Conditionally Linear Models for Non-Homogeneous Spatial Random Fields

Ricardo T. Lemos and Bruno Sansó *

Abstract

We consider parsimonious representations of non-homogeneous spatial random fields. We focus on processes that can be represented as linear combinations of basis functions. As the basis functions are allowed to depend on unknown parameters, we identify such models with conditionally linear processes. We present a detailed description of an approach that uses discrete process convolutions with spatially varying, compactly supported kernels. We discuss the similarities and differences between this approach and the predictive Gaussian process approach. We also discuss the problem of obtaining decompositions of a spatial random field, as well as spatio-temporal extensions of our spatial models.

Key Words: Spatial Random Fields; Spatial Interpolation; Bayesian Inference.

1 Introduction

Following the seminal work of Matheron (Matheron, 1963), the field of geostatistics has seen an impressive development and has found successful applications in a large number of fields. With the advancement and increasing availability of geographical information systems as well as observations from remote sensors, large datasets of spatially referenced scientific data have become very common. In addition, computer models for the simulation, analysis and forecast of spatio-temporal phenomena, like weather, climate, dispersion of pollutants, dynamics of oceans' characteristics, among others, produce massive output that often need to be summarized or compared and combined with observations. Gelfand et al. (2010) provide a collection of state of the art statistical methods for spatial data.

Geostatistical methods consider random fields that are indexed in space, usually 2D or 3D, and more generally space and time. Intuitively, proximity between observations should

*Ricardo Lemos is a JIMAR-PFEL research oceanographer, National Marine Fisheries Service, Pacific Grove, California, U.S.A. E-mail rtl@net.sapo.pt, Bruno Sansó is 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

provide information for inference about unobserved values of the field. This is formalized by the use of covariance structures that depend on location. Covariance functions play a fundamental role in the traditional kriging methods, as explained, for example, in Cressie (1993). Unfortunately, the need to consider increasingly large data sets requires computations involving very large covariance matrices. These computations may not be feasible, even with the use of clever computational tricks. An application for observations at, say, 10,000 locations will produce a covariance matrix with 50,005,000 possibly different values. Such a large data structure would need to be stored, decomposed and operated with, possibly within an iterative procedure.

The problem of fitting a spatial random field for a large number of locations has been considered by a number of authors. For regularly spaced data, and under the assumption that the covariance matrix has strong symmetric properties like isotropy, it is possible to perform very fast computations with the help the Fast Fourier Transform (Rue and Held, 2005, Section 2.6). Fast computations can also be performed taking advantage of sparsity induced by compactly supported covariance functions (Gneiting, 2002). Sparsity of the inverse of the covariance matrix, or precision matrix, can be used for fast computations for Gaussian Markov random fields (Rue and Held, 2005). Suitably chosen bases of orthogonal functions, such as wavelets or Fourier expansions, have been used in Nychka et al. (2002) and Paciorek (2007). Careful use of multiresolution properties for random fields at different levels of aggregation are exploited in Tzeng et al. (2005), to obtain very fast prediction methods for fields on a grid. Spectral methods for the approximation of the likelihood, known as Whittle's approximation, have been used for several decades (Whittle, 1954). A recent extension to irregularly spaced data in presented in Fuentes (2007).

The traditional geostatistical approach to spatial data assumes that the field is isotropic. This implies very strong symmetries of the corresponding covariance function. Such an assumption is convenient from the inferential point of view, as it reduces the formulation of the covariance to just a handful of parameters. While isotropy results in surprisingly flexible predictive surfaces, it can be a very unrealistic assumption in may cases. An alternative is to assume isotropy on a non-linearly transformed space, as proposed in Sampson and Guttorp (1992). This approach uses cross-validation and has been extended to likelihood-based methods (Damian et al., 2001; Schmidt and O'Hagan, 2003). Brown et al. (1995) use multivariate normal models to estimate general covariance matrices. Yet another approach to achieve anisotropy is to consider locally isotropic models convolved with kernels. Fuentes (2002) proposes processes with spatially varying covariance parameters and Higdon et al. (1999) proposes location dependent convolving kernels. This is the approach used in Lemos and Sansó (2009), which will be illustrated and expanded in this paper. A related technique is proposed in Paciorek and Schervish (2006) where large classes of non-stationary covariance functions are created using convolutions. Piecewise isotropic process are used in conjunction with tree structures (Gramacy and Lee, 2008) or partitions obtained from tesselations (Kim et al., 2005), to obtain processes that are globally anisotropic.

The focus of this paper is on conditionally linear spatial processes (CLP). That is, processes that can be represented as linear combinations of basis functions. As such bases may

depend on unknown parameters, we denote the representation as conditionally linear. Fixed rank kriging processes (Cressie and Johannesson, 2008), predictive processes (Banerjee et al., 2008) and process convolution (Higdon, 2007) are all examples of CLPs. Here we consider hierarchical extensions to allow for spatially varying kernels, as presented in the next section. CLPs allow for effective reduction of the dimension of the parameter space without imposing unrealistic properties on the resulting covariance of the process. In Section 2.1 we discuss the ability of CLP to capture the main modes of spatial variability of a random field. In Section 2.3 we discuss the Monte Carlo based methods to fit the proposed models. In Section 2.2 we discuss extensions to space-time models. In Section 3 we present an example as well as comparisons with previously published results for some testbed data. In the last section we present our conclusions and discussion.

