ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Journal of the Royal Statistical Society
Applied Statistics
Series C

Appl. Statist. (2017)

Segmentation of sea current fields by cylindrical hidden Markov models: a composite likelihood approach

Monia Ranalli and Francesco Lagona,

University of Roma Tre, Italy

Marco Picone

Istituto Superiore per la Protezione e la Ricerca Ambientale, Rome, Italy

and Enrico Zambianchi

'Parthenope' University of Naples and Consorzio Nazionale Interuniversitario per le Scienze del Mare, Rome, Italy

[Received December 2016. Revised July 2017]

Summary. Motivated by segmentation issues in studies of sea current circulation, we describe a hidden Markov random field for the analysis of spatial cylindrical data, i.e. bivariate spatial series of angles and intensities. The model is based on a mixture of cylindrical densities, whose parameters vary across space according to a latent Markov field. It enables segmentation of the data within a finite number of latent classes that represent the conditional distributions of the data under specific environmental conditions, simultaneously accounting for unobserved heterogeneity and spatial auto-correlation. Further, it parsimoniously accommodates specific features of environmental cylindrical data, such as circular–linear correlation, multimodality and skewness. Because of the numerical intractability of the likelihood function, estimation of the parameters is based on composite likelihood methods and essentially reduces to a computationally efficient expectation–maximization algorithm that iteratively alternates the maximization of a weighted composite likelihood function with weights updating. These methods are tested on simulations and exploited to segment the sea surface of the Gulf of Naples by means of meaningful circulation regimes.

Keywords: Abe–Ley density; Composite likelihood; EM algorithm; Gulf of Naples; Hidden Markov random field; Marine currents

1. Introduction

Detailed knowledge of coastal currents is crucial for valid integrated coastal zone management, as well as for applications in case of pollution events and in search-and-rescue operations at sea. Among the various available ocean observing technologies, high frequency radars (HFRs) have unique characteristics that make them play a key role in coastal observatories (Bellomo *et al.*, 2015; Falco *et al.*, 2016). As land-based remote sensing instruments, these state of the art devices provide synoptic maps of surface current fields at high spatial resolution. For this reason, following the example of the US ocean observing system (`https://ioos.noaa.gov/project/hf-radar/`), HFR networks are becoming an en-

Address for correspondence: Monia Ranalli, Department of Political Sciences, University of Rome Tre, Via Gabriello Chiabrera 199, Rome 000145, Italy.
E-mail: monia.ranalli@uniroma3.it

 0035–9254/17/67000

vironmental priority everywhere and their number is quickly increasing (Rubio *et al.* (2017), and references therein).

HFR data can be conveniently described as bivariate spatial series of angles and intensities that respectively indicate the direction and the speed of the current at every point of a spatial lattice, which partitions the area of interest according to a specific resolution. Spatial series with a mixed circular–linear support are often referred to as *cylindrical* spatial series (Abe and Ley, 2016), because the pair of an angle and an intensity can be represented as a point on a cylinder. The statistical analysis of cylindrical spatial series is complicated by the unconventional topology of the cylinder and by the difficulties in modelling the cross-correlations between angular and linear measurements across space. Additional complications arise from the skewness and the multimodality of the marginal distributions of the data. Indeed, intensities are typically negatively skewed and directional data are rarely symmetric; multimodality may arise as well as the data often being observed under heterogeneous, space varying conditions.

HFR data are not the only example of cylindrical spatial series. In environmental studies, additional examples include spatial series of either wind speed and direction (Modlin *et al.*, 2012) or height and direction of ocean waves (Wang and Gelfand, 2014; Wang *et al.*, 2015). Spatial cylindrical data arise also in medical imaging (Abraham *et al.*, 2013; Klauenberg and Lagona, 2007) and in ecological studies of telemetry data of animal movement (Hanks *et al.*, 2015). Despite their popularity, though, specific methods for the analysis of spatial cylindrical data have been relatively unexplored. Proposals in this context are limited to Bayesian hierarchical models, which require specific assumptions on the prior distribution of the parameters of interest and *ad hoc* Markov chain Monte Carlo algorithms to compute the posterior distribution of the parameters. For example, in a study of hurricane winds, Modlin *et al.* (2012) proposed a Bayesian hierarchical model, specified by combining a circular conditional auto-regressive model for hurricane wind direction, based on the wrapped circular distribution, and a spatial Gaussian auto-regressive model for the logarithm of hurricane wind speed. In a study of the output of a deterministic model of sea motion, Wang *et al.* (2015) integrated a Gaussian conditional distribution of wave height given wave direction and a Bayesian geostatistical model for wave direction, by exploiting a projected normal spatial process. These approaches were motivated by specific issues that respectively arise in the analysis of hurricane winds, characterized by inward spiralling winds that rotate about a zone of low pressure, and in the analysis of smooth deterministic circulation outputs of numerical models, computed across a broad area of interest.

Sea motion in coastal areas provides, however, a different setting. Coastal currents are shaped and constrained by the orography of the site. As a result, coastal circulation is much more irregular than hurricane-type patterns and it is inaccurately represented by traditional numerical models, which do not incorporate orographic information and work well in the open sea. The development of a physical model that well represents sea motion in coastal areas can be a formidable task if the orography of the site is irregular. A more practical approach relies on decomposing an observed circulation pattern into a small number of local regimes whose interpretation is easier than the global pattern. Previous work in this direction has been done by exploiting methods that are based on empirical orthogonal function analysis (Buffoni *et al.*, 1997; Lillibridge and Mariano, 2013), which extracts the principal components of the circulation data. This approach does not account for the cylindrical nature of the data and provides principal components that are often difficult to interpret.

In this paper we propose an alternative approach that accounts for the cylindrical nature of the data and enables interpretable segmentation. Specifically, we assume that the joint distribution of the data is well approximated by a mixture of a small number of cylindrical densities, with space

varying parameters that are driven by a latent spatial process. In particular, we approximate the data distribution with a mixture of Abe–Ley cylindrical densities, whose parameters vary across space according to a Potts model. In its simplest version, the Potts model is a one-parameter Markov random field (MRF), i.e. a multinomial process in discrete space which fulfils a spatial Markovian property (Guyon, 1995). It segments an area of interest according to an interaction parameter that captures the correlation between adjacent observations and controls the smoothness of the segmentation. The Abe–Ley density (Abe and Ley, 2016) is a five-parameter bivariate density on the cylinder. It parsimoniously accommodates correlated and skew cylindrical data by means of parameters that can be easily interpreted in terms of traditional concepts such as location, shape, scale, skewness and concentration. A mixture of Abe–Ley densities therefore provides a flexible distributional extension to allow for multimodal cylindrical data. By assuming that the mixture parameters vary according to the segmentation provided by a Potts MRF, we obtain a cylindrical hidden MRF (HMRF) that provides a further extension to capture unobserved spatial heterogeneity and, simultaneously, to allow for spatial correlation.

HMRFs are popular models in spatial statistics, since the seminal papers by Besag (Besag, 1975, 1977). They can be seen as a spatial extension of the hidden Markov models that are exploited in time series analysis. Hidden Markov models have recently been proposed for the analysis of cylindrical time series (Lagona *et al.*, 2015). Furthermore, a cylindrical HMRF has been exploited in a study of HFR data in the northern Adriatic sea (Lagona and Picone, 2016). However, the use of cylindrical HMRFs is limited by the intractability of the likelihood function. Lagona and Picone (2016) adapted a mean field approximation for Gaussian HMRFs (Celeux *et al.*, 2003) to the cylindrical setting and developed a computationally intensive expectation–maximization (EM) algorithm. Unfortunately, however, the method is numerically unstable and little is known about the distributional properties of the estimators.

In this paper, we take a composite likelihood (CL) approach to estimate a cylindrical HMRF. CL methods have been proved to be a good solution to balance statistical and computational efficiency in many fields (Varin *et al.*, 2011; Lindsay, 1988). The underlying naive idea dates back to the pseudolikelihoods of Besag (1974) and the partial likelihood of Cox (1975). The CLs are typically constructed by adding individual component likelihoods, each of which corresponds to a marginal or conditional event (Lindsay, 1988; Varin *et al.*, 2011). This strategy, on one hand, provides feasible and fast estimation methods. On the other hand, some dependence between observations is lost, resulting in a loss of statistical efficiency, but consistency of the CL estimators still holds under regularity conditions (Molenberghs and Verbeke, 2005). Under these conditions, furthermore, CL estimators are asymptotically normal with covariance matrix given by the inverse of a sandwich matrix, known as Godambe information (Godambe, 1960) rather than the usual Fisher information matrix for maximum likelihood estimators. In addition, CL-based methods have some further desirable properties: they are robust estimation methods (Xu and Reid, 2011) and their estimators are fully efficient and identical to the maximum likelihood estimators in exponential families under a certain closure property (Mardia *et al.*, 2009). Recent applications of CL methods to spatial and space–time data include the work by Okabayashi *et al.* (2011) and Eidsvik *et al.* (2014) respectively.

The rest of the paper is organized as follows. Section 2 briefly describes the two spatial series that we use for illustrating the methods proposed. Section 3 summarizes the structure of the proposed cylindrical HMRF and Section 4 illustrates the CL methods that we suggest for estimation. In Section 5 we discuss the main output of the simulations that have been carried out to examine the distributional properties of the estimation method; further results are included in the on-line appendix. Section 6 is devoted to the results that have been obtained by the proposed methods on real data. Section 7 finally summarizes relevant points of discussion.

2. Marine currents in the Gulf of Naples

The Gulf of Naples is a semienclosed marginal basin of the central Tyrrhenian Sea (Mediterranean Sea; see Cianelli *et al.* (2012) for an overview). It is a coastal area characterized by striking environmental contrasts: one of the most intensely urbanized coastlines in the whole Mediterranean, with massive industrial settlements, the very polluted Sarno river mouth, a number of distributed sewage outlets, coexisting with the extremely scenic coastal landscapes of the Sorrento Peninsula, of the Islands of Capri, Procida and Ischia and with unique underwater archaeological treasures (e.g. Baiae and Gaiola). For this reason, the Gulf of Naples has been subject to intense monitoring of its meteorological and oceanographic conditions and, more generally, of the status of its marine ecosystem. In particular, starting in 2004 an HFR system has been installed along its coastline, consisting first of two, and from 2008 of three, transceiving antennas operating at 25 MHz, providing hourly data of the surface current field at 1-km^2 horizontal resolution. Such a system has shed light on very rich, multiple-scale surface dynamics and on the mechanisms driving water renewal of individual subbasins of the gulf (Menna *et al.*, 2007; Uttieri *et al.*, 2011; Cianelli *et al.*, 2015). Moreover, these data have been exploited in numerical models to enhance their predictive skills through state of the art assimilation schemes (Iermano *et al.*, 2016).

The functioning principle of HFRs is based on resonant backscatter, resulting from coherent reflection of a transmitted electromagnetic wave by ocean surface waves whose wavelength is half of the transmitted electromagnetic wavelength. This effect, modulated by the dual Doppler shift that is associated with the wave motion and with the presence of an underlying velocity field, enables measurement of the latter. As a result, every station can provide only the radial component of the surface currents with respect to the antenna location. Two, at least, or even better more stations (to ensure better statistics, to minimize gaps due to physical obstacles or to electromagnetic disturbances, to lower geometric dilution of precision) are needed to combine the radial information to obtain a current vector field.

