The Annals of Applied Statistics
2012, Vol. 6, No. 4, 1478–1498
DOI: 10.1214/12-AOAS576

SPATIAL ANALYSIS OF WAVE DIRECTION DATA USING WRAPPED GAUSSIAN PROCESSES

BY GIOVANNA JONA-LASINIO, ALAN GELFAND[1] AND MATTIA JONA-LASINIO

University of Rome Sapienza, Duke University and Leibniz University

Directional data arise in various contexts such as oceanography (wave directions) and meteorology (wind directions), as well as with measurements on a periodic scale (weekdays, hours, etc.). Our contribution is to introduce a model-based approach to handle periodic data in the case of measurements taken at spatial locations, anticipating structured dependence between these measurements. We formulate a wrapped Gaussian spatial process model for this setting, induced from a customary *linear* Gaussian process.

We build a hierarchical model to handle this situation and show that the fitting of such a model is possible using standard Markov chain Monte Carlo methods. Our approach enables spatial interpolation (and can accommodate measurement error). We illustrate with a set of wave direction data from the Adriatic coast of Italy, generated through a complex computer model.

1. Introduction. Directional or angular data arise, for instance, in oceanography (wave directions), meteorology (wind directions) and biology (study of animal movement). They also arise from periodic data, for example, event times might be wrapped to a weekly period to give a circular view (eliminating end effects) of the pattern of event times. Here, we assume the data is recorded in degrees or angles on a circle. This is not a limitation, as any circular scale (e.g., $[0, L)$ or $[-L/2, L/2)$) can be transformed to $[0, 2\pi)$ by a modulus transformation. Handling such data creates difficulties due to the restriction of support to the unit circle, $[0, 2\pi)$, and to the sensitivity of descriptive and inferential results to the starting point on the circle. Hence, analysis of directional data is more challenging than for linear data. There exists a substantial literature on circular data [see, e.g., Mardia (1972), Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001) or Fisher (1993)], but, broadly, it is confined to descriptive statistics and limited inference for simple univariate models.

The contribution of this paper is to introduce a fully model-based approach, that is, Bayesian hierarchical modeling, to handle angular data, enabling full inference regarding all model parameters and prediction under the model. Our focus is on multivariate directional observations arising as angular data measurements taken

Received October 2011; revised May 2012.

[1]Supported in part by NSF Grants DMS-09-14906 and CDI 0940671.

Key words and phrases. Bayesian kriging, Gaussian processes, hierarchical model, latent variables.

at spatial locations, anticipating structured dependence between these measurements. Thus, we formulate an attractive spatial process model for directional data, the wrapped Gaussian process, induced from a linear (customary) Gaussian process. We illuminate the dependence structure. We show how to implement kriging of mean directions and concentrations in this setting. We work within a hierarchical Bayesian framework and show that introduction of suitable latent variables facilitates Markov chain Monte Carlo model fitting. We offer an adaptive truncation strategy for simulation of these latent variables.

Directional data has a long history. Early contributors to the theoretical development include Watson and Stephens [Watson (1961), Stephens (1963, 1970)]. Kent (1978) studied complex circular distributions. The books of Mardia (1972) and Mardia and Jupp (2000) present approaches, distribution theory and inference for such data. In Fisher (1993) we find comprehensive discussion, with particular attention to nonparametric methods. Computational procedures such as MCMC methods and the EM algorithm have enabled analysis for directional data to become less descriptive and more inferential. Examples include linear models [Harrison and Kanji (1988), Fisher (1993), Fisher and Lee (1992), Kato, Shimizu and Shieh (2008)], linear models in a Bayesian context [Guttorp and Lockhart (1988), Damien and Walker (1999)] and models for circular time series [Breckling (1989), Coles (1998), Mardia and Jupp (2000), Ravindran (2002), Hughes (2007), Fisher and Lee (1994), Holzmann et al. (2006)]. Recently, Kato [Kato (2010)], building upon earlier work [Kato, Shimizu and Shieh (2008)], has proposed a discrete time Markov process for circular data using the Möbius circle transformation, connecting it with an early Markov process model of Fisher and Lee (1994). We offer a process model for locations in d-dimensional space but focus on the 2-dimensional case.

There is little in the way of formal multivariate theory for circular data, particularly in the fully Bayesian setting. In this regard, perhaps the work of Coles (1998) is the closest to ours. He also employs wrapped distributions, noting that, in the Gaussian case, they can be readily given a multivariate extension. Coles mostly works with independent replicates of multivariate circular data in low dimension with an unknown covariance matrix and develops some theory and examples for the time series setting. However, he mentions possible extensions to the spatial setting but offers no development, in particular, no thoughts on regression or kriging (Sections 3.5 and 3.6 below). Casson and Coles (1998) include spatial dependence in looking at the direction of maximum wind speed. With little detail, they propose conditionally independent directions modeled with a von Mises distribution, introducing spatial structure in the modal direction and concentration parameters. We introduce spatial structure directly on the angular variables, with a single, potentially high-dimensional multivariate observation but driven by a spatial process model, yielding a high-dimensional covariance matrix with structured dependence as a function of perhaps two or three parameters.

Our motivating example is drawn from marine data. Often, such data are based on outputs from deterministic models, usually climatic forecasts computed at several *spatial* and *temporal* resolutions. Wave heights and outgoing wave directions, the latter being measured in degrees relative to a fixed orientation, are the main outputs of marine forecasts. Numerical models for weather and marine forecasts need statistical post-processing; wave heights, like wind speed, being linear variables, can be treated in several ways [Kalnay (2002), Wilks (2006), Jona Lasinio et al. (2007)]. Wave directions, being angular variables, cannot be treated according to standard post-processing techniques [see Engel and Ebert (2007), Bao et al. (2010) and references therein]. In Bao et al. (2010) bias correction and ensemble calibration forecasts of surface wind direction are proposed. The authors use circular–circular regression as in Kato, Shimizu and Shieh (2008) for bias correction and Bayesian model averaging with the von Mises distribution for ensemble calibration. However, their approach does not explicitly account for spatial structure. In our setting, wave direction data is viewed differently from wind direction data. The former is only available as an angle, while the latter is customarily associated with wind speed, emerging as the resultant of North–South and East–West wind speed components.

Eventually, we plan to do joint spatio-temporal modeling of wave height and wave direction (linear and circular data), fusing numerical model output with observed buoy data. As a first step, here, we take up static spatial modeling for the WAve Model (WAM) data (see Section 4.2), deferring the joint modeling, dynamics and data fusion for a future paper.

The format of the paper is as follows. Section 2 reviews wrapped distributions for univariate circular data. Section 3 moves on to the wrapped Gaussian process. Section 4 takes up the wave direction application and Section 5 offers a summary and future directions.

2. Univariate wrapped distributions. The von Mises (or circular normal) distribution [see Mardia (1972), Mardia and Jupp (2000)] is the most common specification for the univariate case. It is extensively studied and inference techniques are well-developed, but a multivariate extension is still a work in progress. A few recent papers [Mardia, Taylor and Subramaniam (2007), Mardia et al. (2008)] report applications of bivariate and trivariate von Mises distributions, with procedures that are complex and computationally intensive. However, for usual spatial settings, multivariate distributions of dimension 50, 100 or more arise.