2 Conditionally linear spatial processes

A typical setting for the analysis of a spatial random field is given by the model

$$y(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^T\boldsymbol{\beta} + \omega(\boldsymbol{s}) + \varepsilon(\boldsymbol{s}) \tag{1}$$

where $\varepsilon(\boldsymbol{s}) \sim N(0, 1/\tau)$ is random uncorrelated noise, with τ denoting measurement precision; $\boldsymbol{x}(\boldsymbol{s})$ are location dependent covariates, linearly related to the response via a set of q-dimensional parameters $\boldsymbol{\beta}$, and $\omega(\boldsymbol{s})$ is a Gaussian process with mean 0 and covariance function $v(\cdot,\cdot)$. Here $\boldsymbol{s} \in \mathcal{S}$, which is a subset of $\mathbb{R}^d$, for d typically equal to 2 or 3. If we assume that $y(s)$ is observed at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ for a large n, then likelihood based inference for $y(s)$ requires the factorization of the $n \times n$ matrix $\boldsymbol{V}$, where $V_{i,j} = v(\boldsymbol{s}_i, \boldsymbol{s}_j)$. These computations could be unfeasible, in particular when Equation (1) is embedded in a hierarchical structure and iterative methods are used.

An interesting proposal for the reduction of the dimensionality of the problem is presented in Banerjee et al. (2008). There, a set of knots $\boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_m^*$ is chosen. Letting $\boldsymbol{\omega}^* = (\omega(\boldsymbol{s}_1^*), \ldots, \omega(\boldsymbol{s}_m^*))$ and $\boldsymbol{V}^*$ be the corresponding covariance matrix, a new process is defined by $E(\omega(\boldsymbol{s})|\boldsymbol{\omega}^*) = \boldsymbol{v}(\boldsymbol{s})^T(\boldsymbol{V}^*)^{-1}\boldsymbol{\omega}^*$ where $\boldsymbol{v}(\boldsymbol{s}) = (v(\boldsymbol{s}, \boldsymbol{s}_1^*), \ldots, v(\boldsymbol{s}, \boldsymbol{s}_m^*))$. The reduction in the dimension is achieved by substituting $E(\omega(\boldsymbol{s})|\boldsymbol{\omega}^*)$ for ω in Equation (1). The former approach is closely related to the one proposed in Cressie and Johannesson (2008) that consists on assuming that the covariance function v is given as $v(\boldsymbol{s}, \boldsymbol{s}') = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{K}\boldsymbol{B}(\boldsymbol{s}')$. Here $\boldsymbol{K}$ is a $r \times r$ positive definite matrix and $\boldsymbol{B}(\boldsymbol{s}) = (B_1(\boldsymbol{s}), \ldots, B_r(\boldsymbol{s}))^T$ is a set of basis functions, not necessarily orthogonal. Clearly, for $r = m$, $K_{ij} = v(\boldsymbol{s}_i^*, \boldsymbol{s}_j^*)$ and $B_j(\boldsymbol{s}) = [\boldsymbol{v}(\boldsymbol{s})^T(\boldsymbol{V}^*)^{-1}]_j$ we obtain the covariance induced by the approach in Banerjee et al. (2008). Furthermore, in both cases we have an equivalent linear representation of the process. In fact, following Stroud et al. (2001), we can write

$$\omega(\boldsymbol{s}) = \sum_{j=1}^{m} B_j(\boldsymbol{s})\gamma_j = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{\gamma}\ , \quad \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)^T \sim N_m(0, \boldsymbol{K})\ , \tag{2}$$

which implies that $\mathrm{cov}(\omega(\boldsymbol{s}), \omega(\boldsymbol{s}')) = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{K}\boldsymbol{B}(\boldsymbol{s}')$. Note that this corresponds to a non-isotropic covariance function, even if $\boldsymbol{K}$ is obtained by evaluating an isotropic covariance.

Writing $B_j(\boldsymbol{s}) = b(\boldsymbol{s} - \boldsymbol{s}_j^*)$ and $\gamma_j = \gamma(\boldsymbol{s}_j^*)$ we have a process convolution (Higdon, 2007). Equation (2) can be extended by letting $B_j(\boldsymbol{s})$ depend on some unknown parameters, say $\boldsymbol{\phi}$. Furthermore, we can let those parameters depend on location and denote them as $\boldsymbol{\phi}(\boldsymbol{s})$. In this case it is natural to assume that the parameters correspond to random fields defined on $\mathcal{S}$. The term conditionally linear spatial processes refers to the fact that, conditional on $\boldsymbol{\phi}(\boldsymbol{s})$, Equation (2) corresponds to a linear combination of basis functions.

CLPs provide relatively parsimonious yet flexible spatial models and can be naturally extended to spatio-temporal settings. Lemos and Sansó (2009) considered sea surface temperature data for the North Atlantic over a span of 30 years, and estimated location-dependent means, long-term trends and time-varying cycles. The kernel of choice in this case is the (circular) Bézier kernel with range r, given as

$$b_c(\boldsymbol{s}; \phi_1) = \left(1 - \left(\frac{||\boldsymbol{s}||}{r}\right)^2\right)^{L_p + \phi_1(U_p - L_p)} \times \mathbb{1}_{||s||<r},\ \phi_1 \in (0,1)\ .$$

L_p and U_p define the pre-set lower and upper bounds of the power. This kernel has compact support, which induces a sparse structure in $\boldsymbol{B}(\boldsymbol{s})$ and allows for fast computations of the model's linear structure. Additionally, the parameter ϕ_1 controls the smoothness of the resulting random field. In fact, the convolution of the kernel with itself produces a covariance that is $\lfloor 2(L_p + \phi_1(U_p - L_p))\rfloor$ times differentiable (Brenning, 2001).