A vector map (or field) decomposes the current's field into the u- and v-components (Cartesian representation) of the sea surface at each observation point of a spatial lattice, where u corresponds to the west–east and v to the south–north current component. Joint modelling of u and v is, however, typically complicated by cross-correlations that vary dramatically in different parts of the spatial domain (Reich and Fuentes, 2007). We therefore model sea current fields by using polar co-ordinates. Specifically, the observed current field is represented as a cylindrical spatial series, obtained by computing for each observation site the speed $y = \sqrt{(u^2 + v^2)} \in (0, \infty)$ of the current (metres per second) and the direction $x = \tan 2^{-1}\{(u, v)\} \in (0, 2\pi]$ of the current (radians), where $\tan 2^{-1}$ is the inverse tangent function with two arguments and x follows the geographical convention, clockwise from north (0) to east ($\pi/2$).

The analysis that is presented in this paper is based on two cylindrical spatial series, which include current speed and direction across a grid of 300 points, relating to two vector maps recorded on March 6th at 5 p.m. (field A) and on March 8th at midnight (field B).

These data are shown in Fig. 1 and represent two typical scenarios of current circulation in the gulf. Field A represents a south-eastward current induced by a large-scale northerly wind blowing over the central Tyrrhenian Sea. Such a pattern segregates surface waters in the south-easterly sector of the gulf, preventing exchanges with the Tyrrhenian interior. Field B, in contrast, represents a more spatially variable current field, associated with easterly winds and favouring water renewal in the whole southern portion of the gulf.

Fig. 1 displays both the joint distribution of speeds and directions (Figs 1(a) and 1(b)) and the resulting spatial patterns (Figs 1(c) and 1(d)) for both scenarios. The top scatter plots should

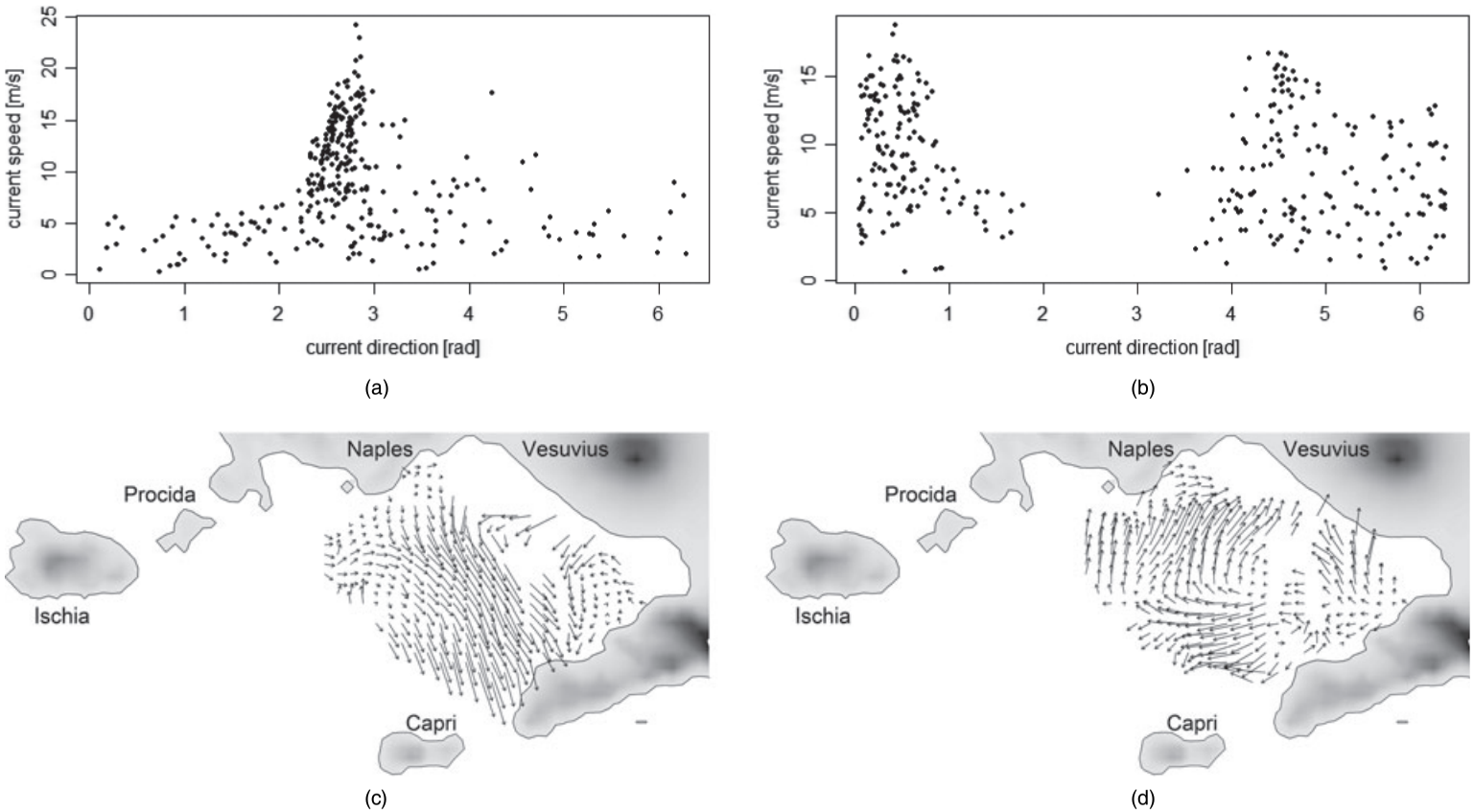


Fig. 1. Currents directions and speeds in the Gulf of Naples, recorded by HFRs at 5 p.m. on March 6th, 2009 (field A), and at midnight on March 8th, 2009 (field B): (a), (b) planar scatter plots of directions towards which the current travels (0, $\pi/2$, π and $3/2\pi$ indicate north, east, south and west respectively) and speeds; (c), (d) vector fields of the observed data, georeferenced and overlapped on a map of the gulf (the orientation of each arrow indicates the current direction at each observation site; the arrow length is proportional to the current speed at that site); (a), (c) field A; (b), (d) field B

be interpreted with care by recalling that, for simplicity, the data are plotted on the plane, by unwrapping their cylindrical domain. Scatter plot interpretation is additionally complicated by the weak correlation between directions and speeds, the skewness and the multimodality of the data. Multimodality, weak correlation and skewness are often held responsible for the inaccuracy of numerical models in coastal areas such as the Gulf of Naples and they are traditionally explained as the result of the complex orography of the area. The gulf orography shapes and constrains the circulation of water, yielding the spatial discontinuities that are clearly shown by the pictures at the bottom of Fig. 1. This motivates the development of special segmentation methods, such as the method that we propose, that can detect spatial discontinuities by means of latent classes, conditionally on which the distribution of the data takes a shape that is easier to interpret than the shape that is taken by the marginal distribution.

3. A cylindrical hidden Markov random field

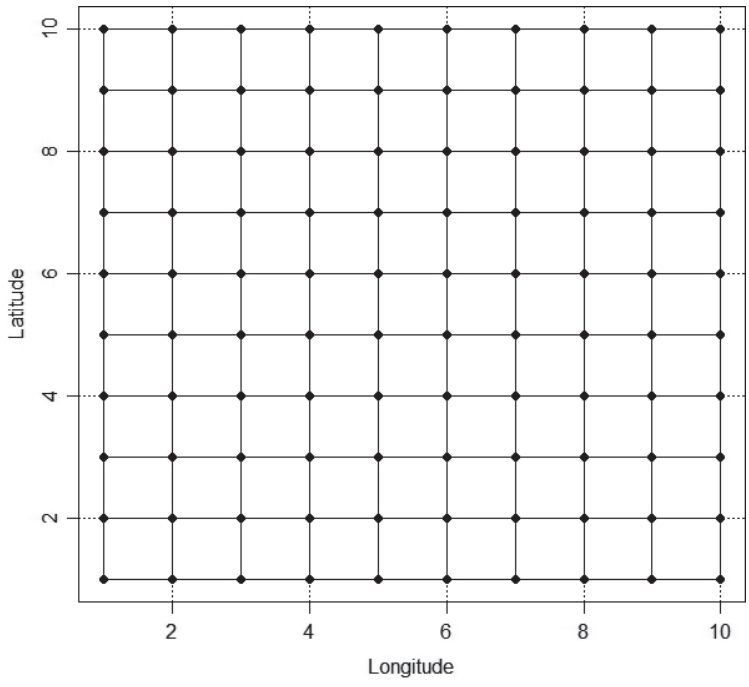
The proposed cylindrical HMRF is obtained by hierarchically integrating a parametric MRF (the Potts model) and a cylindrical density (the Abe–Ley density). We therefore first introduce the Potts and the Abe–Ley model separately. Then, we describe our proposal as a hierarchical combination of these two models.

3.1. The Potts model

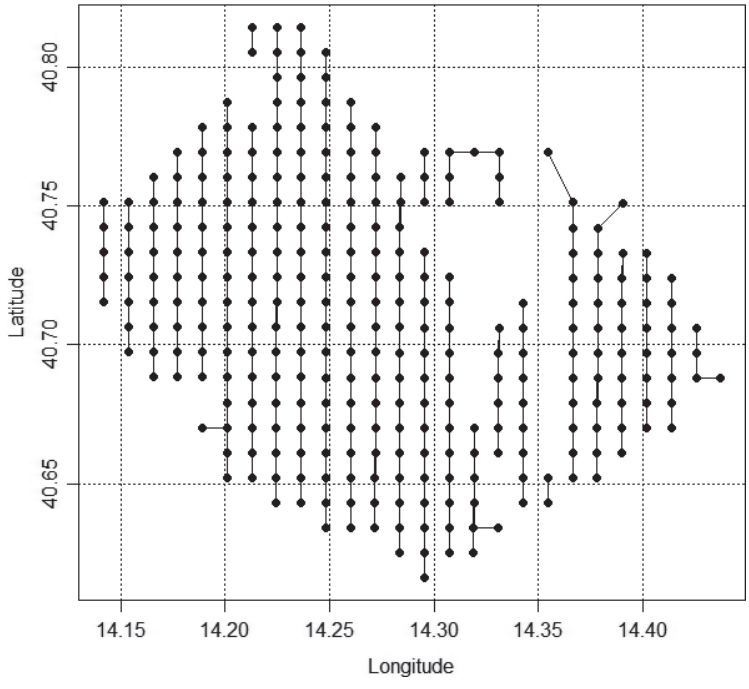
The Potts model is a multinomial process in discrete space (lattice) with K classes. Given a lattice that partitions an area of interest according to n observation sites $i = 1, \ldots, n$, a sample that is drawn from a spatial multinomial process is a segmentation of this area, obtained by associating each site with a segmentation label $k = 1, \ldots, K$. Formally, each observation site i is associated with a multinomial random variable $\boldsymbol{\xi}_i = (\xi_{i1} \ldots \xi_{ik} \ldots \xi_{iK})$ with one trial and K classes, where ξ_{ik} is a Bernoulli random variable that is equal to 1 if i is labelled by k and 0 otherwise. A specific segmentation of the area can be accordingly represented as a sample drawn from the multinomial process $\boldsymbol{\xi} = (\boldsymbol{\xi}_i, i = 1, \ldots, n)$.

