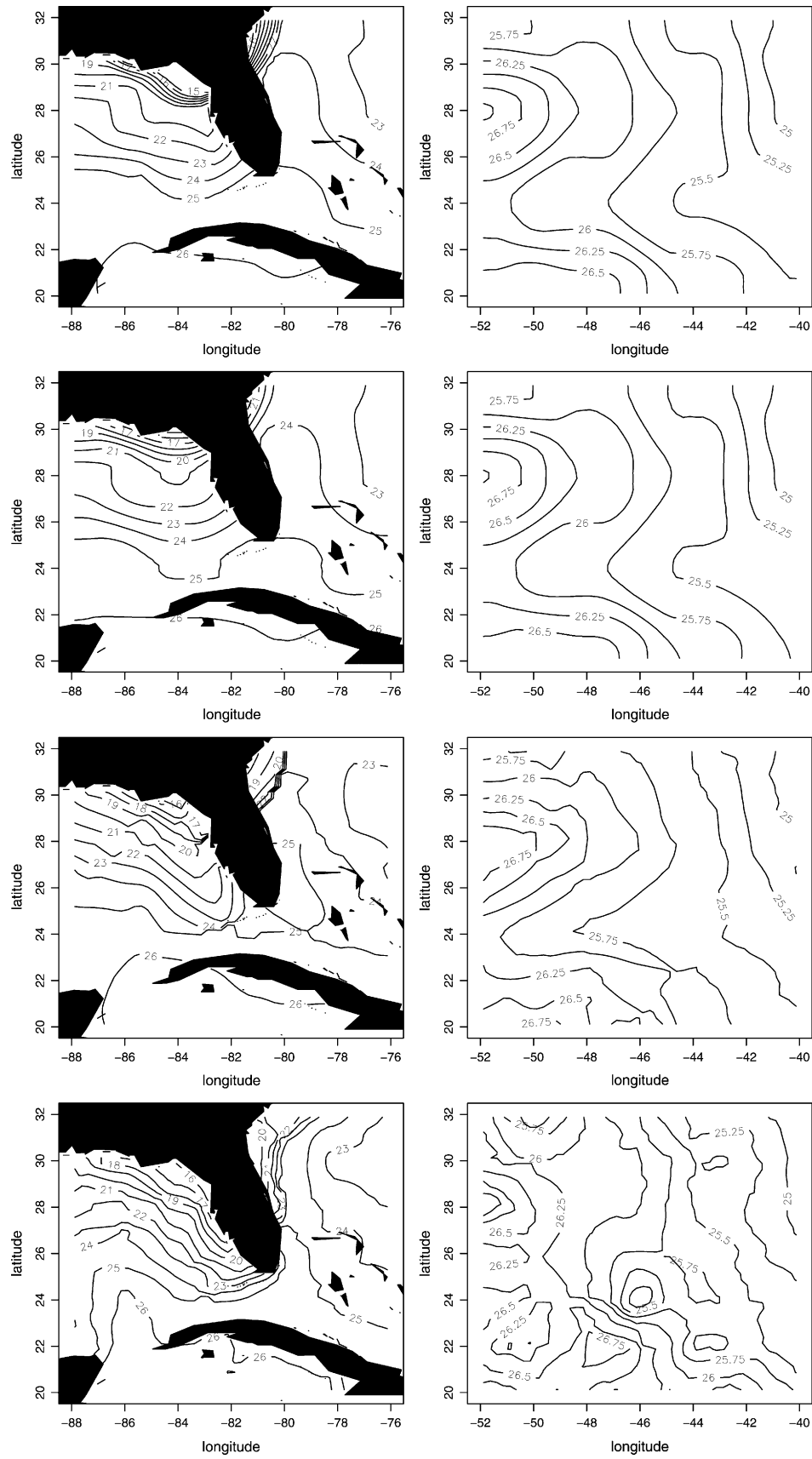


Figure 9. Posterior mean January climatologies for R1 (left panels) and July climatologies for R2 (right panels), resulting from the low resolution, small kernels model (first row), the low resolution, Bézier kernels model (second row), the low resolution, large kernels model (third row), and the high resolution, large kernels model (fourth row).