Bézier kernels avoid the need of tapering covariance matrices (Gneiting, 2002; Kaufman et al., 2008) to obtain families of Gaussian processes with varying degrees of smoothness and compactly supported correlation functions. Non-isotropic kernels can be obtained by varying spatially the shape of the support. As an example, for $d = 2$, a non-circular kernel can be written as

$$b_n(\boldsymbol{s} - \boldsymbol{s}_j^*; \boldsymbol{\phi}) = \left(1 - ||\boldsymbol{s} - \boldsymbol{s}_j^*||^2_{\boldsymbol{\Sigma}}\right)^{L_p + \phi_1(U_p - L_p)} \times \mathbb{1}_{||\boldsymbol{s} - \boldsymbol{s}_j^*||_{\boldsymbol{\Sigma}}<1}$$

where the latter three components of $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3, \phi_4)$ define a distance given as

$$||\boldsymbol{s} - \boldsymbol{s}_j^*||_{\boldsymbol{\Sigma}} \equiv \sqrt{\left(\boldsymbol{s} - \boldsymbol{s}_j^*\right)^T \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{s} - \boldsymbol{s}_j^*\right)}.$$

The ellipsoidal shape is controlled by the parameters in

$$\boldsymbol{\Sigma}^{-1} \equiv \begin{pmatrix} \Psi_1 + \Psi_2 \cos(2\pi\phi_4) & \Psi_2 \sin(2\pi\phi_4) \\ \Psi_2 \sin(2\pi\phi_4) & \Psi_1 - \Psi_2 \cos(2\pi\phi_4) \end{pmatrix}$$

defined as

$$\boldsymbol{\Psi} = \frac{1}{2}\left(\frac{1}{a^2} + \frac{1}{A^2}, \frac{1}{a^2} - \frac{1}{A^2}\right), \quad a = L_s + \phi_2(U_s - L_s), \quad A = a + \phi_3(U_s - a).$$

So, for $\phi_2, \phi_3, \phi_4 \in (0,1)$, the semi-minor and semi-major axes a and A, respectively, belong to (L_s, U_s), and $a \leq A$. L_s and U_s are the pre-set limits for the support of the kernel.

By varying $\boldsymbol{\phi}$ spatially one obtains a random field whose smoothness depends on location, and is defined by ellipsoidal kernels whose elongations and support also depend on location. The spatial variation of $\boldsymbol{\phi}$ is obtained using a normalized Bézier kernel, say b_n^*, as

$$\phi_k(\boldsymbol{s}) = \sum_{j=1}^{m} b_n^*(\boldsymbol{s} - \boldsymbol{s}_j^*; \boldsymbol{u})\rho_k(\boldsymbol{s}_j^*) \ , k = 1, \ldots, 4,$$

where $\boldsymbol{u} = (u, 1, 0, 0)$, implying spherical kernels with radii U_s and some pre-set smoothness parameter u. Normalization implies that $\phi_k(\boldsymbol{s})$ is a convex combination of $\rho_k(\boldsymbol{s}_j^*)$ and thus, if $\rho_k(\boldsymbol{s}_j^*) \in (0,1)\,\forall j$, then $\phi_k(\boldsymbol{s}) \in (0,1)$, $\forall \boldsymbol{s} \in S$, so that we have defined a bounded spatial process for each k.

In summary, the main random quantities in this analysis are the regression coefficients $\beta_i, i = 1, \ldots, q$, and the latent point processes γ_j and $\rho_{j,k}$, $j = 1, \ldots, m$, $k = 1, \ldots, 4$. The other unknown parameters may be derived from these. Measurement precision τ is usually considered known or subject to substantial prior information. Regarding the prior distributions, in the absence of information about $\boldsymbol{K}$, in our applications we use independent improper priors, $p(\gamma_j) \propto 1$. An alternative choice that will induce some spatial dependence is the use of a Markov random field, as considered in Higdon (2007). Also, each $\rho_{j,k}$, receives an independent standard uniform prior, $\rho_{j,k} \sim U(0,1)$.

2.1 Uncorrelated expansions

The well known Kahunen-Loève (KL) expansion (see, for example, Yaglom, 1986) provides a linear representation of a random field $\omega(\cdot)$, with covariance function $v(\cdot,\cdot)$. It is given by $\omega(\boldsymbol{s}) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\xi_j(\boldsymbol{s})\alpha_j$, for a set of orthogonal functions $\xi_1, \xi_2, \ldots$, uncorrelated random variables α_j, and non-negative λ_j that satisfy the integral equation $\int v(\boldsymbol{s}, \boldsymbol{s}')\xi_j(\boldsymbol{s}')d\boldsymbol{s}' = \lambda_j\xi_j(\boldsymbol{s}')$.