In its simplest form, the Potts model is a special multinomial process which depends on a single interaction parameter ρ and on a neighbourhood structure among the observation sites. A neighbourhood structure on $S = \{1, \ldots, i, \ldots, n\}$, say $E \subset S^2$, is a symmetric and non-reflexive binary relationship, such that $(i, j) \in E \Rightarrow (j, i) \in E$ (symmetry) and $(i, i) \notin E$ (non-reflexivity). According to the structure E, each spatial index i is associated with a neighbourhood $N(i) = \{j \in S : (i, j) \in E\}$ of adjacent observation points. There are different ways to specify a spatial neighbourhood structure (Bivand *et al.* (2008), chapter 9). A viable strategy relies on defining the neighbours of each site i as the nearest neighbours, according to the Euclidean distance. Formally, we define $N(i) = \{j \in S : d(i, j) = d_i^*\}$, where $d_i^* = \min_{j \neq i} d(i, j)$, and $d(i, j)$ is the Euclidean distance between sites i and j. The structure E can be interpreted as the set of the edges of an undirected graph with vertices in S. Fig. 2 displays the graphs that are obtained by the above definition of neighbourhoods in the case of a 10×10 regular grid and in the case-study that is considered in this paper. A neighbourhood structure is often conveniently specified by means of an $n \times n$ symmetric adjacency matrix $\mathbf{C}$, whose generic entry c_{ij} is equal to 1 if $(i, j) \in E$ and 0 otherwise (diagonal entries c_{ii} are all equal to 0, because of the non-reflexivity of E). Under the Potts model considered, the joint distribution of the segmentation $\boldsymbol{\xi} = (\boldsymbol{\xi}_1 \ldots \boldsymbol{\xi}_n)$ is given by

$$p(\boldsymbol{\xi}) = \frac{\exp\left\{(\rho/2) \sum_{i=1}^{n} \sum_{j:c_{ij}=1} \boldsymbol{\xi}_i^{\mathrm{T}} \boldsymbol{\xi}_j\right\}}{W(\rho)}, \tag{1}$$



(a)



(b)

Fig. 2. Nearest neighbours adjacency structure for (a) a 10×10 regular lattice and for (b) the irregular lattice of 300 sites that covers the study area of the Gulf of Naples

where $W(\rho)$ is the normalizing constant. This model could be extended by introducing an external field and assuming that this external component depends on some covariates. Unfortunately, reliable covariates are not available in our application, and we cannot introduce any kind of prior information about classes that might be expected.

Computation of $W(\rho)$ requires special algorithms that are feasible only for small spatial lattices and whose complexity increases with the number K of latent classes (Bartolucci and Besag, 2002; Reeves and Pettitt, 2004; Friel and Rue, 2007). This complicates inference under the Potts model, even in the simple one-parameter form that we consider for segmenting coastal circulation.

When $\rho = 0$, the Potts model reduces to a sequence of independent multinomial distributions. Otherwise, the multinomial components of the process are spatially dependent. In particular, under distribution (1), the univariate conditional distributions

$$p(\xi_{ik} = 1 \mid \boldsymbol{\xi}_1 \dots \boldsymbol{\xi}_{i-1}, \boldsymbol{\xi}_{i+1}, \dots \boldsymbol{\xi}_n) = \frac{\exp\left(\rho \xi_{ik} \sum_{j:c_{ij}=1} \xi_{jk}\right)}{\sum_{k=1}^{K} \exp\left(\rho \xi_{ik} \sum_{j:c_{ij}=1} \xi_{jk}\right)} \tag{2}$$

depend only on the observations in the neighbourhood (the spatial Markov property). As a result, this model is a special one-parameter MRF. When ρ exceeds the critical value $\rho_{\text{crit}} = \log(1 + \sqrt{K})$, a phase transition occurs and the realizations of the Potts model tend towards states in which nearly all the values $\boldsymbol{\xi}_i$ are the same (Guyon, 1995).

3.2. The Abe–Ley cylindrical density

A cylindrical sample is a pair $z = (x, y)$, where $x \in [0, 2\pi)$ is a point in the circle and y is a point on the positive semiline $[0, \infty)$. Recently, a new density called an Abe–Ley density, has been proposed by Abe and Ley (2016). It is defined on the cylinder $[0, 2\pi) \times [0, \infty)$ and it takes the form

$$f(x, y) = \frac{\alpha \beta^{\alpha}}{2\pi \cosh(\kappa)} \{1 + \lambda \sin(x - \mu)\} y^{\alpha - 1} \exp[-(\beta y)^{\alpha} \{1 - \tanh(\kappa) \cos(x - \mu)\}]. \tag{3}$$

This density has some interesting features. First, the parameters can be easily interpreted: $\alpha > 0$ is a shape parameter, $\beta > 0$ is a scale parameter, $\mu \in [0, 2\pi)$ is a circular location parameter, $\kappa > 0$ is a circular concentration parameter and $\lambda \in [-1, 1]$ is a circular skewness parameter. Fig. 3 displays the shape of the Abe–Ley density considering various values of κ and λ. The parameter κ plays the role of circular concentration and a circular–linear dependence parameter. Independence is attained when $\kappa = 0$, in which case density (3) becomes the product of the linear Weibull and the circular cardioid distribution with location $\mu + \pi/2$ and concentration λ, as depicted in Figs 3(a)–3(c). Second, the normalizing constant is numerically tractable. Third, the univariate marginal and conditional distributions exist in closed form. The conditional distribution of the intensity y given the angular direction x is Weibull with shape α and direction-dependent scale β_x, where $\beta_x = \beta\{1 - \tanh(\kappa) \cos(x - \mu)\}^{1/\alpha}$. In contrast, the conditional distribution of the direction x given the intensity y is a skew von Mises distribution with location parameter μ and concentration parameter κ_y, where $\kappa_y = (\beta y)^{\alpha} \tanh(\kappa)$.

3.3. A hidden Markov random field for cylindrical data

A cylindrical spatial series can be represented as a bivariate series of angles x_i and intensities

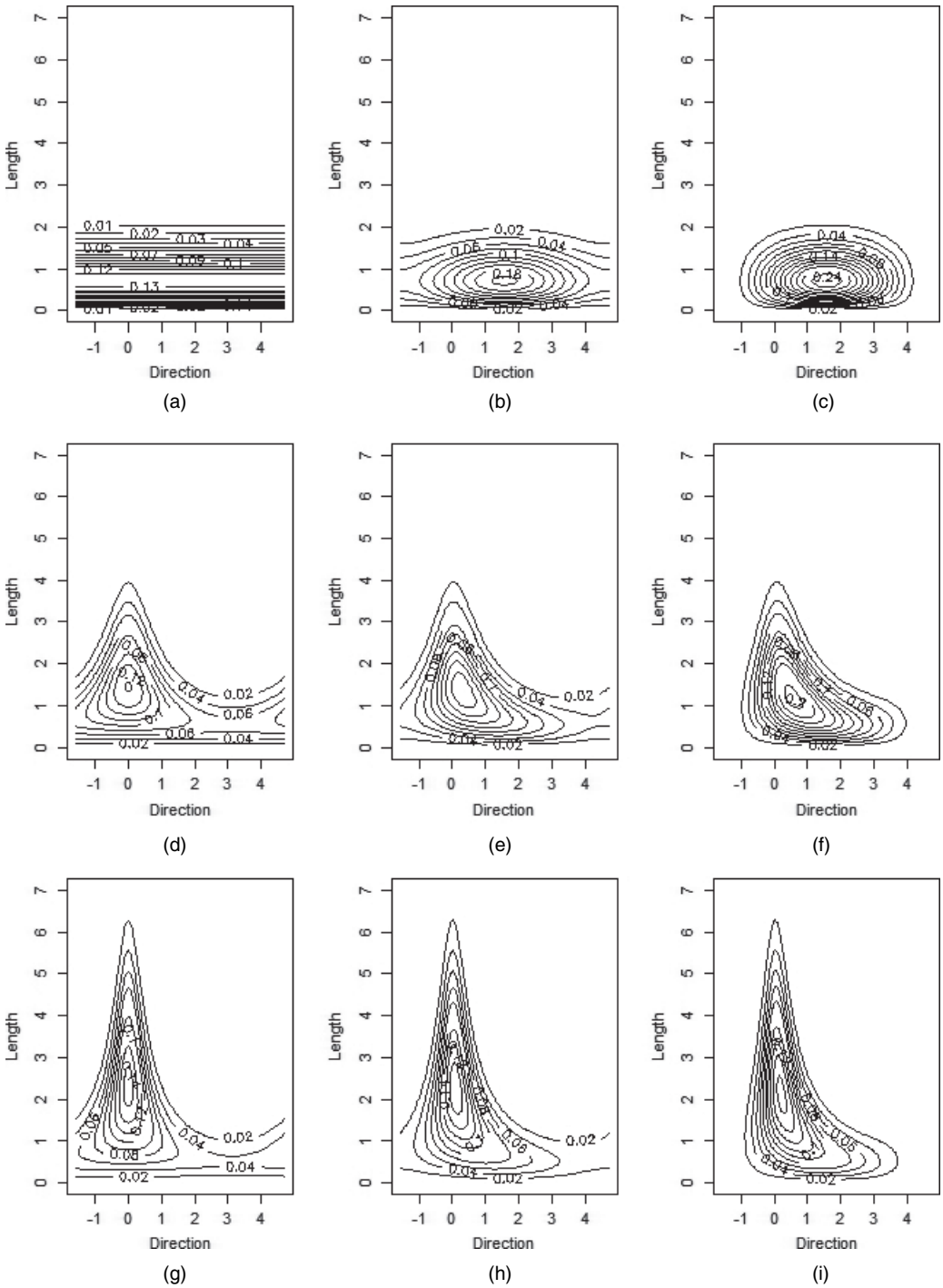


Fig. 3. Contour plots of Abe–Ley cylindrical density over $[0,5)[\pi/2,3/2\pi)$ for $(\alpha,\beta,\mu)=(2,1,0)$ with (a) $(\kappa,\lambda)=(0,0)$ (Weibull and uniform), (b) $(\kappa,\lambda)=(0,0.5)$ (Weibull and cardioid), (c) $(\kappa,\lambda)=(0,1)$ (Weibull and cardioid), (d) $(\kappa,\lambda)=(1,0)$, (e) $(\kappa,\lambda)=(1,0.5)$, (f) $(\kappa,\lambda)=(1,1)$, (g) $(\kappa,\lambda)=(1.5,0)$, (h) $(\kappa,\lambda)=(1.5,0.5)$ and (i) $(\kappa,\lambda)=(1.5,1)$

y_i, observed at n observation points, say $\mathbf{z} = (\mathbf{z}_i, i = 1, \ldots, n)$, $\mathbf{z}_i = (x_i, y_i)$, $x_i \in [0, 2\pi)$ and $y_i \in [0, \infty)$. The proposed HMRF is specified by assuming that these cylindrical observations are conditionally independent, given a segmentation generated by a latent Potts model. Precisely, we assume that the conditional distribution of the observed process, given the latent segmentation $\boldsymbol{\xi}$, takes the form of a product density, say

$$f(\mathbf{z} \mid \boldsymbol{\xi}; \boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{k=1}^{K} f(\mathbf{z}_i; \boldsymbol{\theta}_k)^{\xi_{ik}}, \tag{4}$$

where the vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$ includes K label-specific parameters and $f(\mathbf{z}; \boldsymbol{\theta}_k)$, $k = 1, \ldots, K$, are K Abe–Ley bivariate densities (3). The joint density of the observed data and the unobserved class memberships is therefore given by

$$f(\mathbf{z}, \boldsymbol{\xi}; \boldsymbol{\theta}, \rho) = f(\mathbf{z} \mid \boldsymbol{\xi}; \boldsymbol{\theta})\, p(\boldsymbol{\xi}; \rho). \tag{5}$$