For the wrapping approach, let Y be a random variable on $\mathbb{R}$, henceforth referred to as a *linear* random or unwrapped variable, with probability density function $g(y)$ and distribution function $G(y)$. The induced wrapped variable (X) of period 2π is given by

$$X = Y \bmod 2\pi. \tag{1}$$

Evidently, $0 \le X < 2\pi$. The associated *directional* probability density function $f(x)$ is obtained by wrapping $g(y)$, via the transformation $Y{=}X + 2K\pi$, around a circle of unit radius, with K being the *winding number*. It takes the form of a doubly infinite sum,

$$f(x) = \sum_{k=-\infty}^{\infty} g(x + 2k\pi), \qquad 0 \le x < 2\pi. \tag{2}$$

From (2), we see that the joint distribution of (X, K) is $g(x + 2k\pi)$ with $x \in [0, 2\pi)$ and $k \in \mathbb{Z} \equiv \{0, \pm 1, \pm 2, \ldots\}$. That is, $Y \Leftrightarrow (X, K)$; Y determines (X, K) and vice versa. Then, marginalization over k produces (2). From this joint distribution the marginal distribution of K is $P(K = k) = \int_0^{2\pi} g(x + 2k\pi)\, dx$. Additionally, $K|X = x$ is such that $P(K = k|X = x) = g(x + 2k\pi)/\sum_{j=-\infty}^{\infty} g(x + 2j\pi)$ while the conditional distribution of $X|K = k$ is $g(x + 2k\pi)/\int_0^{2\pi} g(x + 2k\pi)\, dx$. Hence, the wrapped distributions are easy to work with, treating K as a latent variable. In the context of simulation-based model fitting, sampling the full conditional distribution for K will be required. The end of Section 2.1 provides an automatic truncation approximation to facilitate such sampling.

Expectations under $f(x)$ are generally difficult to calculate; a more convenient variable to work with is the associated complex random variable on the unit circle in the complex plane, $Z = e^{iX}$. In particular, for integer p, $E(e^{ipx}) = \psi_Y(p)$, where ψ is the characteristic function of Y [Jammalamadaka and SenGupta (2001)].

2.1. *The wrapped normal distribution.* Under (2), the wrapped normal distribution arises with g a normal density indexed by parameter $\boldsymbol{\theta}$ consisting of mean, μ, and variance σ^2. In fact, we can envision $\mu = \tilde{\mu} + K_\mu$, where $\tilde{\mu}$ is the mean direction, a parameter on $[0, 2\pi)$ and K_μ is the associated winding number; as above, μ determines $\tilde{\mu}$. We can reparametrize σ^2 to $c = e^{-\sigma^2/2} < 1$ where c is referred to as the concentration parameter [Jammalamadaka and SenGupta (2001), pages 27–28]. We write the wrapped normal distribution of X as $\mathrm{WN}(\mu, \sigma^2)$ with the probability density function,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \exp\left\{\frac{-(x - \mu + 2k\pi)^2}{2\sigma^2}\right\}, \qquad 0 \le x < 2\pi. \tag{3}$$

From above, $E(Z^p) = \int_0^{2\pi} e^{ipx} f(x)\, dx = e^{ip\mu - p^2\sigma^2/2}$. Thus, we have

$$E(Z) = e^{-\sigma^2/2}(\cos\mu + i \sin\mu) \tag{4}$$

so that the resultant length of Z is the foregoing concentration parameter c. Furthermore, if $E\cos x = c\cos\mu = c\cos\tilde{\mu}$ and $E\sin x = c\sin\mu = c\sin\tilde{\mu}$, with $\tilde{\mu}$ the mean direction, as above, then $\tilde{\mu} = \arctan^*(E\sin X, E\cos X)$.[2]

To implement Markov chain Monte Carlo model fitting with wrapped normal distributions, we introduce K as a latent variable (Section 2.2). Hence, we will have to sample K's, one for each location, at each iteration. It is difficult to sample over the support $\{0, \pm1, \pm2, \ldots\}$. However, it is well known that (3) can be approximated with only a few terms. For instance, Mardia and Jupp (2000) comment that, for practical purposes, the density can be approximated by truncation to $k \in \{-1, 0, 1\}$ when $\sigma^2 \geq 2\pi$, while for $\sigma^2 < 2\pi$ using only $k = 0$ gives a reasonable approximation. We can be more precise. Suppose we translate from X to $X' = (X + \pi) \bmod 2\pi - \pi$ to achieve symmetric support, $[-\pi, \pi)$, with corresponding translation of μ to μ'. Then, suppressing the primes for convenience, with φ denoting the unit normal density function,

$$\int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} \frac{1}{\sigma}\varphi\left(\frac{x + 2k\pi - \mu}{\sigma}\right) dx = \sum_{k=-\infty}^{\infty} \int_{-\pi}^{\pi} \frac{1}{\sigma}\varphi\left(\frac{x + 2k\pi - \mu}{\sigma}\right) dx$$
$$= \sum_{k=-\infty}^{\infty} \int_{((2k-1)\pi-\mu)/\sigma}^{((2k+1)\pi-\mu)/\sigma} \varphi(z)\, dz.$$

Careful calculation reveals that, if $k_U = 1 + \lfloor \frac{3\sigma}{2\pi} \rfloor = -k_L$ (where $\lfloor a \rfloor$ denotes the integer nearest to a rounded toward 0), then $(2k_U + 1)\pi - \mu > 3\sigma$ and $(2k_L - 1)\pi - \mu < -3\sigma$. As a result,

$$\sum_{k=-\infty}^{\infty} \int_{((2k-1)\pi-\mu)/\sigma}^{((2k+1)2\pi-\mu)/\sigma} \varphi(z)\, dz > \sum_{k=k_L}^{k_U} \int_{((2k-1)\pi-\mu)/\sigma}^{((2k+1)\pi-\mu)/\sigma} \varphi(z)\, dz$$
$$> \int_{-3}^{3} \varphi(z)\, dz = 0.997. \tag{5}$$

Expression (5) facilitates MCMC model fitting since it allows us to determine the number of terms needed for good approximation as a function of σ, for example, if $\sigma < 2\pi/3$, then $k \in \{-1, 0, 1\}$; if $2\pi/3 \leq \sigma < 4\pi/3$, then $k \in \{-2, -1, 0, 1, 2\}$.

So, K can be large if and only if σ^2 can be large. Under a simulation-based model fitting the pair will not be well-identified unless we introduce an informative prior on σ^2. Moreover, when the WN concentration c is small (σ^2 large), it becomes difficult to discriminate the WN from the uniform circular distribution. With

[2]From Jammalamadaka and SenGupta [(2001), page 13] $\arctan^*(S, C)$ is formally defined as $\arctan(S/C)$ if $C > 0$, $S \geq 0$; $\pi/2$ if $C = 0$, $S > 0$; $\arctan(S/C) + \pi$ if $C < 0$; $\arctan(S/C) + 2\pi$ if $C \geq 0$, $S < 0$; undefined if $C = 0$, $S = 0$.

simulation experiments generating 1000 samples from WN's using several combinations of sample sizes and variance values, uniformity tests such as Rayleigh, Kuiper–Watson and Rao fail to discriminate between the WN and the uniform distribution for $\sigma^2 = 3.252$ with small sample sizes ($n = 30$), for $\sigma^2 = 4.02$ when $n = 100$ and $\sigma^2 = 7.11$ when $n = 1000$. Hence, it will be worthwhile to employ exploratory data analysis, for example, through the foregoing tests (available, e.g., in the CircStats library of R), with moments estimators for $\tilde{\mu}$ and σ^2, in order to assess whether to model using a WN. We recall these moments estimators: with angular data $x_1, \ldots, x_n$ and the foregoing notation, let $\bar{C} = \frac{1}{n}\sum_{i=1}^n \cos x_i$ and $\bar{S} = \frac{1}{n}\sum_{i=1}^n \sin x_i$. Setting $\bar{C} = \hat{c}\cos\hat{\tilde{\mu}}$ and $\bar{S} = \hat{c}\sin\hat{\tilde{\mu}}$, we obtain moments estimators for c and $\tilde{\mu}$ as $e^{-\hat{\sigma}^2/2} = \hat{c} = \sqrt{\bar{C}^2 + \bar{S}^2}$ and $\hat{\tilde{\mu}} = \arctan^*(\bar{S}, \bar{C})$.