The CLP defined as in Equation (2) does not provide a full KL expansion of the random field, but it does provide a representation based on independent coefficients that correspond to the main modes of spatial variability. Take the spectral decomposition of $\boldsymbol{K}$, say $\boldsymbol{K} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T$, where $\boldsymbol{P}$ is orthogonal and $\boldsymbol{\Lambda}$ is diagonal with elements λ_j. Let $\boldsymbol{\alpha} \sim N_m(0, \boldsymbol{\Lambda})$, then we have that $\boldsymbol{\gamma} = \boldsymbol{P}\boldsymbol{\alpha}$. Thus,

$$\omega(\boldsymbol{s}) = \boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{\gamma} = \left(\boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{P}\right)\boldsymbol{\alpha} = \boldsymbol{\psi}(\boldsymbol{s})^T\boldsymbol{\alpha} = \sum_{j=1}^{m} \psi_j(\boldsymbol{s})\alpha_j \ . \tag{3}$$

As Λ is a diagonal matrix, we have that the coefficients of the representation in (3) are independent. Furthermore, if we order the elements of Λ to have $\lambda_1 > \lambda_2 > \ldots \lambda_m$, then $\psi_1(\boldsymbol{s})$ corresponds to the factor with the largest variability, $\psi_2(\boldsymbol{s})$ to the factor with the second largest, and so on. This provides a spatial factor analysis of the original random field.

2.2 Extensions to spatio-temporal models

CLPs can be easily extended to spatio-temporal models by making the vector of coefficients dependent on time, say $\boldsymbol{\gamma}_t$. So we write

$$y_t(\boldsymbol{s}) = \boldsymbol{x}_t^T \boldsymbol{\beta} + \boldsymbol{B}^T \boldsymbol{\gamma}_t + \varepsilon_t(\boldsymbol{s}) \tag{4}$$

for the time series of observations at location $\boldsymbol{s}$. The evolution of $\boldsymbol{\gamma}_t$ can be defined in a Markovian fashion by a linear equation like

$$\boldsymbol{\gamma}_t = \boldsymbol{G}_t \boldsymbol{\gamma}_{t-1} + \boldsymbol{v}_t \tag{5}$$

for some matrices $\boldsymbol{G}_t$ and a serially uncorrelated sequence of random vectors $\boldsymbol{v}_t$. Stroud et al. (2001) use dynamic linear models (West and Harrison, 1997) to perform inference for the model defined by Equations (4) and (5). Cressie et al. (2010) develop the so called fixed rank filtering using similar models. Sansó et al. (2008) study a number of families of spatio-temporal model that fit within the framework of CLPs.

To gain more insight on the properties of some simple CLPs, consider the spatio-temporal extension of the model in Equation (2), $\omega_t(\boldsymbol{s}) = \sum_{j=1}^m B_j(\boldsymbol{s})\gamma_{t,j} = \boldsymbol{B}(\boldsymbol{s})^T \boldsymbol{\gamma}_t$, with evolution equation given by $\boldsymbol{\gamma}_t = \boldsymbol{\gamma}_{t-1} + \boldsymbol{v}_t$. Collapsing the two expressions we have

$$\omega_t(\boldsymbol{s}) = \sum_{j=1}^m B_j(\boldsymbol{s})\gamma_{t-1,j} + \sum_{j=1}^m B_j(\boldsymbol{s}) v_{t,j} \ . \tag{6}$$

In words, the process of interest at time t and location $\boldsymbol{s}$ is a weigthed average of the values of a latent process at the previous time step, for locations that are in a neighborhood of $\boldsymbol{s}$, plus a spatially correlated, but serially uncorrelated noise. Kernels with spatially varying shapes and support, like the ones proposed in Section 2, capture the interactions between time and space. They provide parsimonious descriptions of potentially complicated dynamics. Xu et al. (2005) model the spatio-temporal dynamics of a process of interest by a discrete time, continuous space, linear integro-difference equation (IDE). An IDE is given by

$$\omega_t(\boldsymbol{s}) = \int_S B_{\boldsymbol{s}}(s; \boldsymbol{\phi}) \omega_{t-1}(\boldsymbol{s}) d\boldsymbol{s} \ .$$

$B_{\boldsymbol{s}}(s; \boldsymbol{\phi})$ is the "redistribution" kernel, responsible for describing the interactions between time and space in the evolution of the process. Clearly there are similarities between an IDE and the model in Equation (6). A word of caution is due, though, as only in the case of kernels with a very tight support can $\boldsymbol{\gamma}_t$ be interpreted as the restriction of $\omega_t(\boldsymbol{s})$ to a grid.

2.3 Model fitting

Model fitting can use standard Markov chain Monte Carlo (MCMC) methods, which may be improved if we take advantage of conditional linearity and sparseness. More specifically,

when the parameters of the basis functions are fixed, CLPs become linear models with given design matrices ($\boldsymbol{X}$). Thus, the full conditional distribution of $\boldsymbol{\gamma}$ results in

$$(\boldsymbol{\gamma}|\boldsymbol{\beta}, \boldsymbol{\Phi}, \boldsymbol{Y}) \sim N\left(\tau\left(\boldsymbol{K}^{-1} + \tau(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}, \left(\boldsymbol{K}^{-1} + \tau(\boldsymbol{X}^T\boldsymbol{X})^{-1}\right)^{-1}\right)$$

with $\boldsymbol{\Phi} = \boldsymbol{\phi}(\boldsymbol{s}_1), \ldots, \boldsymbol{\phi}(\boldsymbol{s}_n)$, $\boldsymbol{Y} = y(\boldsymbol{s}_1), \ldots, y(\boldsymbol{s}_n)$ and $X(i,j) = B_j(\boldsymbol{s}_i)$.