By integrating this distribution with respect to the segmentation $\boldsymbol{\xi}$, we obtain the likelihood function of the unknown parameters

$$L(\boldsymbol{\theta}, \rho) = \sum_{\boldsymbol{\xi}} f(\mathbf{z}, \boldsymbol{\xi}; \boldsymbol{\theta}, \rho). \tag{6}$$

4. Composite-likelihood-based inference

4.1. Expectation–maximization algorithm

Direct maximization of the likelihood function (6) in unfeasible. As a result, we propose to estimate the parameters by maximizing a surrogate function, namely a composite log-likelihood function. Our proposal is based on the specification of a cover $\mathbb{A}$ of the set $S = \{1 \ldots n\}$ of the observation sites, i.e. a family of (not necessarily disjoint) subsets $A \subseteq S$ such that $\cup_{A \in \mathbb{A}} = S$. For each subset A, we define

$$L_A(\boldsymbol{\theta}, \rho) = \sum_{\boldsymbol{\xi}_A} f(\mathbf{z}_A, \boldsymbol{\xi}_A; \boldsymbol{\theta}, \rho)$$

as the contribution of the data in A to the CL function, where

$$f(\mathbf{z}_A, \boldsymbol{\xi}_A; \boldsymbol{\theta}, \rho) = p(\boldsymbol{\xi}_A; \rho) \prod_{i \in A} \prod_{k=1}^{K} f(\mathbf{z}_i; \boldsymbol{\theta}_k)^{\xi_{ik}}$$

whereas

$$p(\boldsymbol{\xi}_A, \rho) = \frac{\exp\left\{ (\rho/2) \sum_{i \in A} \sum_{j \in A} c_{ij} \boldsymbol{\xi}_i^{\mathrm{T}} \boldsymbol{\xi}_j \right\}}{W(\rho)}$$

and, finally,

$$W(\rho) = \sum_{\boldsymbol{\xi}_A} p(\boldsymbol{\xi}_A; \rho).$$

Given a choice of the cover $\mathbb{A}$, we propose to estimate the parameters by maximizing the composite log-likelihood function

$$\mathrm{cl}(\boldsymbol{\theta}, \rho; \mathbb{A}) = \sum_{A \in \mathbb{A}} \log\{L_A(\boldsymbol{\theta}, \rho)\}. \tag{7}$$

In particular, we suggest an EM algorithm for maximization, based on the complete-data composite log-likelihood function

$$\mathrm{cl}_{\mathrm{c}}(\boldsymbol{\theta}, \rho; \mathbb{A}) = \sum_{A \in \mathbb{A}} \{\mathrm{cl}_{\mathrm{c}}^{A}(\boldsymbol{\theta}) + \mathrm{cl}_{\mathrm{c}}^{A}(\rho)\},$$

where

$$\mathrm{cl}_{\mathrm{c}}^{A}(\boldsymbol{\theta}) = \sum_{i \in A} \sum_{k=1}^{K} \xi_{ik} \log\{f(\mathbf{z}_i; \boldsymbol{\theta}_k)\}$$

and

$$\mathrm{cl}_{\mathrm{c}}^{A}(\rho) = \sum_{\boldsymbol{\xi}_A} \boldsymbol{\xi}_A \log\{p(\boldsymbol{\xi}_A; \rho)\}.$$

At the $(q+1)$th iteration of the algorithm, we compute the expected value of the complete-data composite log-likelihood with respect to the predictive distribution of the segmentation. This E-step reduces to the computation of the predictive probabilities

$$\hat{\boldsymbol{\xi}}_A = p(\boldsymbol{\xi}_A \mid \mathbf{z}_A, \hat{\rho}_q, \hat{\boldsymbol{\theta}}_q) = \frac{p(\boldsymbol{\xi}_A; \hat{\rho}_q) f(\mathbf{z}_A; \hat{\boldsymbol{\theta}}_{qk})}{\sum_{\boldsymbol{\xi}_A} p(\boldsymbol{\xi}_A; \hat{\rho}_q) f(\mathbf{z}_A; \hat{\boldsymbol{\theta}}_{qk})} \tag{8}$$

for each $A \in \mathbb{A}$. Suitable marginalization of expression (8) provides the univariate probabilities $\hat{\boldsymbol{\xi}}_i = p(\boldsymbol{\xi}_i \mid \mathbf{z}_A, \hat{\rho}_q, \hat{\boldsymbol{\theta}}_q)$, for each $i \in A$. We then maximize the expected complete-data composite log-likelihood (M-step). Because this function is the sum of two components that depend on different sets of parameters, the M-step reduces to the separate maximization of the two functions, namely

$$Q(\boldsymbol{\theta}; \mathbb{A}) = \sum_{A \in \mathbb{A}} \sum_{i \in A} \sum_{k=1}^{K} \hat{\xi}_{ik} \log\{f(\mathbf{z}_i; \boldsymbol{\theta}_k)\}, \tag{9}$$

$$Q(\boldsymbol{\rho}; \mathbb{A}) = \sum_{A \in \mathbb{A}} \sum_{\boldsymbol{\xi}_A} \hat{\boldsymbol{\xi}}_A \log\{p(\boldsymbol{\xi}_A; \rho)\}. \tag{10}$$

Maximization of equation (9) can be undertaken by an unconstrained maximization algorithm, after a suitable reparameterization of the parameters involved. Precisely, we maximize equation (9) over the parameter vector $\boldsymbol{\theta}_k = (\theta_{1k}, \theta_{2k}, \theta_{3k}, \theta_{4k}, \theta_{5k})$, where $\theta_{1k} = \log(\alpha_k)$, $\theta_{2k} = \log(\beta_k)$, $\theta_{3k} = \log(\kappa_k)$, $\theta_{4k} = \tan(\mu_k/2)$ and, finally, $\theta_{5k} = \tanh^{-1}(\lambda_k)$, by exploiting a quasi-Newton procedure like that provided, for example, by the function `optim` in R. Maximization of equation (10) over the domain $(0, \rho_{\mathrm{crit}})$ can be carried out by a constraint optimization algorithm like that provided by the option `L-BFGS-B` method of `optim`.

Both the E- and the M-step of the algorithm proposed involve summations over all the $K^{|A|}$ possible values that $\boldsymbol{\xi}_A$ can take. As a result, the numerical tractability of these steps dramatically decreases with the cardinality of the largest subset of the cover $\mathbb{A}$. On the one side, this would suggest choosing a cover with many small subsets. On the other side, a cover that includes a few large subsets is expected to provide a CL function that is a better approximation of the likelihood function. Because summations over $\boldsymbol{\xi}_A$ become cumbersome for $|A| \geqslant 3$, a natural strategy is a cover that includes subsets with two elements. When $\mathbb{A}$ includes all the subsets of two elements, then equation (7) reduces to the pairwise likelihood function (Varin *et al.*, 2011). In a spatial setting, a pairwise likelihood can be further simplified by discarding all the pairs (i, j) that are not in the neighbourhood structure E. This choice provides a computationally efficient EM algorithm, without sacrificing the good distributional properties that are expected from a CL estimator.

4.2. Further computational details

It is well known that the EM algorithm suffers from two drawbacks: it is sensitive to the choice of starting points and it could converge to local maxima. These two aspects are strictly linked to each other. To avoid local maxima we follow a short-runs strategy, by running the EM algorithm from 50 random initializations, and stopping the algorithm without waiting for full convergence, i.e. when the relative increase in two consecutive composite log-likelihoods is less than 10^{-2}. The best solution is taken as the starting point to run the EM algorithm until full convergence, i.e. when the difference in two consecutive composite log-likelihoods is less than 10^{-5}.

Using an i7 processor (2.50 GHz), and depending on the scenarios, the computational time of a single short run was rarely greater than 20 s, whereas a single long run could take up to 121 s. Therefore the computational cost of the estimation strategy proposed essentially depends on the number of short runs. Computational speed can be improved by choosing a small number of short runs, at the price of a high risk of convergence to a local maximum. More efficiently, computational speed can be improved by accelerating the last steps of each EM run by direct CL maximization. In principle, equation (7) could be directly maximized by using a suitable optimization algorithm, chosen for example from among the optimization methods that are available with the R optimization routine `optim`. We, however, found that, if the initial conditions are sufficiently close to the maximum CL estimate, then direct maximization works and is generally very fast. Otherwise, if initialized outside a neighbourhood of the maximum CL estimate, direct optimization often crashes. It is possible that the methods of `optim` are not suitable for the quite unconventional case of a mixture of cylindrical distributions. Nevertheless, our finding supports a hybrid strategy that starts with EM iterations and then replaces the last iterations of the EM algorithm by direct maximization. In this way, we obtain a hybrid algorithm that essentially compensates the large circle of convergence provided by the EM algorithm with the high speed of direct numerical maximization. Our experiments indicate that this idea can save up to 30% of the computational time.

Standard errors could in principle be obtained by numerically approximating the observed Godambe matrix (Godambe, 1960), which is, however, known to suffer from numerical instability. It requires both the numerical approximation of variability and sensitivity matrices, and the inversion of the variability matrix (i.e. the covariance of the CL score), which is usually a large size matrix. A viable alternative relies on parametric bootstrap methods, to obtain quantiles of the distribution of the estimates. In this paper, we refitted the model to $R = 200$ bootstrap samples, which were simulated from the estimated model parameters. We then computed the 2.5% and the 97.5% quantiles of the empirical distribution of each bootstrap estimate. Simulation of the cylindrical HMRF is straightforward, by taking advantage of standard simulation routines that are available for the Potts model and the Abe–Ley distribution. Specifically, we exploit the Swendsen–Wang algorithm (Swendsen and Wang, 1987) (which is available in the R package `potts`) to simulate a configuration of segmentation labels. Given a configuration of segmentation labels, a cylindrical observation $\mathbf{z}_i = (\mathbf{x}_i, \mathbf{y}_i)$ is drawn at each lattice site i, according to the appropriate Abe–Ley distribution, evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_{k_i}$, where k_i is the segmentation label that has been generated by the Potts model at location i. Under equation (3), the marginal distribution of the angular variable x is a sine-skewed wrapped Cauchy distribution, whereas the conditional distribution of y given x is Weibull. We therefore draw a cylindrical sample by first drawing from a sine-skewed wrapped Cauchy distribution and then drawing a sample from the Weibull distribution. Abe and Ley (2016) suggested a simple routine to simulate from a sine-skewed wrapped Cauchy distribution and we follow their proposal.