2.2. *Model fitting within a Bayesian framework.* For data $\{x_1, x_2, \ldots, x_n\}$, as suggested above, it is easiest to write the full Bayesian model in terms of the joint distribution $\{(X_i, K_i), i = 1, 2, \ldots, n\}$ given μ, σ^2 with a prior on μ and σ^2, that is, as $\Pi_i \frac{1}{\sigma}\varphi((x_i + 2\pi k_i - \mu)/\sigma)[\mu, \sigma^2]$. Hence, the posterior for this model involves the latent $\{K_i\}$ (only employed to facilitate model fitting) as well as μ and σ^2. The K_i's will be updated in implementing a Gibbs sampler, but only the posterior samples of μ (which will provide posterior samples of $\tilde{\mu}$) and σ^2 will be of interest. To update the K_i, since we are given μ and σ^2, we can use (5) to implement *adaptive* truncation, that is, we can take $m = 1 + \lfloor \frac{3\sigma}{2\pi} \rfloor$ and $k = \{-m, \ldots, 0, \ldots, m\}$. Then,

$$\Pr(K_i = k_i | \mu, \sigma, x_i) \approx \frac{\varphi((x_i + 2k_i\pi - \mu)/\sigma)}{\sum_{k_i=-m}^{m} \varphi((x_i + 2k_i\pi - \mu)/\sigma)} \tag{6}$$
$$k_i = -m, \ldots, 0, \ldots, m.$$

Thus, we achieve constant maximum approximation error at each iteration.

The discussion of the previous section helps in prior specification. First, as it is customary, we assume μ and σ^2 are independent. For μ we would adopt a normal distribution, say, something like $N(\mu_0, \sigma_0^2)$. Recalling that $\mu = \tilde{\mu} + K_\mu$, this induces a WN prior on $\tilde{\mu}$ but also makes it clear that we cannot learn about μ from the X_i's, that is, in (6), we cannot identify the k_i's and k_μ, hence the k_i's and μ. However, we can learn about $\tilde{\mu}$. Furthermore, due to the conjugacy, we obtain a familiar normal for the full conditional for μ, that is, $N(\frac{\sigma_0^2\sum_i(x_i+2\pi k_i)+\sigma^2\mu_0}{n\sigma_0^2+\sigma^2}, \frac{\sigma^2\sigma_0^2}{\sigma^2+n\sigma_0^2})$. For σ^2, from the previous section, we suggest a right truncated inverse Gamma with known scale β_0 and shape α_0 and truncation defined according to $\hat{\sigma}^2$ and n, as in Section 2.1. For example, if the sample size is $n = 30$, the inverse gamma can be right truncated at π. Then, the full conditional for σ^2 will be a right truncated inverse gamma with shape parameter $\alpha_0 + n/2$ and scale parameter $\beta_0 + \frac{1}{2}[\sum_{i=1}^n (x_i + 2k_i\pi - \mu)^2]$. With such priors, the k_i are still updated as in (6).

The MCMC is straightforward, though convergence for μ and the K_i's will not be achievable due to the identifiability problem. However, we can perform usual convergence diagnostics on the $\mu + 2\pi K_i$ on $\tilde{\mu}$ and on σ^2's, all of which are well identified. Then, the posterior samples of σ^2 inform about posterior features for σ^2, and hence for c. The posterior samples of μ yield posterior samples of $\tilde{\mu}$; for the latter, we can adopt whatever posterior centrality summary we wish. Attractively, we can directly create a $1-\alpha$ credible set, that is, a symmetric posterior credible arc [Fisher (1993)]. This is merely the arc that contains the central $1-\alpha$ proportion of the posterior samples.

3. Wrapped Gaussian processes. Here, we show how a Gaussian process model for linear spatial data induces a spatial process model for wrapped data. We examine some of the properties of the induced wrapped process, in particular, the induced covariance structure. We discuss model fitting for directional data obtained at a collection of spatial locations using this process. Again, we adopt a hierarchical modeling approach, describing model fitting within the Bayesian framework using MCMC. We briefly look at regression in the context of wrapped periodic data. Finally, we show how to implement Bayesian kriging, that is, spatial prediction, within this framework. As with Bayesian kriging in the context of usual Gaussian processes, we are able to implement such prediction as a post-model fitting activity.

3.1. *Multivariate wrapped distributions*. There is surprisingly little literature on multivariate directional data modeling. Bivariate circular distributions (whence the support is a torus) are discussed in Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001). In Kato, Shimizu and Shieh (2008) bivariate circular Cauchy distributions are considered in the circular–circular regression framework (Section 3.5). In this setting, there is effort to define a sensible measure of correlation between such pairs and to test for independence for such pairs (Section 3.3). We shall see that things simplify when we work with wrapped normal distributions. Our multivariate motivation is a setting where the directional data are wave directions at locations and there is anticipated spatial dependence between the angular variables.

As a general strategy, it is perhaps easiest to obtain a multivariate wrapped distribution for, say, $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ starting with a multivariate linear distribution for $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$. In particular, suppose $\mathbf{Y} \sim g(\cdot)$, where $g(\cdot)$ is a n-variate distribution on $\mathbb{R}^n$. Usually, g is a family of distributions indexed by, say, $\boldsymbol{\theta}$; a convenient choice for $g(\cdot)$ is an n-variate normal distribution. Let $\mathbf{K} = (K_1, K_2, \ldots, K_n)$ be such that $\mathbf{Y} = \mathbf{X} + 2\pi\mathbf{K}$, analogous to the univariate case. Then, the joint distribution of $\mathbf{X}$ and $\mathbf{K}$ is $g(\mathbf{x} + 2\pi\mathbf{k})$ for $0 \le x_j < 2\pi$, $j = 1, 2, \ldots, n$ and $k_j \in \mathbb{Z}$, $j = 1, 2, \ldots, n$. The marginal distribution of $\mathbf{X}$ is an n-fold doubly infinite sum of $g(\mathbf{x} + 2\pi\mathbf{k})$ over $\mathbb{Z}^n$. Its form is intractable to work with,

even for moderate p. Again, we introduce latent K_j's to facilitate the model fitting [see Coles (1998) in this regard and Section 3.3 below].

We say that $\mathbf{X}$ has a p-variate wrapped normal distribution when $g(\cdot;\boldsymbol{\theta})$ is a multivariate normal where $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu}$ an $n \times 1$ vector of mean directions and Σ a positive definite matrix. Using standard results, the conditional distribution of Y_j given $\{Y_l, l \neq j\}$ and $\boldsymbol{\theta}$ is immediate, hence, as well, the distribution of X_j, K_j given $\{X_l, K_l, l \neq j\}$ and $\boldsymbol{\theta}$.

3.2. *Wrapped spatial Gaussian processes.* A Gaussian process on $\mathbb{R}^d$ induces a wrapped Gaussian process on $\mathbb{R}^d$. In particular, the Gaussian process (GP) is specified through its finite dimensional distributions which in turn induce the finite dimensional distributions for the wrapped process. Hence, we are returned to the multivariate wrapped distributional models of the previous subsection. In particular, if $s \in \mathbb{R}^d$ and $Y(s)$ is a GP with mean $\mu(s)$ and covariance function, say, $\sigma^2\rho(s - s';\phi)$, where ϕ is a decay parameter, then, for locations $s_1, s_2, \ldots, s_n$, $\mathbf{X} = (X(s_1), X(s_2), \ldots, X(s_n)) \sim \mathrm{WN}(\boldsymbol{\mu}, \sigma^2\mathbf{R}(\phi))$, where $\boldsymbol{\mu} = (\mu(s_1), \ldots, \mu(s_n))$ and $R(\phi)_{ij} = \rho(s_i - s_j;\phi)$. In the sequel we utilize a stationary, in fact, isotropic covariance function but with regard to the model fitting (see below); other choices could be investigated similarly. We note that the multivariate wrapped modeling in Coles (1998) employs replications to learn about a general $\boldsymbol{\Sigma}$ for the multivariate model. We do not require replications due to the structured spatial dependence introduced through the GP. Also, to implement a spatial regression model for angular response with linear covariates, it is necessary to introduce a monotone link function $h(\cdot)$ from R^1 to $(-\pi, \pi)$ with $h(0) = 0$, for example, $h(\cdot) = \arctan(\cdot)$ [Lee (2010)]. In the sequel, we confine ourselves to the case where $\mu(s) = \mu$.