Table 7. Model comparison criteria for R1 and R2

	DIC _{R1}	PPLC _{R1}	DIC _{R2}	PPLC _{R2}
Low resolution, small kernels model	131,028	11,306	10,467	983
Low resolution, Bézier kernels model	131,470	11,341	10,476	994
Low resolution, large kernels model	129,797	11,085	10,422	954
High resolution, large kernels model	128,416	10,766	10,502	987

article 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, that is, with each layer and property being analyzed 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 with other methods (e.g., Levitus, Antonov, and Boyer 1994; Grey, Haines, and Troccoli 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 before averaging them into monthly climatologies, similarly to Higdon (1998). In this manner, it incorporates our knowledge that temperature fluctuates over time, and mitigates biasing toward 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, nonstationarity and location dependent anisotropy are other features that should be accounted for when constructing an 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), whereas 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 an MCMC setting for posterior sampling, the number of iterations in the procedure becomes subject to convergence analyses. The initial conditions become of no importance, as our experiment with 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 processes that control the kernels’ shape in the model (κ and ρ), we follow other suggestions made by Higdon (1998) to improve spatio-temporal 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 dataset: with 13 computing processors working on 8° longitudinal strips of the North Atlantic, the MCMC algorithm is able to run five times faster than with just one. To ensure the kernels vary smoothly over space, we also allow the processes 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 that 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 datasets, 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 depart from traditional climatological analyses, where fluid dynamics are not used, and make it more similar to ocean data assimilation models.

APPENDIX: FULL CONDITIONAL DISTRIBUTIONS

$$\bullet (\theta_{m,y}(s) | \dots) \sim N(D_\theta d_\theta, D_\theta)$$

$$D_\theta^{-1} = \frac{1}{\sum_j K[s-j, \Lambda(s)] \exp(\sigma(j))} + \frac{1}{\tau_i^2}$$

$$d_\theta = \frac{\sum_j K[s-j, \Lambda(s)] (\alpha(j) + \beta_i(j) \mathbf{w}_i^T + \eta(j)(t-180))}{\sum_j K[s-j, \Lambda(s)] \exp(\sigma(j))} + \frac{x_{i,m,y}(s)}{\tau_i^2},$$

$$\bullet (\alpha(j) | \dots) \sim N(D_\alpha d_\alpha, D_\alpha)$$

$$D_\alpha^{-1} = \frac{1}{\tau_\alpha^2} + \sum_s \frac{K^2[s-j, \Lambda(s)]}{\sum_{j' \in J} K[s-j', \Lambda(s)] \exp(\sigma(j'))}$$

$$d_\alpha = \frac{\alpha(N(j)) + \alpha(S(j)) + \alpha(E(j)) + \alpha(W(j))}{4\tau_\alpha^2}$$

$$+ \sum_s \frac{K[s-j, \Lambda(s)] (\theta_{m,y}(s) - C_\alpha)}{\sum_{j' \in J} K[s-j', \Lambda(s)] \exp(\sigma(j'))},$$

$$C_\alpha = K[s-j, \Lambda(s)] (\beta_i(j) \mathbf{w}_i^T + \eta(j)(t-180)) + \sum_{j' \in J \setminus \{j\}} K[s-j', \Lambda(s)] (\alpha(j') + \beta_i(j') \mathbf{w}_i^T + \eta(j')(t-180))$$

- $(\eta(j) | \dots) \sim N(D_\eta d_\eta, D_\eta)$
- $D_\eta^{-1} = \frac{1}{\tau_\eta^2} + \sum_s \frac{K^2[s - j, \Lambda(s)]}{\sum_{j' \in J} K[s - j', \Lambda(s)] \exp(\sigma(j'))}$
- $d_\eta = \frac{\eta(N(j)) + \eta(S(j)) + \eta(E(j)) + \eta(W(j))}{4\tau_\eta^2}$
- $+ \sum_s \frac{K[s - j, \Lambda(s)](\theta_{m,y}(s) - C_\eta)}{\sum_{j' \in J} K[s - j', \Lambda(s)] \exp(\sigma(j'))}$
- $C_\eta = K[s - j, \Lambda(s)](\alpha(j) + \beta_i(j)w_i^T)$
- $+ \sum_{j' \in J \setminus \{j\}} K[s - j', \Lambda(s)](\alpha(j') + \beta_i(j')w_i^T$
- $+ \eta(j') (t - 180))$
- $(\tau_\alpha^2 | \dots) \sim IG\left(\frac{1}{2}(n_0 + n_J), \frac{1}{2}\left(n_0 S_0 + \sum_j \left(\alpha(j) - \frac{\alpha(N(j)) + \alpha(S(j)) + \alpha(E(j)) + \alpha(W(j))}{4}\right)^2\right)\right)$
- Let τ_k^2 and β_k designate the k -th component
- of the vectors $\tau_{\beta_1}^2$ and $\beta_1, k = 1, \dots, 4$. Then
- $(\tau_k^2 | \dots) \sim IG\left(\frac{1}{2}(n_0 + n_J), \frac{1}{2}\left(n_0 S_0 + \sum_j \left(\beta_k(j) - \frac{\beta(N(j)) + \beta(S(j)) + \beta(E(j)) + \beta(W(j))}{4}\right)^2\right)\right)$
- $(\tau_\eta^2 | \dots) \sim IG\left(\frac{1}{2}(n_0 + n_J), \frac{1}{2}\left(n_0 S_0 + \sum_j \left(\eta(j) - \frac{\eta(N(j)) + \eta(S(j)) + \eta(E(j)) + \eta(W(j))}{4}\right)^2\right)\right)$

[Received January 2008. Revised June 2008.]