4.3. Model selection

The number K of latent classes is chosen by selecting the model minimizing the CL Bayes information criterion C-BIC (Gao and Song, 2010). It combines the goodness of fit for a given model (minus twice the composite log-likelihood) and the penalty term of model complexity, namely

$$\text{C-BIC} = -2\,\text{cl}(\hat{\boldsymbol{\theta}}, \hat{\rho}; \mathbf{x}) + \text{tr}\{\mathbf{s}(\hat{\boldsymbol{\theta}}, \hat{\rho})\mathbf{s}(\hat{\boldsymbol{\theta}}, \hat{\rho})'\mathbf{I}(\hat{\boldsymbol{\theta}}, \hat{\rho})^{-1}\}\log(n), \tag{11}$$

where $\mathbf{s}(\cdot)$ is the observed score, and $\mathbf{I}(\cdot)$ is the observed information matrix. It extends the traditional Bayes information criterion to the CL framework, where the identity $\mathbf{s}(\hat{\boldsymbol{\theta}}, \hat{\rho})\mathbf{s}(\hat{\boldsymbol{\theta}}, \hat{\rho})' = \mathbf{I}(\hat{\boldsymbol{\theta}}, \hat{\rho})$ does not hold. In this paper, the empirical estimates of $\mathbf{s}(\hat{\boldsymbol{\theta}}, \hat{\rho})\mathbf{s}(\hat{\boldsymbol{\theta}}, \hat{\rho})'$ and $\mathbf{I}(\hat{\boldsymbol{\theta}}, \hat{\rho})^{-1}$ were computed by exploiting the R package `numDeriv` to estimate both the score and the Hessian of the observed CL. Although this method seems viable for small values of K, it tends to become numerically unstable as K increases.

5. Simulation study

A simulation study was carried out to explore the distributional properties of the proposed estimation method across a battery of $12 = 2 \times 2 \times 3$ contrived but realistic parameter groups. Table 1 displays the parameter values of each scenario. We considered HMRFs with $K = 2$ and $K = 3$ latent classes which are associated with either strongly separated or weakly separated cylindrical densities and affected by three levels of spatial dependence: no spatial dependence, weak spatial dependence and strong spatial dependence. These eight studies were respectively repeated on two square lattices with 10×10 and 30×30 sites. Each study was carried out by drawing 500 samples from the appropriate cylindrical hidden Markov model.

Figs 4 and 5 show the distributions of the parameter estimates, respectively obtained with $K = 2$ and $K = 3$ classes, using a 30×30 regular grid. The results that were obtained in the case of a 10×10 regular grid are in the on-line 'supporting information' file (Figs 1 and 2), which further includes the bias, the standard deviation, the mean, the median and the root-mean-squared error that have been obtained in the simulation study (Tables A1–A4 and Tables 1–4 in the on-line supporting information file).

Overall, all the distributions look quite symmetric and show little bias. Exceptions are the

Table 1. True values of model parameters over various scenarios

$K = 2$; strongly separated cylindrical classes					
$\alpha_1 = 0.5$	$\beta_1 = 0.1$	$\lambda_1 = 0.5$	$\kappa_1 = 0.9$	$\mu_1 = 0$	$\rho = 0, 0.29, 0.44$
$\alpha_2 = 0.8$	$\beta_2 = 0.9$	$\lambda_2 = 0.5$	$\kappa_2 = 0.9$	$\mu_2 = 1.5$	
$K = 2$; weakly separated cylindrical classes					
$\alpha_1 = 0.5$	$\beta_1 = 0.3$	$\lambda_1 = 0.5$	$\kappa_1 = 0.75$	$\mu_1 = 0$	$\rho = 0, 0.29, 0.44$
$\alpha_2 = 0.75$	$\beta_2 = 0.6$	$\lambda_2 = 0.5$	$\kappa_2 = 0.75$	$\mu_2 = 1.0$	
$K = 3$; strongly separated cylindrical classes					
$\alpha_1 = 0.5$	$\beta_1 = 0.1$	$\lambda_1 = 0.5$	$\kappa_1 = 0.9$	$\mu_1 = 0$	$\rho = 0, 0.34, 0.50$
$\alpha_2 = 0.8$	$\beta_2 = 0.9$	$\lambda_2 = 0.5$	$\kappa_2 = 0.9$	$\mu_2 = 1.5$	
$\alpha_3 = 1.3$	$\beta_3 = 1.7$	$\lambda_3 = 0.5$	$\kappa_3 = 0.9$	$\mu_3 = 3$	
$K = 3$; weakly separated cylindrical classes					
$\alpha_1 = 0.5$	$\beta_1 = 0.3$	$\lambda_1 = 0.5$	$\kappa_1 = 0.75$	$\mu_1 = 0$	$\rho = 0, 0.34, 0.50$
$\alpha_2 = 0.75$	$\beta_2 = 0.6$	$\lambda_2 = 0.5$	$\kappa_2 = 0.75$	$\mu_2 = 1.0$	
$\alpha_3 = 1.0$	$\beta_3 = 0.9$	$\lambda_3 = 0.5$	$\kappa_3 = 0.75$	$\mu_3 = 2$	

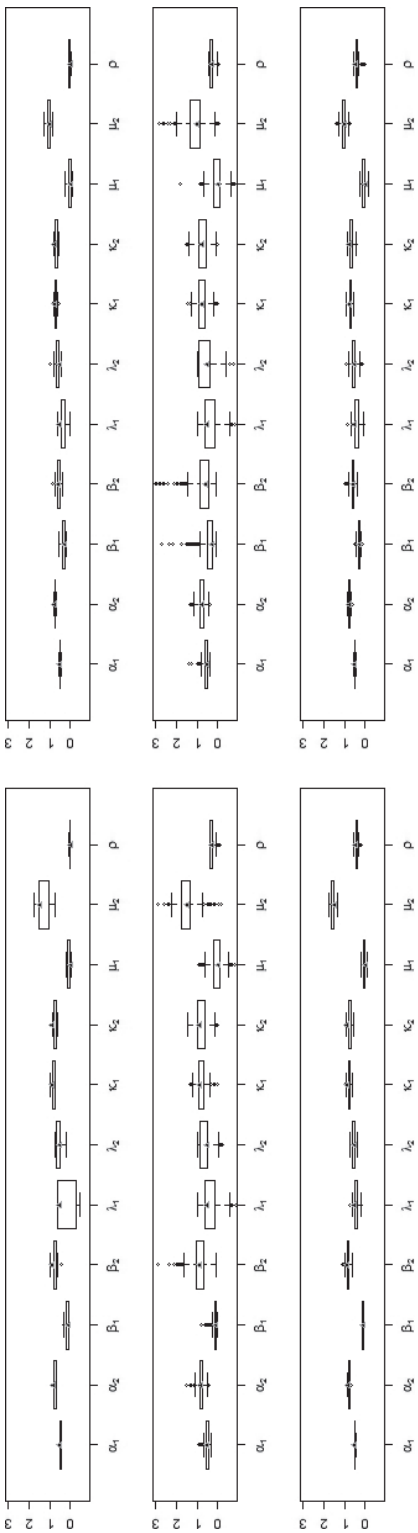


Fig. 4. Boxplots of parameter estimates (500 random samples with $K = 2$; grid 30 × 30 considering three different values for ρ; Δ, true values): (a) strongly separated classes; (b) weakly separated classes

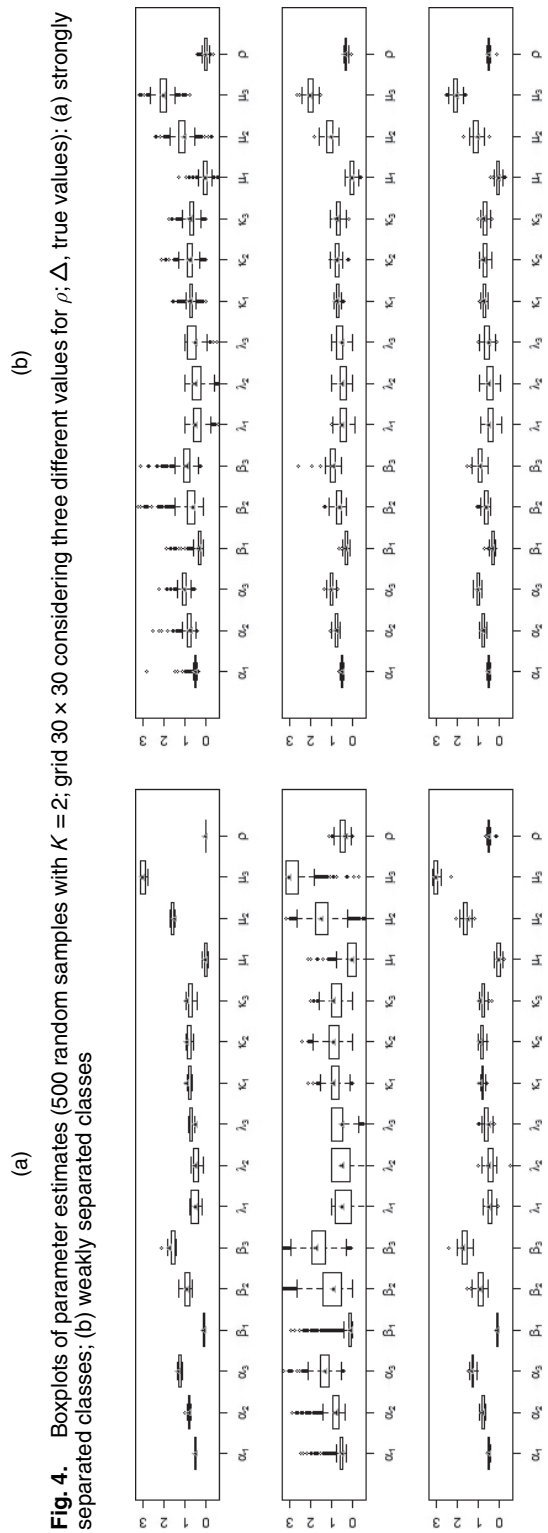


Fig. 5. Boxplots of parameter estimates (500 random samples with $K = 3$; grid 30 × 30 considering three different values for ρ; Δ, true values): (a) strongly separated classes; (b) weakly separated classes

estimate of the asymmetry parameter λ and the concentration parameter κ. Furthermore, in the case of a 30×30 lattice, the bias and the variance of all parameter estimates are smaller under all three dependence structures (both when $K = 2$ and when $K = 3$) than their counterparts obtained for a 10×10 lattice. When the latent classes are weakly separated, the estimates are affected by a larger variability than in the case of strongly separated classes. In keeping with known results about the efficiency of the CL *versus* the efficiency of the full likelihood (Xu and Reid, 2011; Cox and Reid, 2004), the quality of the CL estimates decreases when the spatial dependence parameter takes moderate values and then improves again as spatial dependence increases, reflecting a smaller discrepancy between the CL and the full likelihood. However, the bias of the dependence parameter κ and the skewness parameter λ is typically larger than the bias of the other estimates.

6. Application

We have estimated a number of cylindrical HMRFs from the two circulation fields that are displayed in Fig. 1, by varying the number K of components from 2 to 5. Table 2 shows the log-likelihood values and the C-BIC-statistics that were obtained for each model and each circulation field.

Table 2 suggests a model with $K = 2$ components for circulation field A, correctly reflecting a scenario where strong remote forcing is associated with south-eastward currents. A model with $K = 3$ components is instead selected for field B, where eastern and southern currents of similar speed are associated with low intensity flows whose direction is scattered across the gulf.