3.3. *Fitting a wrapped GP model.* Model fitting for a wrapped GP within a Bayesian framework can be done using MCMC. First, suppose a linear GP model of the form $Y(s_i) = \mu + w(s_i), i = 1, 2, \ldots, n$, where $w(s_i)$ is a mean 0 GP with covariance function $\sigma^2\rho(s - s';\phi)$. Consider an exponential covariance and a prior on $\boldsymbol{\theta} = (\mu, \sigma^2, \phi)$ of the form $[\mu][\sigma^2][\phi]$ which is normal, inverse Gamma and uniform, respectively. Because of the well-known identifiability issue with σ^2 and ϕ [Zhang (2004)], it is often best to update them as a pair on the log scale using a Metropolis–Hastings step with a bivariate normal proposal, following the implementation scheme underlying the R package spBayes [Finley, Banerjee and Carlin (2007)]. For the wrapped GP, the approach follows that of Section 2 by introducing a latent vector of K's. Again, the induced wrapped GP provides $\mathbf{X} \sim \mathrm{WN}(\mu\mathbf{1}, \sigma^2\mathbf{R}(\phi))$, where $R(\phi)_{jk} = \rho(s_j - s_k;\phi)$ so that the joint distribution of $\mathbf{X}, \mathbf{K}$ takes the form $N(\mathbf{X} + 2\pi\mathbf{K}|\mu\mathbf{1}, \sigma^2\mathbf{R}(\phi))$. From above, we suggest a normal prior for μ, a truncated inverse gamma prior for σ^2 and, for the decay parameter, we have employed a uniform prior with support allowing small ranges up to ranges a bit larger than the maximum distance over the region. The

full conditionals for μ and σ^2 are similar to those in Section 2. The full conditional for ϕ is unpleasant because it is buried in the covariance matrix associated with the n-variate normal distribution for $\mathbf{X} + 2\pi\mathbf{K}$. Again, we have found it best to update σ^2 and ϕ as a pair on the log scale as in the linear case. In fact, with such joint sampling, very large values of σ^2 are rejected in the M–H step, so, in fact, we do not need to impose any truncation on the prior for σ^2.

Finally, the full conditionals for the K_i arise from the conditional distribution of $Y_i = X_i + 2\pi k_i | \{Y_j = X_j + 2\pi K_j, j \neq i\}; \boldsymbol{\theta}$. The form is analogous to (6) with μ and σ^2 replaced by the conditional mean and variance μ_i and σ_i^2 which are functions of $\{X_j, j \neq i\}$, $\{K_j, j \neq i\}$ and θ. The adaptive approximation of Section 2.1 can be employed.

Moments estimates associated with the wrapped Gaussian process are useful for the same two purposes as in the independence setting. One is to help specify priors to facilitate inference stability, following the discussion of Section 2. The other is to provide a sensible range for starting values to begin the MCMC model fitting. Under normality, again, $E(e^{iX(s)}) = \exp(i\mu - \sigma^2/2)$. So, moment estimators for mean direction, uncertainty and concentration are as in the independence case.

3.4. *Induced correlation for the circular variables.* We have defined the wrapped GP in terms of the linear GP which has a covariance or correlation structure (and the parameters which specify them). For a bivariate circular variable there is no unique choice of correlation measure. However, in Jammalamadaka and SenGupta [(2001), Chapter 8], we find considerable discussion of suitable measures for the correlation between two circular variables. Jammalamadaka and Sarma (1988) propose a measure which satisfies many of the properties of the product moment correlation coefficient. For the wrapped bivariate normal for variables X_1, X_2 with covariance matrix $\left(\begin{smallmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{smallmatrix}\right)$ it simplifies to $\rho_c(X_1, X_2) = \sinh(\rho\sigma^2)/\sinh(\sigma^2)$. Hence, with a valid covariance function $\sigma^2\rho(s, s')$, the induced correlation function for the wrapped Gaussian process is $\rho_c(s, s') = \frac{\sinh(\sigma^2\rho(s,s'))}{\sinh(\sigma^2)}$. Figure 1 provides a picture of the exponential correlation function for various choices of the decay parameter ϕ and the corresponding correlation for the wrapped process. For a given distance, association is similar but a bit weaker under the latter.

As an aside, though we don't build any covariance matrices using $\sinh(\rho(u))$, we note that it is a valid covariance function on $\mathbb{R}^d$ if $\sigma^2\rho(u)$ is. This is evident since $\sinh(v) = \sum_{n=1}^{\infty} \frac{v^{2n+1}}{(2n+1)!}$. Because $\rho(u)$ is a valid correlation function, hence a characteristic function by Bochner's theorem, $\rho^p(u)$ is a characteristic function for any integer $p > 0$. So, $\sinh(\rho(u))$ is a mixture (with suitable normalization) of characteristic functions, hence a characteristic function, therefore, again, by Bochner's theorem, a valid covariance function itself.

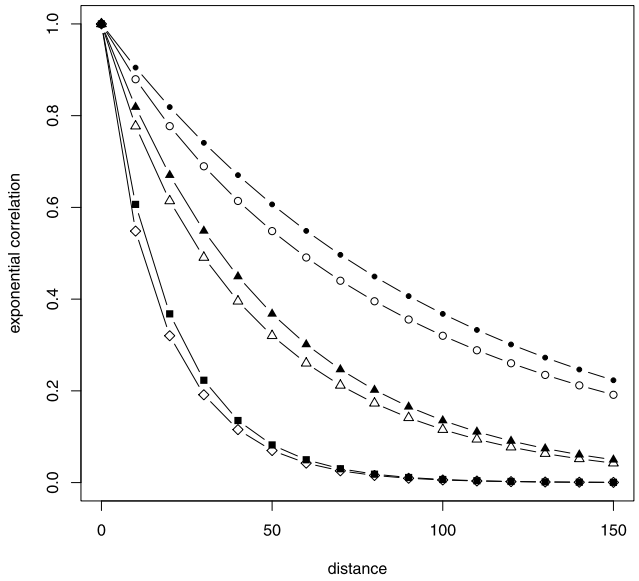


FIG. 1. *Exponential spatial correlation function* (*solid circle* $\phi = 0.01$, *solid triangle* $\phi = 0.02$, *solid square* $\phi = 0.05$) *and the corresponding wrapped correlation* (*empty circle* $\phi = 0.01$, *empty triangle* $\phi = 0.02$, *empty square* $\phi = 0.05$) *for* 3 *values of the decay parameter* ϕ.

3.5. *Circular–circular regression with WN models*. For general bivariate directional data models, say, $g(X_1, X_2)$, circular–circular regression for, say, X_1 given X_2, is discussed in Jammalamadaka and Sarma (1993) using trigonometric polynomial approximations. In the case of wrapped bivariate normal distributions, we can obtain the regression explicitly, avoiding approximation. We also note that Kato, Shimizu and Shieh (2008) consider circular–circular regression curves obtained under a Möbius circle transformation [See also Downs and Mardia (2002) in this regard]. Incidentally, they discuss a bivariate circular distribution which is specified through the product of a conditional times a marginal wrapped Cauchy distribution.

We return to the wrapped bivariate normal for variables X_1, X_2 with mean μ_1, μ_2 and covariance matrix $\left(\begin{smallmatrix}\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2\end{smallmatrix}\right)$. The regression model we seek is $E(e^{iX_2}|X_1;\boldsymbol{\theta})$ in the form $c(X_1;\boldsymbol{\theta})e^{i\mu(X_1;\boldsymbol{\theta})}$, where $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma^2, \rho)$. Then, we can interpret, respectively, $\mu(X_1;\boldsymbol{\theta})$ as the conditional mean direction and $c(X_1;\boldsymbol{\theta})$ as the conditional concentration for X_2 given X_1.