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," *International Journal of Climatology*, 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," *Freiburg Online Geoscience*, 6, 1–108.
- Brooks, S., and Gelman, A. (1998), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.
- Brown, P. J., Le, N. D., and Zidek, J. V. (1995), "Multivariate Spatial Interpolation and Exposure to Air Pollutants," *The Canadian Journal of Statistics*, 22, 489–509.
- Casey, K. S., and Cornillon, P. (2001), "Global and Regional Sea Surface Temperature Trends," *Journal of Climate*, 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.
- (2007), "Approximate Likelihood for Large Irregularly Spaced Spatial Data," *Journal of the American Statistical Association*, 102, 321–331.
- Gamerman, D., and Lopes, H. F. (2006), *Markov Chain Monte Carlo—Stochastic Simulation for Bayesian Inference* (2nd ed.), London: Chapman and Hall.
- Gaspari, G., and Cohn, S. E. (1999), "Construction of Correlation Functions in Two and Three Dimensions," *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757.
- Gelfand, A., and Ghosh, S. (1998), "Model Choice: A Minimum Posterior Predictive Loss Approach," *Biometrika*, 85, 1–11.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis* (2nd ed.), London, UK: Chapman & Hall/CRC.
- Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472.
- Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments," in *Bayesian Statistics 7*, eds. J. M. Bernardo, J. O. Berger, P. Dawid, A. F. M. Smith, and M. West, Oxford: Clarendon Press.
- Gouretski, V. V., and Janke, K. (1999), "A Description and Quality Assessment of the Historical Hydrographic Data for the South Pacific Ocean," *Journal of Atmospheric and Oceanic Technology*, 16 (11), 1791–1815.
- Gouretski, V. V., and Koltermann, K. P. (2007), "How Much Is the Ocean Really Warming," *Geophysical Research Letters*, 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," *Journal of Climate*, 13 (15), 2697–2711.
- Heidelberger, P., and Welch, P. (1983), "Simulation Run Length Control in the Presence of an Initial Transient," *Operations Research*, 31, 1109–1144.
- 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 *Quantitative Methods for Current Environmental Issues*, eds. C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-haarawi, London: Springer Verlag, pp 37–56.
- Higdon, D. M., Swall, J., and Kern, J. (1999), "Non-Stationary Spatial Modeling," in *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, P. Dawid and A. F. M. Smith, Oxford, UK: Oxford University Press, pp 761–768.
- Hjort, N. L., Dahl, F. A., and Steinbakk, G. H. (2006), "Post-Processing Posterior Predictive P Values," *Journal of the American Statistical Association*, 101, 1157–1174.
- Holbrook, N. J., and Bindoff, N. L. (2000), "A Statistically Efficient Mapping Technique for Four-Dimensional Ocean Temperature Data," *Journal of Atmospheric and Oceanic Technology*, 17, 831–846.
- Johannesson, G., Cressie, N., and Huang, H.-C. (2007), "Dynamic Multi-resolution Spatial Models," *Environmental and Ecological Statistics*, 14, 5–25.
- Johnson, V. (2007), "Bayesian Model Assessment Using Pivotal Quantities," *Bayesian Analysis (Online)*, 2, 719–734.
- Kushnir, Y. (1994), "Interdecadal Variations in North Atlantic Sea Surface Temperature and Associated Atmospheric Conditions," *Journal of Climate*, 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," *Progress in Oceanography*, 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 Research*, 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," *Statistical Modelling*, 2, 315–332.
- Paciorek, C. J. (2007), "Bayesian Smoothing with Gaussian Processes Using Fourier Basis Functions in the Spectral GP package," *Journal of Statistical Software*, 19, 1–38.
- 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," *Journal of Climate*, 18, 4562–4581.
- R Development Core Team (2005), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing, 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," *Statistical Science*, 7, 493–497.
- (1992b), "How Many Iterations in the Gibbs Sampler?" in *Bayesian Statistics 4*, eds. J. M. Bernardo, J. O. Berger, P. Dawid, and A. F. Smith, Oxford, UK: Oxford University Press, pp. 765–776.
- 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," *Journal of Climate*, 15, 1609–1625.
- Ridgway, K. R., Dunn, J. R., and Wilkin, J. L. (2002), "Ocean Interpolation by Four-Dimensional Weighted 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*, doi: 10.1002/env.895.
- Sampson, P. D., and Guttorp, P. (1992), "Nonparametric Estimation of Non-stationary Spatial Covariance Structure," *Journal of the American Statistical Association*, 87, 108–119.
- Schmidt, A. M., and O'Hagan, A. (2003), "Bayesian Inference for Non-stationary Spatial Covariance Structures Via Spatial Deformations," *Journal of the Royal Statistical Society, Ser. B*, 65, 743–758.
- Smith, B. J. (2005), "*Bayesian Output Analysis program (BOA) for MCMC*. R package version 1.1.5-2.
- Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002), Bayesian Measures of Model Complexity and Fit (with discussion). *Journal of the Royal Statistical Society, Ser. B*, 64: 583–639.
- Tzeng, S., Huang, H.-C., and Cressie, N. (2005), "A Fast, Optimal Spatial-Prediction Method for Massive Datasets," *Journal of the American Statistical Association*, 100, 1343–1357.
- West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models*, (2nd ed.), New York: Springer Verlag.
- Wikle, C. K., Millif, R. F., Nychka, D., and Berliner, L. M. (2001), "Spatio-temporal Hierarchical Bayesian Modeling: Tropical Ocean Winds," *Journal of the American Statistical Association*, 96, 382–397.
- Xun, Q. (2005), "Representations of Inverse Covariances by Differential Operators," *Advances in Atmospheric Sciences*, 22, 181–198.