For each model, Table 3 displays the maximum CL estimates and the 2.5% and 97.5% bootstrap quantiles of the parameters of the conditional cylindrical distributions within each state. It further includes the maximum CL estimate and the bootstrap quantiles of the spatial dependence parameter of the Potts model.

Table 3 provides two general pieces of evidences that support the distributional choices of this paper. First, for each model, and within each state, the circular–linear dependence parameter κ is significant. This supports the choice of a cylindrical density and indicates that, at least in these cases, a conditional independence assumption between univariate distributions of circular and linear variables is unrealistic. Second, for each model, the spatial dependence parameter ρ is significant. This supports the inclusion of a spatial process to account for spatial autocorrelation and indicates that the assumption of spatial independence is unrealistic.

The rest of Table 3 should be interpreted with the help of Figs 6 and 7, which overlap the state-specific densities on the data points, filled with grey levels according to the posterior membership probabilities under each state.

Table 2. Log-likelihoods and C-BIC statistics†

Number of components	*Results for field A*		*Results for field B*	
	Composite log-likelihood	*C-BIC*	*Composite log-likelihood*	*C-BIC*
2	−1131.358	*3114.588*	−1219.373	3294.301
3	−1111.528	3438.180	−1207.382	*3270.332*
4	−1105.844	5536.709	−1199.468	3524.616
5	−1098.162	3305.920	−1183.930	3505.404

†The best models are indicated by italics.

Table 3. Estimates and 2.5% and 97.5% quantiles of two cylindrical HMRFs with two and three states

State	*Parameter*	*Results for field A*			*Results for field B*		
		Estimate	*2.5% quantile*	*97.5% quantile*	*Estimate*	*2.5% quantile*	*97.5% quantile*
1	α	3.76	2.49	3.82	3.26	3.06	4.55
	β	0.20	0.18	0.25	0.19	0.15	0.21
	λ	−1.00	−1.00	1.00	0.50	0.40	1.00
	κ	2.58	2.01	2.98	1.83	1.01	2.53
	μ	2.72	0.10	2.74	4.52	4.16	5.20
2	α	1.68	1.58	2.29	6.56	3.57	6.98
	β	0.20	0.16	0.23	0.11	0.10	0.15
	λ	−0.27	−0.56	0.24	−1.00	−1.00	−0.55
	κ	0.70	0.60	1.34	1.96	1.54	2.13
	μ	3.14	2.78	3.16	0.42	−0.51	1.02
3	α				2.80	2.30	3.60
	β				0.24	0.21	0.27
	λ				−1.00	−1.00	−0.95
	κ				1.30	0.85	1.99
	μ				0.69	0.56	1.18
	ρ	0.72	0.56	0.83	0.25	0.17	0.29

Field A is described in terms of two conditional distributions that represent well-defined circulation patterns. The first distribution (state 1) is associated with spatially coherent, high speed water flows ($\alpha = 3.76$ and $\beta = 0.20$), partly driven by northern winds that blow across the Tyrrhenian Sea. As a result, most of the data with the highest speed in the sample are clustered within this regime (Figs 6(a) and 6(b)). These currents are highly concentrated ($\kappa = 2.58$) around one modal direction ($\mu = 2.72$). Under this state, the distribution of direction is, however, negatively skewed ($\lambda = -1$), because part of this flow slows down as it easterly approaches the gulf coastline. The second distribution (state 2) is instead associated with currents that are of lower intensity ($\alpha = 1.68$ and $\beta = 0.20$) than those captured by the first component of the model. The small value that is reached by κ (0.70) indicates that these flows spread across the gulf without a privileged direction and that the dependence between speeds and directions is weak. In addition, the asymmetry parameter is not significant, indicating that the distribution of the directions is, under this state, essentially symmetric. The moderate skewness of the speeds under this regime is probably due to a few outliers. The contour plot of this density (Fig. 6(c)) suggests that this state clusters noisy currents that move around the main northern water flow.

In contrast, field B is described in terms of three conditional distributions (Fig. 7). The intensity of the currents that are generated by the first distribution (state 1; $\alpha = 3.26$ and $\beta = 0.19$) is similar to that generated under state 1 in field A. However, there are two important differences. First, the modal direction is toward west ($\mu = 4.52$). Second, the distribution of the directions is not remarkably skew ($\lambda = 0.50$), as this current does not encounter main obstacles when it travels towards the open sea. State 2 is instead associated with high speed currents that travel towards north-east. As expected, the distribution of the directions is significantly skew, as this current approaches the gulf coastline and the current speed at each direction depends on the orography of the coastline. The third distribution (state 3) is finally associated with a transition state between state 1 and state 2 and clusters north-eastern ($\mu = 0.69$) currents of

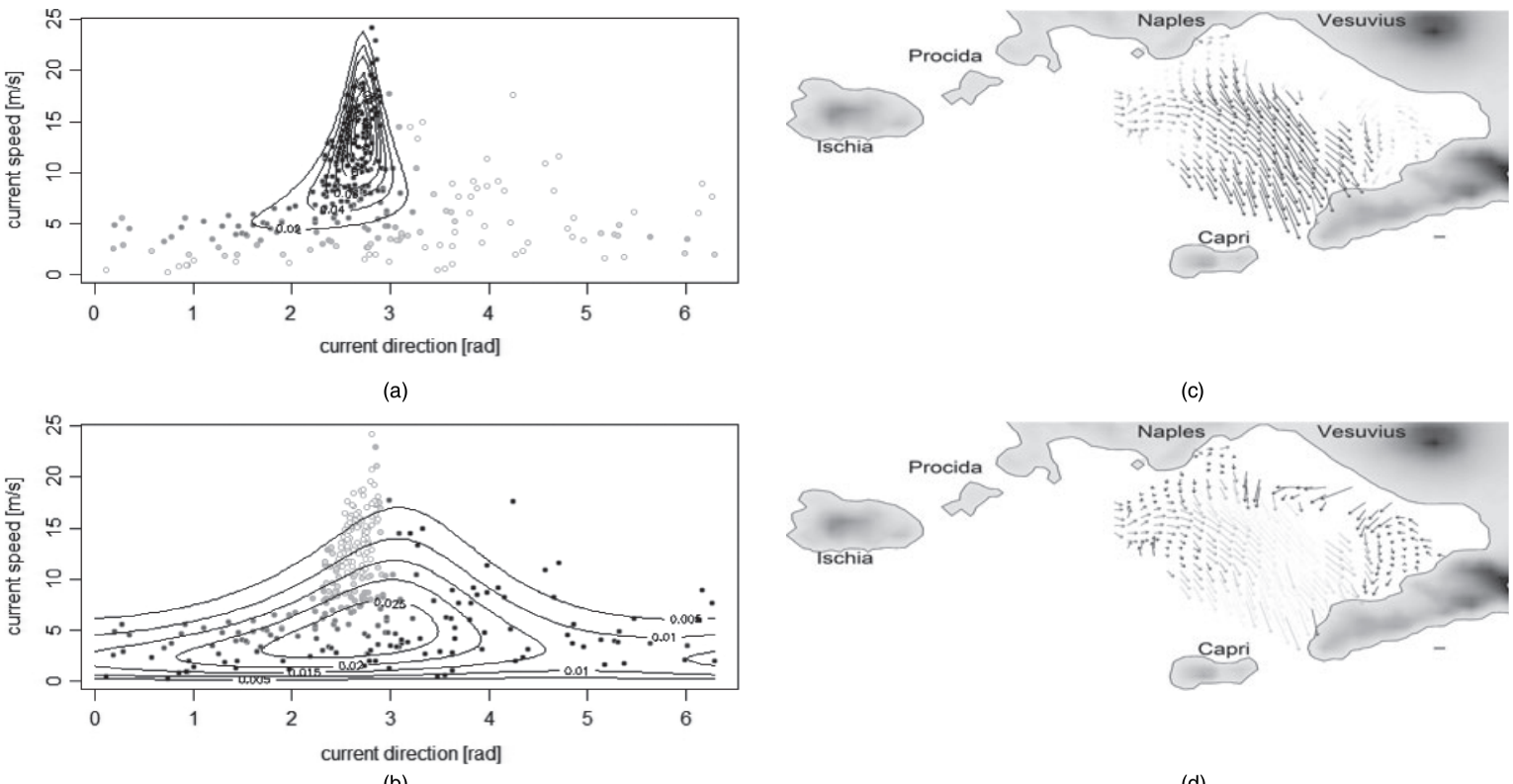


Fig. 6. Field A—(a), (b) state-specific cylindrical distributions and (c), (d) spatial distributions of the data, according to a cylindrical HMRF with two states (points and arrows are coloured with grey levels according to the estimated membership probabilities (black indicates a probability equal to 1)): (a), (c) state 1; (b), (d) state 2

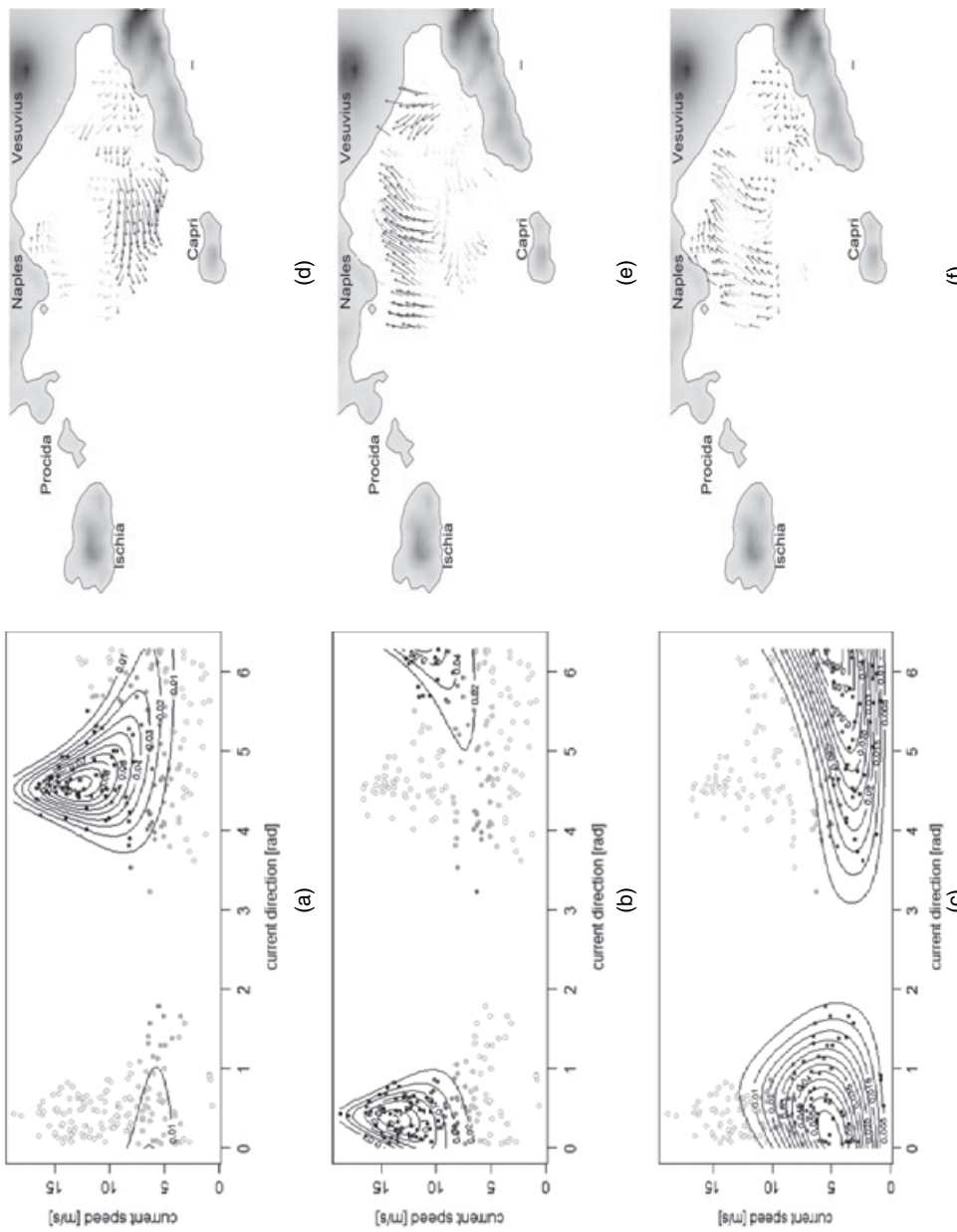


Fig. 7. Field B—(a), (b), (c) state-specific cylindrical distributions and (d), (e), (f) spatial distributions of the data, according to a cylindrical HMRF with three states (points and arrows are coloured with grey levels according to the estimated membership probabilities (black indicates a probability equal to 1)): (a), (d) state 1; (b), (e) state 2; (c), (f) state 3