For the associated unwrapped normal distribution,

$$Y_2|Y_1;\boldsymbol{\theta} \sim N\big(\mu_{2|1}(Y_1;\boldsymbol{\theta}), \sigma^2_{2|1}(\boldsymbol{\theta})\big),$$

where $\mu_{2|1}(Y_1;\boldsymbol{\theta}) = \mu_2 + \rho(Y_1 - \mu_1)$ and $\sigma^2_{2|1}(\boldsymbol{\theta}) = \sigma^2(1-\rho^2)$. Hence, writing $Y_1 = X_1 + 2\pi K_1$, we have $X_2|X_1, K_1;\boldsymbol{\theta} \sim \mathrm{WN}(\tilde{\mu}_{2|1}(X_1 + 2\pi K_1);\boldsymbol{\theta}), \sigma^2_{2|1}(\boldsymbol{\theta}))$,

where $\tilde{\mu}(Y) = \mu(Y) \bmod 2\pi$ and, therefore,

$$E\big(e^{iX_2}|X_1, K_1; \boldsymbol{\theta}\big) = \exp\big(-\sigma^2_{2|1}(\boldsymbol{\theta})/2 + i\tilde{\mu}_{2|1}(X_1 + 2\pi K_1; \boldsymbol{\theta})\big).$$

Next, we have $E(e^{iX_2}|X_1; \boldsymbol{\theta}) = E_{K_1|X_1;\boldsymbol{\theta}} E(e^{iX_2}|X_1, K_1; \boldsymbol{\theta})$, where the conditional distribution for $K|X, \boldsymbol{\theta}$ under the wrapped normal is discussed in Section 2. So, let $p(k; X_1, \boldsymbol{\theta}) = P(K_1 = k|X_1, \boldsymbol{\theta})$. Then,

$$\begin{aligned} E\big(e^{iX_2}|X_1; \boldsymbol{\theta}\big) &= \exp\big(-\sigma^2_{2|1}(\boldsymbol{\theta})/2\big) \\ &\quad \times \sum_{k=-\infty}^{\infty} p(k; X_1, \boldsymbol{\theta}) \exp\big(i\tilde{\mu}_{2|1}(X_1 + 2\pi k; \boldsymbol{\theta})\big). \end{aligned} \tag{7}$$

We see that $c(X_1; \boldsymbol{\theta}) = e^{-\sigma^2_{2|1}(\boldsymbol{\theta})/2}$ and $\exp(i\mu(X_1; \boldsymbol{\theta})) = \sum_{k=-\infty}^{\infty} p(k; X_1, \boldsymbol{\theta}) \times \exp(i\tilde{\mu}_{2|1}(X_1 + 2\pi k; \boldsymbol{\theta}))$. So, $\cos(\mu(X_1; \boldsymbol{\theta})) = \sum_{k=-\infty}^{\infty} p(k; X_1, \boldsymbol{\theta}) \cos(\tilde{\mu}_{2|1}(X_1 + 2\pi k; \boldsymbol{\theta}))$ and $\sin(\mu(X_1; \boldsymbol{\theta})) = \sum_{k=-\infty}^{\infty} p(k; X_1, \boldsymbol{\theta}) \sin(\tilde{\mu}_{2|1}(X_1 + 2\pi k; \boldsymbol{\theta}))$. Making the usual inversion,

$$\mu(X_1; \boldsymbol{\theta}) \equiv \arctan^*\big(\sin\big(\mu(X_1; \boldsymbol{\theta})\big), \cos\big(\mu(X_1; \boldsymbol{\theta})\big)\big). \tag{8}$$

In practice, we would compute $\mu(X_1; \boldsymbol{\theta})$ by appropriate truncation of K. If we fit the bivariate wrapped normal distribution with data (X_{1i}, X_{2i}), $i = 1, 2, \ldots, n$, using MCMC, posterior samples for $\boldsymbol{\theta}$ enable posterior samples for $\mu(X_1; \boldsymbol{\theta})$ and $c(X_1; \boldsymbol{\theta})$ at any X_1.

3.6. *Kriging with wrapped GP models.* Kriging is a customary activity with spatial data. In this context, we would have, as observations, $\mathbf{X} = (X(s_1), X(s_2), \ldots, X(s_p))$ and we would seek to predict $X(s_0)$ at a new location s_0. In fact, we shall argue that this is a straightforward post-model fitting exercise and can be implemented following the ideas of circular–circular regression for the wrapped normal from the previous subsection.

Suppose, for the linear "observations," $\mathbf{Y} = (Y(s_1), Y(s_2), \ldots, Y(s_p))$ along with $Y(s_0)$ we have the joint distribution

$$\begin{pmatrix} \mathbf{Y} \\ Y(s_0) \end{pmatrix} = N\left(\begin{pmatrix} \boldsymbol{\mu} \\ \mu(s_0) \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{R}_{\mathbf{Y}}(\phi) & \boldsymbol{\rho}_{0,\mathbf{Y}}(\phi) \\ \boldsymbol{\rho}^T_{0,\mathbf{Y}}(\phi) & 1 \end{pmatrix}\right). \tag{9}$$

Evidently, we can obtain the distribution for $Y(s_0)|\mathbf{Y}, \boldsymbol{\theta}$, hence the wrapped normal distribution for $X(s_0)|\mathbf{X}, \mathbf{K}, \boldsymbol{\theta}$ and, thus, $E(e^{iX(s_0)}|\mathbf{X}, \mathbf{K}, \boldsymbol{\theta})$, following the previous section. Now, for $E(e^{iX(s_0)}|\mathbf{X}, \boldsymbol{\theta})$ we would need to marginalize over the distribution of $\mathbf{K}|\mathbf{X}, \boldsymbol{\theta}$. This requires a p-fold sum over a multivariate discrete distribution, hopeless for large p even with considerable truncation, for example, a sum over 3^p terms if each K_i is allowed only 3 values.

In fact, within the Bayesian modeling framework we seek $E(e^{iX(s_0)}|\mathbf{X})$. Fitting the spatial wrapped GP model as in Section 3.4 will yield posterior samples, say, $(\boldsymbol{\theta}^*_b, \mathbf{K}^*_b)$, $b = 1, 2, \ldots, B$. Then, as usual, $E(e^{iX(s_0)}|\mathbf{X}) =$

$E_{\mathbf{K},\boldsymbol{\theta}|\mathbf{X}} E(e^{iX(s_0)}|\mathbf{X}, \mathbf{K}, \boldsymbol{\theta})$ and so a Monte Carlo integration yields

$$E\big(e^{iX(s_0)}|\mathbf{X}\big) \approx \frac{1}{B}\sum_b \exp\big(-\sigma^2(s_0, \boldsymbol{\theta}_b^*)/2 + i\tilde{\mu}\big(s_0, \mathbf{X} + 2\pi\mathbf{K}_b^*; \boldsymbol{\theta}_b^*\big)\big), \tag{10}$$

where $\tilde{\mu}(s_0, \mathbf{Y}, \boldsymbol{\theta})$ and $\sigma^2(s_0, \boldsymbol{\theta})$ extend the notation $\tilde{\mu}_{2|1}$ and $\sigma^2_{2|1}$ of the previous section to $Y(s_0)|\mathbf{Y}$. If $g_c(s_0, \mathbf{X}) = \frac{1}{B}\sum_{b^*} \exp(-\sigma^2(s_0, \boldsymbol{\theta}_b^*)/2)\cos(\tilde{\mu}(s_0, \mathbf{X} + 2\pi\mathbf{K}_b^*; \boldsymbol{\theta}_b^*))$ and $g_s(s_0, \mathbf{X}) = \frac{1}{B}\sum_{b^*} \exp(-\sigma^2(s_0, \boldsymbol{\theta}_b^*)/2)\sin(\tilde{\mu}(s_0, \mathbf{X}+2\pi\mathbf{K}_b^*; \boldsymbol{\theta}_b^*))$, then the posterior mean kriged direction is

$$\mu(s_0, \mathbf{X}) = \arctan^*\big(g_{0,s}(\mathbf{X}), g_{0,c}(\mathbf{X})\big) \tag{11}$$

and the associated posterior kriged concentration is

$$c(s_0, \mathbf{X}) = \sqrt{\big(g_c(s_0, \mathbf{X})\big)^2 + \big(g_s(s_0, \mathbf{X})\big)^2}. \tag{12}$$

4. Examples and MCMC implementation. In Section 4.1 we offer some simulation examples, while in Section 4.2 we turn to the motivating spatial wave direction data. We fit the model described in Section 3.2 and implemented the kriging following Section 3.6. We note that MCMC model fitting as described in Section 3.4 is well behaved.