Comment

D. HIGDON

My thanks to Dr. Lemos and Dr. Sansó for their excellent work on modeling sea surface temperatures (SSTs) in the North Atlantic over space and time. The computational challenge of integrating more than one million temperature measurements distributed haphazardly over a space and time model is formidable. I commend the authors on their success in developing a modeling strategy and actually making it work on this difficult problem.

This work has delivered a number of notable innovations:

- A discrete process convolution (DPC) representation of the SST for each month of a 30-year span
- A dynamic linear model to account for temporal variation between months
- A nonstationary spatial dependence structure
- A distributed, computational approach for model estimation that takes advantage of the conditional dependence structure of this integrated space–time model.

Although each of these features is nice, it is their thoughtful combination that makes this effort noteworthy.

The principled statistical modeling approach of the authors has a number of advantages over the ad hoc objective analysis-based approaches: It's adaptive, we learn about spatial and temporal dependence, and a probabilistic description is made of the entire space–time field of surface temperatures. Hence, the resulting climatologies are the result of a statistically based data combination, rather than ad hoc rules for constructing averages over space and time. It is also worth remembering that even one million observations spread out over 40 years across the entire North Atlantic gives rather sparse coverage. Any sensible estimation of this space–time field is going to require thoughtful modeling, as described here, to fill in over sparsely sampled regions.

This analysis treads in an important area for us statisticians to consider: the interface between computational and statistical science. In a sense, the DPC can be thought of as a Gaussian process model with a statistically motivated sparse, approximate representation of a large covariance matrix. Both the spacing of the knots and the specification of the convolution kernel affect the statistical properties of the model as well as what computational strategies should be used.

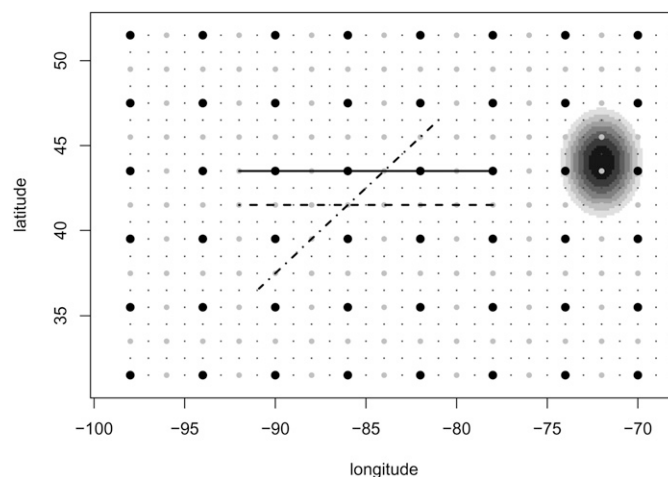


Figure 1. We consider three DPC models constructed from successively denser grids. The successively denser grids are denoted $1 \times$ (4-degree spacing, black dots), $4 \times$ (2-degree spacing, black and gray dots), and $16 \times$ (1-degree spacing, all dots). A single kernel is used for all locations and corresponds to $\omega = (2, 4/\sqrt{2}, 4, 0)$. We will consider properties of the resulting DPC $z(s)$ over the three transects shown in the figure for each of the grid densities. We will consider $z(s)$ as s moves from right to left along the transect. Each transect is 14 "Mercator" degrees in length.