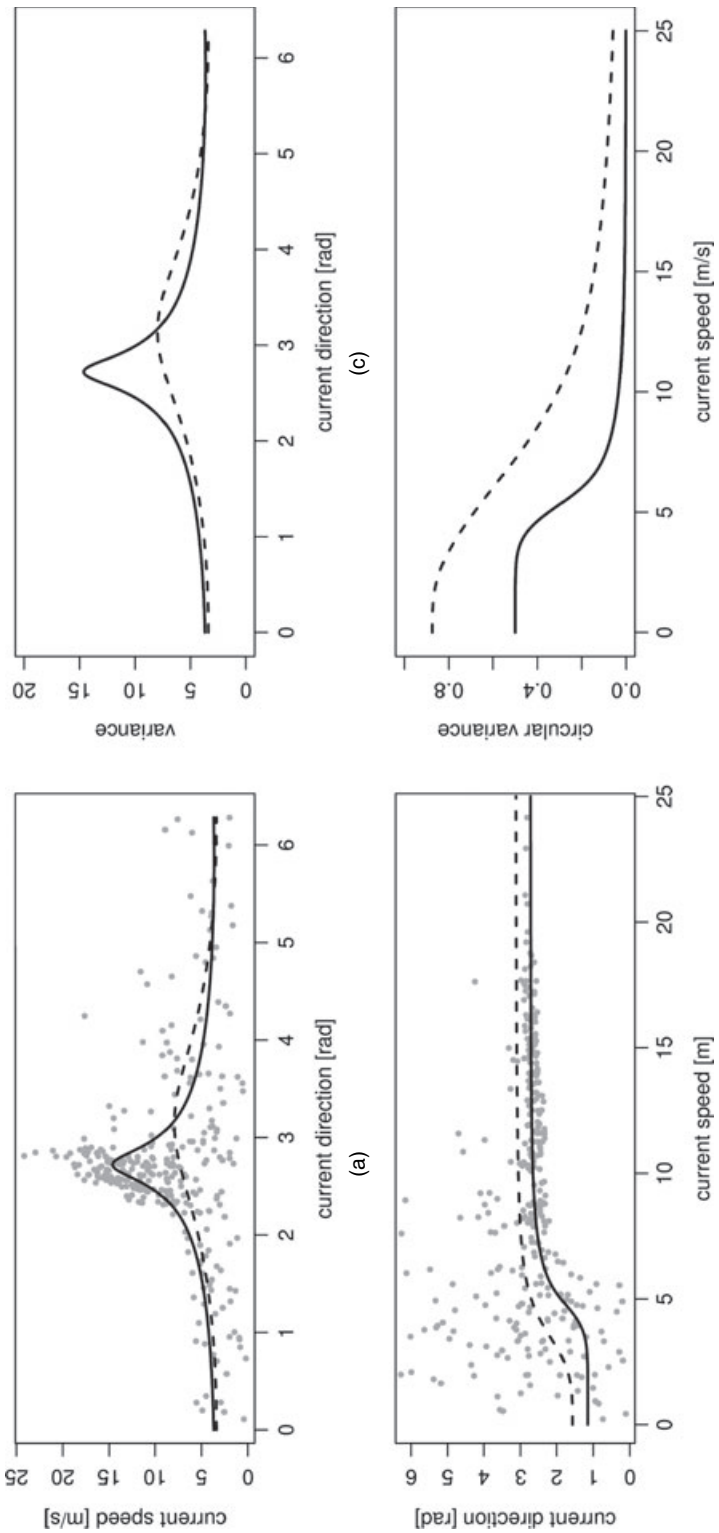


Fig. 8. Field A—(a) state-specific linear–circular (conditional mean of current speed) and (b) circular–linear (conditional mean of current direction) regression functions, and (c) state-specific linear (conditional variance of current speed) and (d) circular (conditional variance of current direction) variance functions: ———, state 1; – – –, state 2

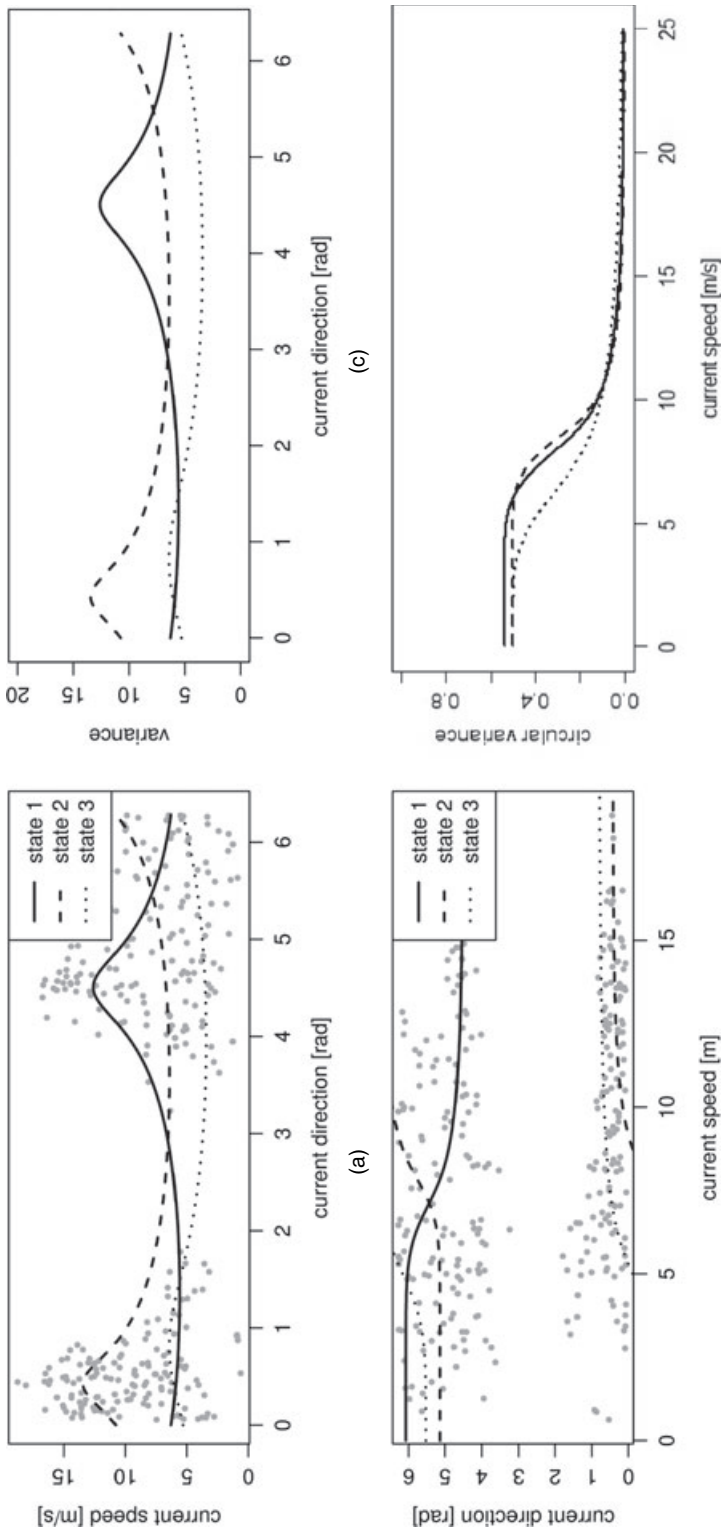


Fig. 9. Field B—(a) state-specific linear–circular (conditional mean of current speed) and (b) circular–linear (conditional mean of current direction) regression functions, and (c) state-specific linear (conditional variance of current speed) and (d) circular (conditional variance of current direction) variance functions: ———, state 1; – – –, state 2; · · · · · ·, state 3

moderate speed ($\alpha = 2.80$ and $\beta = 0.24$). Compared with the data of field A, field B data are more heterogeneous and, as a result, an additional transition component is required to provide a reasonable approximation of the data distribution. Further, the estimated spatial dependence parameter ($\rho = 0.25$) is significantly less than its counterpart of field A ($\rho = 0.72$), indicating that field B is more heterogeneous not only in the variable space but also across the spatial domain.

We recall that, under each latent state, the conditional distribution of the intensity y given the angular direction x is Weibull with state-specific parameters and that the conditional distribution of the direction x given the intensity y is a skew von Mises distribution with state-specific parameters (Section 3.2). The estimates of Table 3 can therefore be exploited to compute regime-specific regressions and variance functions, by respectively evaluating the conditional mean and the conditional variance of speed given direction and direction given speed at the maximum CL estimates. Figs 8(a) and 8(b) and Figs 9(a) and 9(b) respectively display the regression functions of current speed and current direction, under each latent state. For simplicity, these curves are depicted on the plane, but they actually live on the surface of a cylinder. These regression functions not only describe the non-linear functional relationship between current speed and direction in two case-studies but also, and more interestingly, demonstrate that the shape of such relationships dramatically changes with the latent state. This indicates that, in general, current speed should not be used as a predictor of direction (and, vice versa, current direction should not be used as a predictor of current speed), without accounting for the environmental heterogeneity that is detected by the latent states. The variance functions in Figs 8(c) and 8(d) and 9(c) and 9(d) further reflect the distributional skewness of both the linear and the angular components of the data. In particular, the shape that is taken by the conditional variance of direction given speed indicates that the variability of current directions decreases with current speed, although at state-specific paces. Even though this analysis is restricted to 2-hourly snapshots of the surface circulation in this area, our results suggests that in the Gulf of Naples faster currents tend to be concentrated around single modal directions, as shown, for example, by the fields presented by Menna *et al.* (2007). This is generally true for coastal semienclosed areas, representing a characteristic of circulation constrained in confined basins.