4.1. *Simulation examples*. In the simulations we generate samples of size $n = 100$ from an unwrapped GP with constant mean μ and we obtain the directional process by wrapping them onto the unit circle. Locations are generated uniformly using coordinates taken from the real data example of Section 4.2; we work with two different sample sizes by randomly choosing 30 and 70 sites for posterior estimation and the remaining are used for validation. We fix the covariance structure for the linear GP to be exponential with parameters (σ^2, ϕ). Here we report examples generated with $\phi = 0.021$ corresponding to a practical range of 142.86 km (maximum distance spanned by the coordinates is 290.13 km), $\mu = \pi$ and three different variances $\sigma^2 = 0.1, 0.5, 1$ corresponding to concentrations $c = 0.951, 0.779, 0.606$ respectively. The prior for μ is a Gaussian distribution with zero mean and large variance (with an induced wrapped normal prior for $\tilde{\mu}$). For σ^2 we use informative inverse gamma distributions centered on the *true* value and variances 0.01, 0.06 and 0.07, respectively. For the decay parameter the prior is a uniform distribution in $[0.001, 1)$ when $\sigma^2 = 0.1, 0.5$ and in $[0.001, 0.5)$ when $\sigma^2 = 1$. For all priors settings, several variance values were considered to assess behavior under strongly and weakly informative priors. The block sampling of σ^2 and ϕ produces strongly autocorrelated chains for both parameters and slow convergence compared to the independence case. We run the MCMC for 30,000 iterations, we discard the initial 6000 and we apply a thinning of 10, using 2400 samples for estimation.

The precision of estimates is studied by 95% credible intervals, obtained as discussed in Section 2.2. We compute kriging estimates as in (11) and (12) for each specification and compute an average prediction error, defined as the average circular distance[3] between the observed and kriged estimate, over the validation set of observed values. That is, for the validation set $\{x(s_j^*), j = 1, 2, \ldots, m\}$, we compute $\frac{1}{m}\sum_l (1 - \cos(\mu(s_j, \mathbf{X}) - x(s_j^*)))$. We also compare this spatial interpolation to prediction obtained by fitting the independent wrapped normal model proposed in Section 2 and the prior structure described there. Comparison is assessed through average error, computed for the nonspatial model as the average circular distance between the directions in the validation set and the estimated mean direction of the nonspatial model.

In Table 1 posterior estimates and interpolation errors are given for both the spatial and nonspatial models. Both models can recover the mean direction and concentration with wider confidence intervals when c is small. The decay parameter is correctly estimated when $c = 0.951, 0.779$ for both sample sizes but requires larger sample size when $c = 0.606$. There is substantial reduction in average prediction error using the wrapped GP model when there is spatial dependence, while for small ranges (not shown) the spatial model performs comparably to the nonspatial one.

4.2. *Wave direction data analysis*. The majority of studies carried out on marine data are based on outputs from deterministic models, usually climatic forecasts computed at several spatial and temporal resolutions. Since the 1980s, deterministic models have been used for weather forecasting with increasing reliability. Moreover, in the last decade, sea surface wind data projections, produced by meteorological models, have been found accurate enough to be taken as the basis for operational marine forecasts. Wave heights and outgoing wave directions, the latter being angular data measured in degrees, are the main outputs of marine forecasts. The principal provider for global numerical wave forecasts in Europe is the European Center for Medium-Range Weather Forecasts (ECMWF), which runs at global medium range (3–5 up to 10 day forecasts, 55 km spatial resolution) and at high resolution short term (3 days, 28 km resolution WAve Model, WAM) models in the Mediterranean Area. WAM outputs are given in deep waters (more than 100 m depth) on a grid with about 25×25 km cells.

Calibration with buoy data for wave heights was studied [Jona Lasinio et al. (2007)] using a multistep model that included outgoing wave directions as a categorical variable (categorized according to an 8 sector windrose) not allowing for a full exploitation of the available information. The main use of these calibrated data is in starting of runs of shallow water forecast models [SWAN model, see,

[3]We adopt as circular distance, $d(\alpha, \beta) = 1 - \cos(\alpha - \beta)$, as suggested in Jammalamadaka and SenGupta [(2001), page 16].

TABLE 1

Results from simulated data with $\mu = \pi$ and $\phi = 0.013$ for all simulations. Posterior mean and concentration point estimates are obtained averaging the MCMC samples, while the decay point estimate is obtained as modal value

	$\hat{\mu}$ (95% CI)	$\hat{c}$ (95% CI)	$\hat{\phi}$ (95% CI)	Average prediction error
$c = 0.951$				
$n = 30$	3.124	0.926	0.013	0.034
	(2.716, 3.517)	(0.825, 0.965)	(0.005, 0.066)	
$n = 70$	3.135	0.929	0.017	0.023
	(2.800, 3.505)	(0.848, 0.962)	(0.007, 0.037)	
$c = 0.779$				
$n = 30$	3.164	0.708	0.018	0.058
	(2.399, 3.777)	(0.547, 0.811)	(0.009, 0.329)	
$n = 70$	3.229	0.748	0.015	0.085
	(2.518, 3.898)	(0.596, 0.841)	(0.008, 0.032)	
$c = 0.606$				
$n = 30$	2.916	0.594	0.049	0.188
	(2.416, 3.514)	(0.470, 0.693)	(0.022, 0.190)	
$n = 70$	2.928	0.608	0.025	0.099
	(2.261, 3.673)	(0.480, 0.706)	(0.015, 0.048)	
Non spatial model				
$c = 0.951$				
$n = 30$	3.102	0.947		0.048
	(2.981, 3.223)	(0.914, 0.968)		
$n = 70$	3.110	0.948		0.042
	(3.034, 3.188)	(0.928, 0.962)		
$c = 0.779$				
$n = 30$	2.781	0.794		0.240
	(2.535, 3.030)	(0.680, 0.871)		
$n = 70$	2.925	0.749		0.170
	(2.710, 3.140)	(0.649, 0.823)		
$c = 0.606$				
$n = 30$	2.869	0.640		0.335
	(2.514, 3.217)	(0.473, 0.765)		
$n = 70$	2.785	0.677		0.382
	(2.577, 3.001)	(0.578, 0.755)		

e.g., Holthuijsen (2007)]. The output of these models provide the basis for coastal alarm in severe weather and for the evaluation of coastal erosion. These models are built on grids with about 10 km spatial resolution and they include as inputs calibrated and downscaled wave heights, bathymetry, wind speed and direction all given at the same spatial resolution. Wave directions are aligned to the finer grid

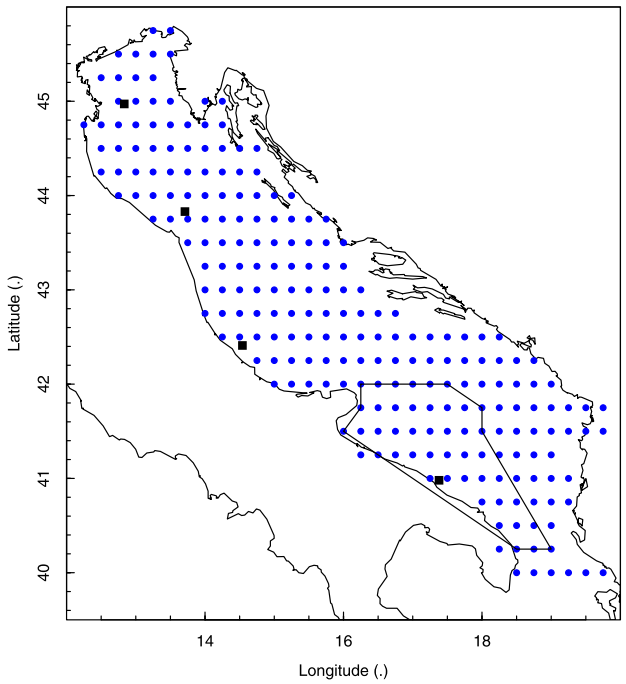


FIG. 2. *WAM grid, Italian Wave Monitoring network in the Adriatic Sea and estimation area, squares denote locations of buoys.*

simply by taking the circular mean of intersected WAM cells. Here we propose a fully model-based solution using a downscaling procedure.