Due to the compact support of the kernels in our proposed model, most of the entries of $\boldsymbol{X}$ are zero. This can be used to speed up computations of the mean and covariance matrix of the above distribution. Of particular interest is the case where $\boldsymbol{\gamma}$ receives an improper prior. Here, the full conditional variance is a symmetric band matrix, under suitable indexing of grid points. The bandwidth is controlled prior to model fitting, through the ratio between the upper support for the kernel (U_s) and the grid spacing. For example, in a one-dimensional problem the relationship between bandwidth (w) and ratio (z) is $w = 4z - 1$. Therefore, it is possible to adapt the MCMC algorithm beforehand, and even to sample (conditionally) independent γs with parallel computing processes. The resulting reduction in computation time is important for large m and especially relevant for spatio-temporal problems, where we have $\gamma_{t,j}$ instead of γ_j.

3 Examples

For this example, we use a classical data set containing 148 sea scallop (*Placopecten magellanicus*) abundance measurements, based on a 1990 survey cruise in the Atlantic continental shelf off Long Island, New York, USA (Ecker and Heltshe, 1994). A recent analysis of this data set is provided by Banerjee et al. (2004).

As shown in the upper left panel in Figure 1, scallop abundance varies widely across the domain. Greater density occurs in the central region and along the southern border, which has SW-NE orientation. To interpolate this data set, using the methods proposed in this paper, we begin by setting up a grid, with 0.5° resolution, which envelops the data (Figure 1, upper left panel). Then, we define the bounds for the convolution kernel: $L_p = 2, U_p = 4$, so that the resulting surface will be fairly smooth. We set $L_s = 0.5°, U_s = 2°$, so that the support is allowed to vary from one to four units of the grid. Finally we set $u = 2$, to obtain a reasonably smooth surface for the processes of kernel parameters. For a matter of simplicity, we do not consider predictor variables ($q = 0$). We specify an improper flat prior for $\boldsymbol{\gamma}$ and uniform $(0, 1)$ priors for each of the $\boldsymbol{\rho}_k$. We fit our model using purposely built Fortran code.

For comparison purposes, we also fit a predictive Gaussian process model, using the function `spLM` from the `R` package `spBayes` (Banerjee et al., 2008). We opt for the Matèrn covariance function, and let the range parameter have a Uniform$(0.1, 3)$ prior, while the smoothness parameter ν receives a Uniform$(0.5, 3.5)$ prior. This configuration yields sufficiently good mixing for all model parameters. For both models we obtain a sample of size 1,000 of $\boldsymbol{\gamma}$, after 1:10 thinning and a burn-in stage of 1,000 iterations.

The upper right panel in Figure 1 presents the interpolated surface that results from fitting the model with the Bézier kernel. Two regions with peak abundance are identified in

the central part of the domain. Steep slopes are found in the south, as well as between the peaks. The lower right panel displays the result using the `spLM` function. Here, there is a fairly leveled SW-NE ridge, which seems to misfit the data in several locations. Gradients are more spatially homogeneous, and predicted peak abundance falls short of the observed values. To investigate if lack of fit could be due to grid coarseness, we repeat the `spLM` procedure, but this time with twice the grid resolution (0.25°). The outcome, shown in the lower right panel of Figure 1, now has roughly the same features as that of the Bézier kernel method.

The conclusion we take from this example is that anisotropic features present in spatial data can be captured using either a coarse grid and Bézier kernels, or a fine enough grid and circular kernels. Each has different costs: the former has several kernel parameters, the later requires more point processes.

An advantage of the Bézier kernel method is that we may study how anisotropy and differentiability depend on location, by looking at the mean posterior shape of the convolution kernels (Figure 2). Along the southern border of the domain, kernels are strongly elliptical, with the major axis being aligned with the SW-NE direction. Kernels are also elliptical in the center of the domain, but now oriented along the NW-SE direction. In northern regions, kernels are almost spherical. The spatial variability in posterior mean kernel eccentricity ($\epsilon = \sqrt{1-(a/A)^2}$) confirms that kernels are elliptical in the southern, central part of the domain, and that they radiate into more spherical kernels from there. Finally, there is a small SW-NE gradient in the posterior mean of the smoothness parameter ϕ_1, which indicates a change in differentiability.

Figure 3 depicts the spectral decomposition of the covariance matrix $\boldsymbol{K}$, as estimated from MCMC samples of $\boldsymbol{\gamma}$. In the Bézier kernel model, the sorted eigenvalues decay exponentially; in the predictive process approach, using either the 0.5° or the 0.25° grid, the largest eigenvalue stands out from the others, which decay slowly, in exponential fashion.

From Equation (3), we obtain m components of spatial variability, $\{\psi_j(s)\}_{j=1}^m$, and display the first two (Figure 3, mid panels). Clearly, the two methods diverge in decomposing data variability.

The representation of the cumulative sum of the ten leading components of spatial variability also reveals different patterns. In the predictive process approach, the overall SW-NE orientation of contour lines is well captured; on the other hand, gradients are weak and the maxima are misplaced. In the Bézier kernel model, a single strong mode appears at 73°W; however, the secondary peak at 72.6°W is not depicted, and abundance in the NW corner is overestimated. An interesting and expectable feature of these two plots (given the distribution of the eigenvalues) is that the former resembles $\psi_1(s)$, while the latter does not.