7. Concluding remarks

We proposed a hierarchical model for spatial cylindrical data that parsimoniously integrates a parametric cylindrical density and a latent MRF. It segments the data according to cylindrical latent classes, by simultaneously accounting for spatial dependence and unobserved heterogeneity. In the two specific cases that are considered in the paper, the model offered a clear-cut description of sea current patterns in terms of intuitively appealing environmental regimes. It correctly captured regime-specific non-linear relationships between the speed and the direction of the currents, through state-specific regression functions and variance functions. It finally provided a classification that reflects the orography of the study area on current dynamics, through regime-specific skew cylindrical densities.

Our proposal is motivated by issues that arise in marine studies, but it can be easily adapted to a wide range of real world cases, including for example environmental studies of wind fields, where the speed and the direction of wind are recorded across space, or ecological studies of animal behaviour, where direction and speed of movements are recorded across space.

A limit of the model is the intractable likelihood function, which complicates parameter estimation. This issue was addressed by taking a CL approach, which is based on the definition of a class of subsets that covers the study area. The numerical tractability of this approach depends on the size of the largest covering subset. This suggested the use of pairs of neighbouring sites as

covering subsets. This method, however, depends on a spatial neighbourhood structure, namely the nearest neighbourhood structure, which is assumed *a priori* on the spatial lattice and which is not necessarily the best choice for defining the CL function. However, the simulation studies that have been carried out show that this method provides estimates with good distributional properties.

A second limit of the model is represented by the assumption of a homogeneous latent field, which depends on a single spatial interaction parameter. We assume that the spatial auto-correlation is constant across space. This assumption is reasonable in studies that involve relatively small areas, such as in the case of the Gulf of Naples that was considered in this paper. Larger studies might require the specification of a non-homogeneous Potts model, with an auto-correlation parameter that depends on space varying environmental covariates.

Despite these limitations, the approach proposed flexibly describes the plasticity of the sea surface, indicating that the joint distribution of current speeds and directions changes under different environmental regimes. Regime switching not only changes directional and linear averages but also, and more interestingly, it shapes the non-linear functional relationship between speed and direction. That said, we remark that our proposal was conceptualized to segment coastal areas at a given point in time, and, as such, it does not include a temporal component of the sea circulation process. It cannot therefore be directly exploited for temporal forecasting. A good method of spatial segmentation is nevertheless a big step forward for reliable forecasting methods in a coastal setting. Our proposal could in fact be integrated with a hidden Markov model for cylindrical time series (Lagona *et al.*, 2015) to open up new perspectives in the analysis of space–time cylindrical data and, more ambitiously, to forecast sea motion in coastal areas. A possible approach in this direction, for example, is the specification of a sequence of conditionally independent hidden Markov fields, whose parameters evolve according to the states of a latent Markov chain. Further research is, however, needed to explore ways to integrate CL methods for spatial cylindrical data with the forward–backward estimation methods that are routinely exploited in models that involve latent Markov chains.

Acknowledgements

We thank the reviewers and the Associate Editor for their helpful comments and suggestions. Francesco Lagona was supported by the 2015 Progetti di Rilevante Interesse Nazionale supported project 'Environmental processes and human activities: capturing their interactions via statistical methods', funded by the Italian Ministry of Education, University and Scientific Research. The work of Enrico Zambianchi was partly funded by Parthenope University individual and competitive research grants (DR 727/2015, 953/2016 and 954/2016).

References

Abe, T. and Ley, C. (2016) A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econmetr. Statist.*, to be published, doi `http://dx.doi.org/10.1016/j.ecosta.2016.04.001`.

Abraham, C., Molinari, N. and Servien, R. (2013) Unsupervised clustering of multivariate circular data. *Statist. Med.*, **32**, 1376–1382.

Bartolucci, F. and Besag, J. (2002) A recursive algorithm for Markov random fields. *Biometrika*, **89**, 724–730.

Bellomo, L., Griffa, A., Cosoli, S., Falco, P., Gerin, R., Iermano, I., Kalampokis, A., Kokkini, Z., Lana, A., Magaldi, M., Mamoutos, I., Mantovani, C., Marmain, J., Potiris, E., Sayol, J. M., Barbin, Y., Berta, M., Borghini, M., Bussani, A., Corgnati, L., Dagneaux, Q., Gaggelli, J., Guterman, P., Mallarino, D., Mazzoldi, A., Molcard, A., Orfila, A., Poulain, P.-M., Quentin, C., Tintoré, J., Uttieri, M., Vetrano, A., Zambianchi, E. and Zervakis, V. (2015) Toward an integrated hf radar network in the Mediterranean sea to improve search and rescue and oil spill response: the tosca project experience. *J. Oper. Oceanogr.*, **8**, no. 2, 95–107.

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B, **36**, 192–236.
Besag, J. (1975) Statistical analysis of non-lattice data. *Statistician*, **24**, 179–195.
Besag, J. (1977) Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika*, **64**, 616–618.
Bivand, R. S., Pebesma, E. J. and Gómez-Rubio, V. (2008) *Applied Spatial Data Analysis with R*. New York: Springer.
Buffoni, G., Falco, P., Griffa, A. and Zambianchi, E. (1997) Dispersion processes and residence times in a semi-enclosed basin with recirculating gyres: an application to the Tyrrhenian sea. *J. Geophys. Res. Oceans*, **102**, 18699–18713.
Celeux, G., Forbes, F. and Peyrard, N. (2003) EM procedures using mean field-like approximations for Markov model-based image segmentation. *Pattn Recogn.*, **36**, 131–144.
Cianelli, D., Falco, P., Iermano, I., Mozzillo, P., Uttieri, M., Buonocore, B., Zambardino, G. and Zambianchi, E. (2015) Inshore/offshore water exchange in the gulf of Naples. *J. Mar. Syst.*, **145**, 37–52.
Cianelli, D., Uttieri, M., Buonocore, B., Falco, P., Zambardino, G. and Zambianchi, E. (2012) Dynamics of a very special Mediterranean coastal area: the Gulf of Naples. In *Mediterranean Ecosystems: Dynamics, Management and Conservation* (ed. G. Williams), pp. 129–150. New York: Nova Science.
Cox, D. R. (1975) Partial likelihood. *Biometrika*, **62**, 269–276.
Cox, D. R. and Reid, N. (2004) A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729–737.
Eidsvik, J., Shaby, B. A., Reich, B. J., Wheeler, M. and Niemi, J. (2014) Estimation and prediction in spatial models with block composite likelihoods. *J. Computnl Graph. Statist.*, **23**, 295–315.
Falco, P., Buonocore, B., Cianelli, D., De Luca, L., Giordano, A., Iermano, I., Kalampokis, A., Saviano, S., Uttieri, M., Zambardino, G. and Zambianchi, E. (2016) Dynamics and sea state in the Gulf of Naples: potential use of high-frequency radar data in an operational oceanographic context. *J. Oper. Oceanogr.*, **9**, suppl. 1, s33–s45.
Friel, N. and Rue, H. (2007) Recursive computing and simulation-free inference for general factorizable models. *Biometrika*, **94**, 661–672.
Gao, X. and Song, P. X.-K. (2010) Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *J. Am. Statist. Ass.*, **105**, 1531–1540.
Godambe, V. P. (1960) An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208–1211.
Guyon, X. (1995) *Random Fields on a Network: Modeling, Statistics, and Applications*. New York: Springer.
Hanks, E. M., Hooten, M. B. and Alldredge, M. W. (2015) Continuous-time discrete-space models for animal movement. *Ann. Appl. Statist.*, **9**, 145–165.
Iermano, I., Moore, A. and Zambianchi, E. (2016) Impacts of a 4-dimensional variational data assimilation in a coastal ocean model of southern Tyrrhenian sea. *J. Mar. Syst.*, **154**, 157–171.
Klauenberg, K. and Lagona, F. (2007) Hidden Markov random field models for {TCA} image analysis. *Computnl Statist. Data Anal.*, **52**, 855–868.
Lagona, F. and Picone, M. (2016) Model-based segmentation of spatial cylindrical data. *J. Statist. Computn Simuln*, **86**, 2598–2610.
Lagona, F., Picone, M. and Maruotti, A. (2015) A hidden Markov model for the analysis of cylindrical time series. *Environmetrics*, **26**, 534–544.
Lillibridge, J. L. and Mariano, A. J. (2013) A statistical analysis of gulf stream variability from 18+ years of altimetry data. *Deep Sea Res.* part II, **85**, 127–146.
Lindsay, B. (1988) Composite likelihood methods. *Contemp. Math.*, **80**, 221–239.
Mardia, K. V., Kent, J. T., Hughes, G. and Taylor, C. C. (2009) Maximum likelihood estimation using composite likelihoods for closed exponential families. *Biometrika*, **96**, 975–982.
Menna, M., Mercatini, A., Uttieri, M., Buonocore, B. and Zambianchi, E. (2007) Wintertime transport processes in the gulf of Naples investigated by hf radar measurements of surface currents. *Nuovo Cim.* C, **30**, 605–622.
Modlin, D., Fuentes, M. and Reich, B. (2012) Circular conditional autoregressive modeling of vector fields. *Environmetrics*, **23**, 46–53.
Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer Science and Business Media.
Okabayashi, S., Johnson, L. and Geyer, C. (2011) Extending pseudo-likelihood for Potts models. *Statist. Sin.*, **21**, 331–347.
Reeves, R. and Pettitt, A. N. (2004) Efficient recursions for general factorisable models. *Biometrika*, **91**, 751–757.
Reich, B. and Fuentes, M. (2007) A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Ann. Appl. Statist.*, **1**, 249–264.
Rubio, A., Mader, J., Corgnati, L., Mantovani, C., Griffa, A., Novellino, A., Quentin, C., Wyatt, L., Schulz-Stellenfleth, J., Horstmann, J., Lorente, P., Zambianchi, E., Hartnett, M., Fernandes, C., Zervakis, V., Gorringe, P., Melet, A. and Puillat, I. (2017) HF radar activity in European coastal seas: next steps toward a pan-European HF radar network. *Front. Mar. Sci.*, **4**, article 8.
Swendsen, R. H. and Wang, J.-S. (1987) Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, **58**, 86–88.

Uttieri, M., Cianelli, D., Nardelli, B. B., Buonocore, B., Falco, P., Colella, S. and Zambianchi, E. (2011) Multi-platform observation of the surface circulation in the Gulf of Naples (southern Tyrrhenian sea). *Ocean Dyn.*, **61**, 779–796.

Varin, C., Reid, N. and Firth, D. (2011) An overview of composite likelihood methods. *Statist. Sin.*, **21**, 1–41.

Wang, F. and Gelfand, A. E. (2014) Modeling space and space-time directional data using projected Gaussian processes. *J. Am. Statist. Ass.*, **109**, 1565–1580.

Wang, F., Gelfand, A. and Jona-Lasinio, G. (2015) Joint spatio-temporal analysis of a linear and a directional variable: space-time modeling of wave heights and wave directions in the Adriatic sea. *Statist. Sin.*, **25**, 25–39.

Xu, X. and Reid, N. (2011) On the robustness of maximum composite likelihood estimate. *J. Statist. Planng Inf.*, **141**, 3047–3054.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Segmentation of sea current fields by cylindrical hidden Markov models: a composite likelihood approach'.