We analyze data from a single time during a storm in the Adriatic Sea. The data are outgoing wave directions produced by the WAM during the analysis, that is, a run of the forecast of the model at time $t = 0$. We use 45 points from the WAM grid, as they cover a fairly homogeneous area during the storm movement. In Figure 2 the entire Adriatic network of buoys is shown together with the WAM grid and the estimation area (delimited by continuous lines). Available data yield a moments estimated sample concentration of $\hat{c} = 0.8447$, in the North–East direction (moments estimated sample mean direction $\hat{\hat{\mu}} = 0.5540$). In the data only five values differ from $\pi/4$ by more than 0.8 radians [they are marked on Figure 4(b)], four are located in the upper north part of the area outside the Gargano Peninsula, one is located near the coast inside the Mattinata Gulf. This last value, being located in the curve of the gulf, may account for a local wave direction inversion. The other four seem to describe some turbulence in the wave field. We keep them in the data set to see how the spatial model deals with this local behavior. We are interested in downscaling these *WAM* values to a grid of 222 cells with 10 km resolution.

Arguably, an angular data model such as a multivariate version of the von Mises distribution might be more natural here than a periodic data model like our wrapped normal. However, because the WN allows convenient extension to the spatial setting and because there is little practical difference between the WN and von Mises on the circle, we work with the latter, as detailed in Section 3. In the absence of covariates, we employ the constant mean direction model (though a trend surface could be investigated). In fitting, we run the MCMC algorithm for 30,000 iterations, with a burn-in of 6000, and compute kriging estimates as in (11) and (12) using 2400, taking one sample every 10 samples. We choose the following prior setting: $\sigma^2 \sim \text{IG}(9, 4)$ (i.e., mode $= 0.4$ and variance $= 0.04$), $\phi \sim \text{Unif}(0.001, 0.1)$ and, for the mean, a Normal distribution with zero mean and large variance. In Figure 3 the posterior learning from the data is shown. Posterior estimates return a posterior mean direction of 0.971 radians ([0.471, 1.814] is a 95% credible interval) and posterior concentration of 0.618 ([0.427, 0.753] is 95% credible interval). The decay parameter, estimated as posterior modal value,

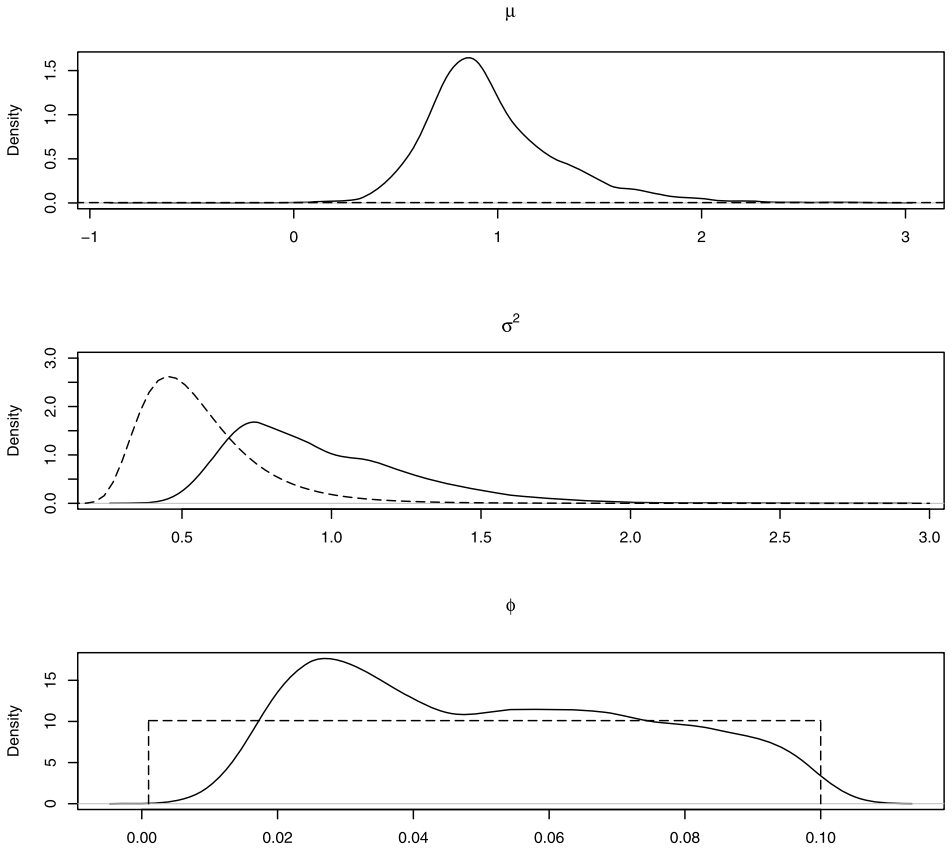
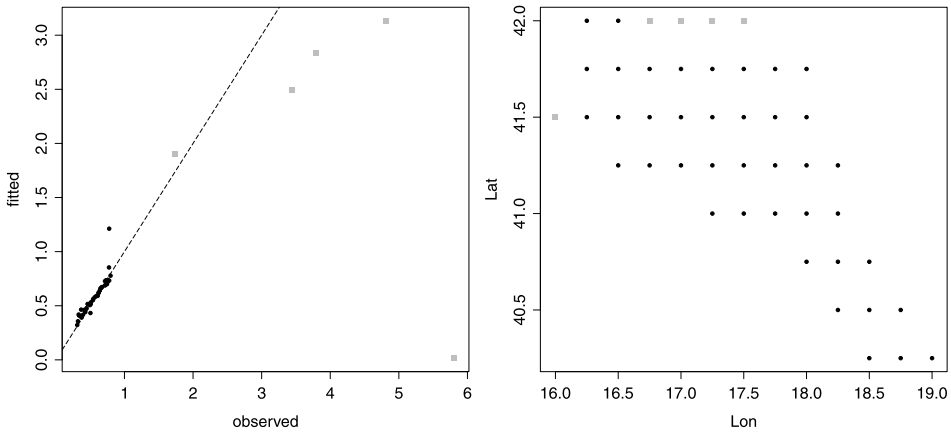


FIG. 3. *Prior* (*dashed*) *and posterior* (*solid*) *distributions for the WAM data.*



(a) (b)

FIG. 4. (a) *Kriged estimate vs. observed using leave-one-out validation*, (b) *locations of the* 5 *outliers present in the data* (*grey squares*).

is 0.023 ([0.017, 0.096] is a 95% credible interval). In Figure 4 results from a leave-one-out validation procedure are shown. It can be seen that the five outliers are shrunk toward the 45° line. The average prediction error is 0.0488.

In Figure 5 an arrow plot of incoming wave directions with length of the arrows proportional to $1 - \hat{c}$ (the longer the vector, the more variable the estimates at that grid point) is shown. In the arrow plot we choose to present the incoming direc-

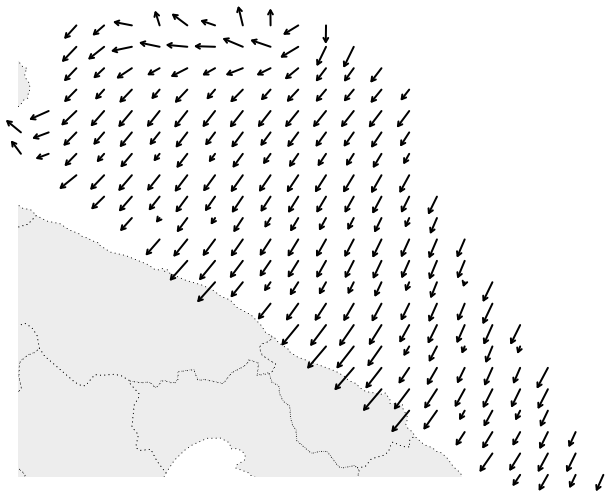


FIG. 5. *Kriging results for WAM data*: *wave directions represented as arrow*, *the length of the arrows is proportional to* 1*-concetration. Here we report incoming waves directions*.

tion[4] instead of the outgoing one to better visualize the model impact of waves on the coast during storms. From Figure 5 we can see that the model captures local features of the wave field, that arrows rotate in the upper part of the plot (following the contour of the Gargano promontory and the presence of the outliers), and that more variable estimates are obtained near the coast line (mimicking similar behavior of wave heights before calibration). Finally, we compare our results with those from the nonspatial model (which we ran for 20,000 iterations, keeping the second half for estimation, using priors as described in Section 2.2). The estimated posterior mean direction is 0.5507 ([0.3278, 0.7810] 95% C.I.) and the posterior concentration is 0.8272 ([0.7748, 0.8667] 95% C.I.). The leave-one-out validation yielded an average predictive error of 0.1553, revealing that the spatial model yielded a reduction of 68%.