4 Discussion and Conclusions

We have presented a general framework for modeling spatial process that encompasses several popular methods for dimension reduction. Representation of a spatial process as a linear combination of a reduced number of basis functions has strong computational advantages.

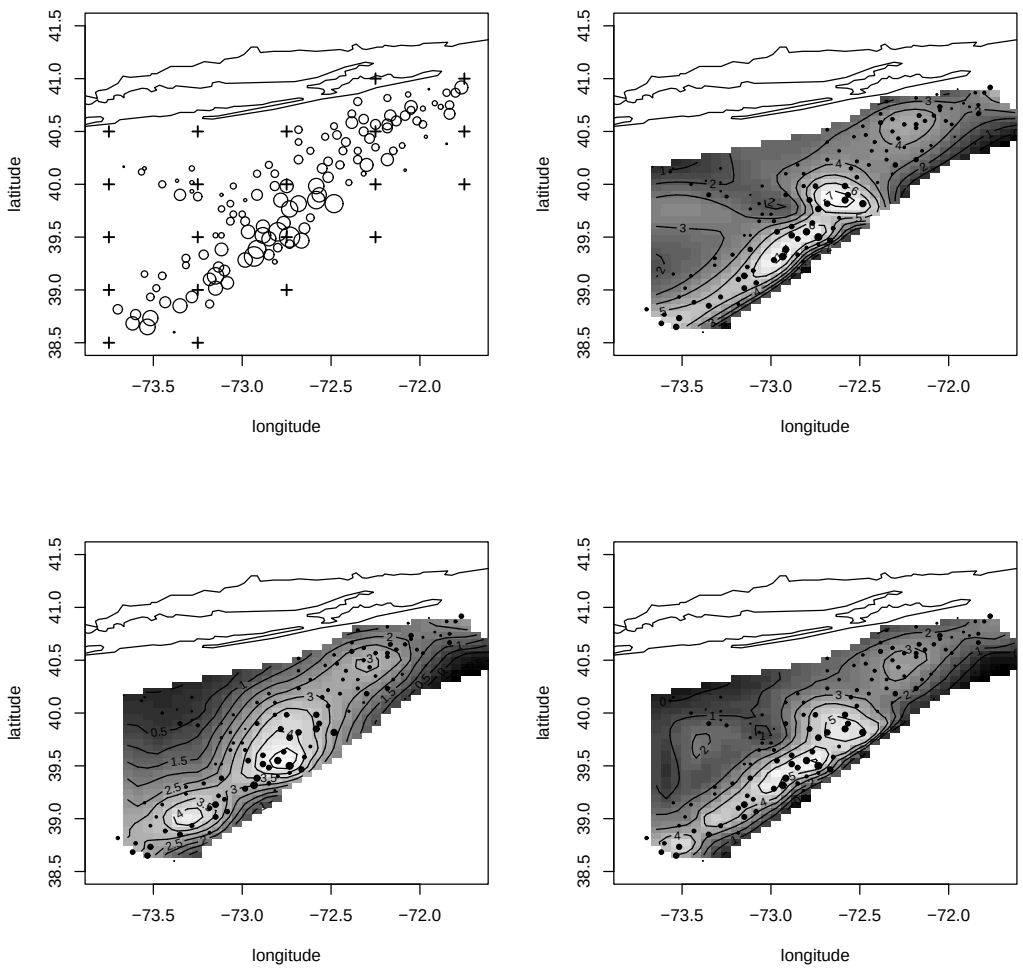

Figure 1: Fitting spatial models to the scallops data set. Upper left panel: original data and 0.5° convolution grid. Upper right: posterior mean of the convolution surface obtained with the Bézier kernel and the 0.5° grid. Lower panels: posterior means obtained with a predictive Gaussian process model using a 0.5° grid (left) and a 0.25° grid. Bubble and dot sizes are proportional to scallop log-abundance.

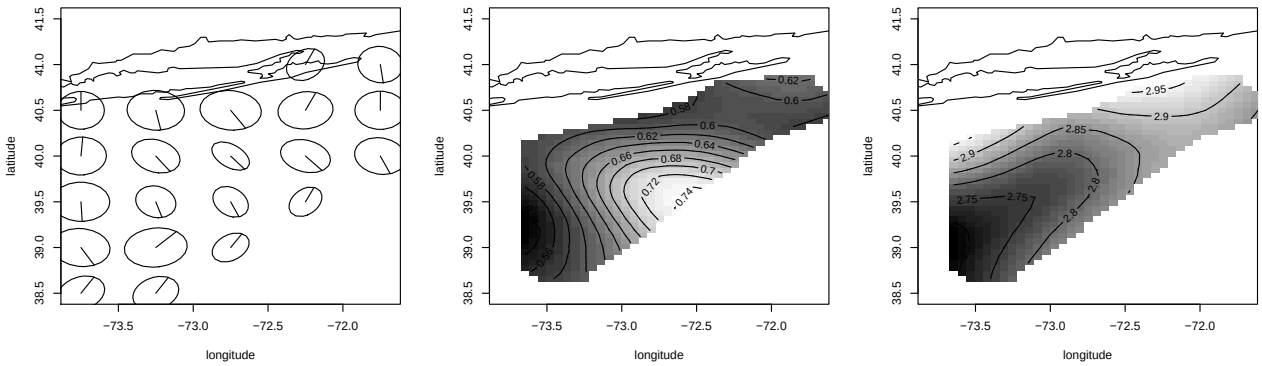

Figure 2: Spatially variable anisotropy and differentiability in scallops abundance. Left panel: posterior mean of the 0.995 contour and semi-major axis of kernels centered at the grid points. Central panel: posterior mean kernel eccentricity. Right panel: posterior mean smoothness parameter ϕ_1.

While providing enough flexibility to accommodate spatially varying features of the process, it is based on parsimonious inference. Additionally, it can be used to decompose the spatial variability in components with independent coefficients. Also, it can be generalized to spatio-temporal models that, in spite of their simplicity, are sophisticated enough to capture realistic space-time dynamics.

Regarding the comparison between predictive processes and DPCs, we observe that predictive processes require only the choice of a covariance function, while DPCs require the choice of a convolution and of a covariance matrix for the latent process. So DPCs are by construction more flexible, but require more inputs. Another property of predictive processes is that, by definition, they interpolate the gridded process $\boldsymbol{\omega}^*$. So, for example, if $\omega(\boldsymbol{s})$ is used to model scallop abundance, then γ_j are scallop abundances, restricted to a grid. Such correspondence can be used to elicit priors and build probabilistic structures for γ_j based on the biological and ecological traits of scallops. A similar property is only valid for DCPs that have very tight support, like the ones considered in Lemos and Sansó (2009). Unfortunately for predictive processes, if $\text{var}(\gamma_j) = \boldsymbol{K}_{ii} = \sigma^2$, then $\text{var}(\omega(\boldsymbol{s})) = \text{var}(\boldsymbol{B}(\boldsymbol{s})^T\boldsymbol{\gamma}) \leq \sigma^2$ implying that maximum variance is achieved at the grid points. In other words, intra-grid points are subject to structural underestimation. Finley et al. (2009) propose a correction to compensate for this, based on adding independent random shocks. By allowing for kernels with support wider than the grid spacing, the DCPs proposed in this paper avoid the type of structural underestimation typical of predictive processes.

A frequent criticism of the methods proposed in this paper is that they are based on an arbitrarily defined grid. Lemos and Sansó (2009) considered the problem of estimating the grid size from a model comparison perspective. That is, for a set possible grid sizes, one can fit the model and then use some model selection criteria to determine the optimum grid size. A more comprehensive approach would be to include the grid size as one of the model parameters. This has the drawback of requiring inferential methods that can handle

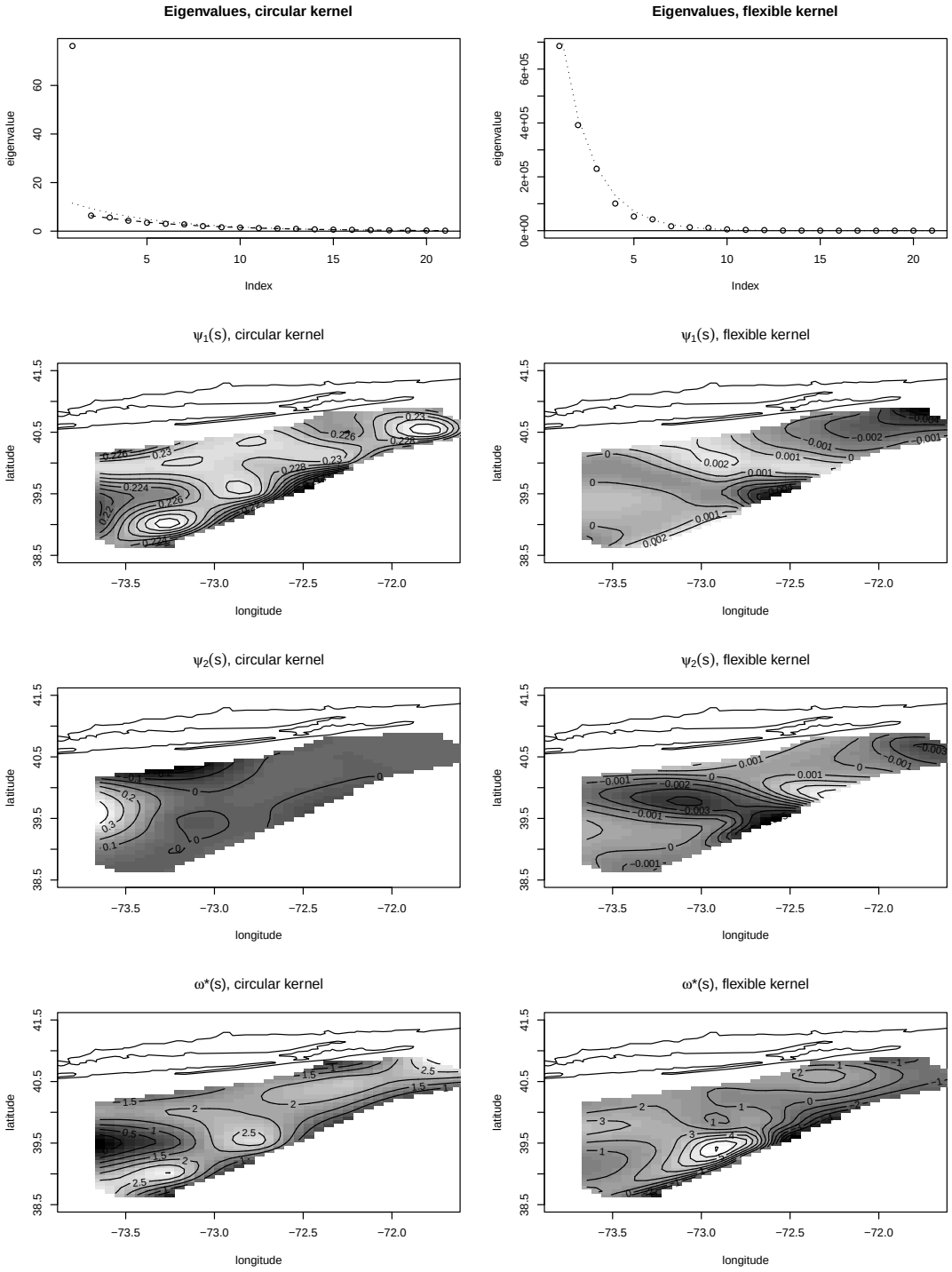

Figure 3: Spatial factor analysis. Top row of panels: sorted eigenvalues of the covariance matrix $\boldsymbol{K}$ for the Predictive Process approach (PP; left) and the Discrete Process Convolution approach (DPC; right), using the 0.5° grid. Stippled and dashed lines denote exponential fits, with the latter not using the leading eigenvalue. Mid rows: first two components of spatial variability, provided by $\psi_j(s)$. Bottom row: sum of the first ten components, $\omega^*(s) = \sum_{j=1}^{10} \psi_j(s)\alpha_j$.

parameter spaces of variable dimension, as the number of components in $\boldsymbol{\gamma}$ depends on the grid size. Of course estimation of the grid size is tied to the use of a regular grid. This can be a strong limitation, as the grid density should be a function of the spatial variability of the process of interest. We are aware of ongoing efforts to develop models that place the grid according to a spatial point process. On the other hand, the approach proposed in this paper puts the emphasis on the kernel shape and support, which are strongly associated with the location and density of grid points. By being able to vary the kernel the model can compensate for the grid regularity and coarseness, as shown in Figure 1. Naturally, any additional sophistication has a computational prize, which is important to bear in mind when dealing with large datasets. Our experience with dense satellite data shows that a high resolution grid and spherical kernels with fixed compact support provide reasonable results and this may be ultimately the most cost effective approach. Still, information on anisotropy and smoothness (as in Figure 2) may prove invaluable for the understanding of underlying dynamic processes.

Acknowledgments

Bruno Sansó was partially supported by National Science Foundation grant DMS-0906765.

References

Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical Modeling and Analysis of Spatial Data*. Chapman and Hall, New York.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, B*, **70**:825–848.

Brenning, A. (2001). Geostatistics without stationarity assumptions within geographical information systems alexander brenning. *Freiberg Online Geoscience*, 6.

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.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large data sets. *Journal of the Royal Statistical Society, B*, **70**:209–226.

Cressie, N., Shi, T., and Kang, E. (2010). Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics*, 19:724–745.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, Revised Edition*. John Wiley and Sons, New York.

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

Ecker, M. and Heltshe, J. (1994). Geostatistical estimates of scallop abundance. In Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest L., and Greenhouse, J., editors, *Case studies in biometry*, pages 125–144. John Wiley & Sons, Inc.

Finley, A., Banerjee, S., Waldman, P., and Ericsson, T. (2009). Hierarchical spatial modeling of additive and dominance genetic variance for large spatial trial datasets. *Biometrics*, 65:441–451. DOI: 10.1111/j.1541-0420.2008.01115.x.

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. *Journal of the American Statistical Association*, 102:321–331.

Gelfand, A., Diggle, P., Fuentes, M., and Guttorp, P., editors (2010). *Handbook of Spatial Statistics*. Chapman and Hall, Boca Raton, USA.

Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83:493–508.

Gramacy, R. and Lee, H. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103:1119–1130.

Higdon, D. (2007). A primer on space-time modeling from a Bayesian perspective. In Finkenstädt, B., Held, L., and Isham, V., editors, *Statistical Methods for Spatio-Temporal Systems*, chapter 6, pages 217–279. Chapman and Hall.

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

Kaufman, C., Schervish, M., and Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, 103:1545–1555.

Kim, H.-M., Mallick, B. K., and Holmes, C. C. (2005). Analyzing non-stationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association*, 100:653–668.

Lemos, R. and Sansó, B. (2009). A spatio-temporal model for mean, anomaly and trend fields of north Atlantic sea surface temperature (with discussion). *Journal of American Statistical Association*, 104:5–25. DOI 10.1198/jasa.2009.0018.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58:1246–1266.

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

Paciorek, C. (2007). Bayesian smoothing with Gaussian processes using Fourier basis functions in the spectralGP 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.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall, London, UK.

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

Sansó, B., Schmidt, A., and Nobre, A. (2008). Spatio-temporal models based on discrete convolutions. *Canadian Journal of Statistics*, 36:239–258.

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.

Stroud, J., Müller, P., and Sansó, B. (2001). Dynamic models for spatio-temporal data. *Journal of the Royal Statistical Society, B*, **63**:673–689.

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.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41:434–449.

Xu, K., Wikle, C., and Fox, N. (2005). A kernel-based spatio-temporal dynamical model for nowcasting radar precipitation. *Journal of the American Statistical Association*, 100:1133–1144.

Yaglom, A. (1986). *Correlation Theory of Stationary and Related Random Functions I: Basic Results*. Springer Series in Statistics. Springer-Verlag, New York.