5. Summary and future work. We have introduced a class of spatial process models which can be used to study circular data that is anticipated to have spatial structure. In particular, adopting the class of wrapped distributions for describing circular data, we have shown how a usual Gaussian process specification can be converted to a wrapped normal process specification. We have demonstrated that fitting for such models can be done straightforwardly within a hierarchical modeling framework using MCMC methods with adaptive truncation. We have also shown how to implement kriging for such models. Our motivating application has revealed the predictive benefit of the spatial modeling.

Particularly with directional data, one would expect there to be concerns regarding measurement error, for instance, with monitors recording wind direction and with buoys measuring wave direction. We are unaware of any measurement error work in the context of directional data but addressing it using wrapped Gaussian processes turns out to be straightforward. It will be captured by a *nugget* similar to the usual geostatistical modeling setting. In fact, a frequent interpretation of the nugget is measurement error [Banerjee, Carlin and Gelfand (2004)]. Suppose the observed angular data are collected with conditionally independent measurement error, that is, $X_{o,j} \sim \mathrm{WN}(Y_{t,j}, \tau^2)$, where $Y_{t,j} = X_{t,j} + 2\pi K_{t,j}$ and $X_{t,j}$ is the true angular direction, with independence across j. As above, we introduce latent $K_{o,j}$ and model the joint distribution of $X_{o,j}$, $K_{o,j}$ as $N(X_{o,j} + 2\pi K_{o,j} | X_{t,j}, \tau^2)$. The latent true directions follow the wrapped GP of Section 3.2. That is,

$$\mathbf{X}_t = \big(X_t(s_1), X_t(s_2), \ldots, X_t(s_p)\big) \sim \mathrm{WN}\big(\mu\mathbf{1}, \sigma^2\mathbf{R}(\phi)\big),$$

where $R(\phi)_{ij} = \rho(s_i - s_j; \phi)$. Introducing $\mathbf{K}_t$, the joint distribution for $\mathbf{X}_t, \mathbf{K}_t$ takes the form $N(\mathbf{X}_t + 2\pi\mathbf{K}_t | \mu\mathbf{1}, \sigma^2\mathbf{R}(\phi))$. The full model takes the form

$$\Pi_j\big[X_{o,j} + 2\pi K_{o,j} | X_{t,j}, \tau^2\big]\big[\mathbf{X}_t + 2\pi\mathbf{K}_t | \mu, \sigma^2\mathbf{R}(\phi)\big][\mu]\big[\sigma^2\big]\big[\tau^2\big][\phi]. \tag{13}$$

[4]Incoming wave directions are obtained with a 180° rotation of the outgoing wave directions.

An issue that requires further investigation is the matter of the assumption of stationarity in the covariance function. It would be useful to develop diagnostics to examine this, paralleling those for linear spatial data. At present, all we can suggest is model comparison using average predictive error as in Section 4.

Future work will investigate promise of extending the wrapped normal process to a wrapped t-process through the usual Gamma mixing that extends a GP to a t-process [see, e.g., Zhang, Wu and Chang (2007)]. Extension to the t-process discussed in Heyde and Leonenko (2005) would be more challenging. It will also lead us to incorporate dynamic structure into our modeling; with regard to our data set, we have wave direction information at various temporal resolutions. Finally, we will explore two data assimilation issues. The first is to fuse the angular data produced by the WAve Model (WAM) with the buoy data (RON) to improve our interpolation of wave direction. The second involves joint modeling of the wave direction data with the available associated wave height data. Hopefully, joint modeling will enable a version of co-kriging to improve the individual interpolations.

Acknowledgments. The authors thank the Coastal Defense Unit, ISPRA, Italy, for providing the data. They are grateful to Daniela Cocchi and Clarissa Ferrari for many discussions and useful suggestions on this topic. The authors wish to thank the reviewers and the editor for very useful comments and suggestions that helped to improve considerably the paper.

REFERENCES

BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall, London.

BAO, L., GNEITING, T., GRIMIT, E. P., GUTTORP, P. and RAFTERY, A. E. (2010). Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review* **138** 1811–1821.

BRECKLING, J. (1989). *The Analysis of Directional Time Series*: *Applications to Wind Speed and Direction*. *Lecture Notes in Statistics* **61**. Springer, Berlin. MR1027836

CASSON, E. and COLES, S. (1998). Extreme hurricane wind speeds: Estimation, extrapolation and spatial smoothing. *Journal of Wind Engineering and Industrial Aerodynamics* **74-76** 131–140.

COLES, S. (1998). Inference for circular distributions and processes. *Statist. Comput.* **8** 105–113.

DAMIEN, P. and WALKER, S. (1999). A full Bayesian analysis of circular data using the von Mises distribution. *Canad. J. Statist.* **27** 291–298.

DOWNS, T. D. and MARDIA, K. V. (2002). Circular regression. *Biometrika* **89** 683–697. MR1929172

ENGEL, C. and EBERT, E. (2007). Performance of hourly operational consensus forecasts (OCFs) in the Australian region. *Weather and Forecasting* **22** 1345–1359.

FINLEY, A. O., BANERJEE, S. and CARLIN, B. P. (2007). spBayes: An R package for univariate and multivariate hierarchical point-referenced spatial models. *Journal of Statistical Software* **19** 1–24.

FISHER, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge Univ. Press, Cambridge.

FISHER, N. I. and LEE, A. J. (1992). Regression models for an angular response. *Biometrics* **48** 665–677. MR1187598

FISHER, N. I. and LEE, A. J. (1994). Time series analysis of circular data. *J. Roy. Statist. Soc. Ser. B* **56** 327–339. MR1281937

GUTTORP, P. and LOCKHART, R. A. (1988). Finding the location of a signal: A Bayesian analysis. *J. Amer. Statist. Assoc.* **83** 322–330. MR0971359

HARRISON, D. and KANJI, G. K. (1988). The development of analysis of variance for circular data. *J. Appl. Stat.* **15** 197–224.

HEYDE, C. C. and LEONENKO, N. N. (2005). Student processes. *Adv. in Appl. Probab.* **37** 342–365. MR2144557

HOLTHUIJSEN, L. H. (2007). *Waves in Oceanic and Coastal Waters*. Cambridge Univ. Press, Cambridge.

HOLZMANN, H., MUNK, A., SUSTER, M. and ZUCCHINI, W. (2006). Hidden Markov models for circular and linear-circular time series. *Environ. Ecol. Stat.* **13** 325–347. MR2242193

HUGHES, G. (2007). Multivariate and time series models for circular data with applications to protein conformational angles. Ph.D. thesis, Univ. Leeds, Leeds, England, UK.

JAMMALAMADAKA, S. R. and SARMA, Y. R. (1988). A correlation coefficient for angular variables. In *Statistical Theory and Data Analysis*, *II* (*Tokyo*, 1986) 349–364. North-Holland, Amsterdam. MR0999196

JAMMALAMADAKA, S. R. and SARMA, Y. R. (1993). Circular regression. In *Statistical Sciences and Data Analysis* (K. Matusita, M. L. Puri and T. Hayakawa, eds.) 109–128. VSP, Utrecht.

JAMMALAMADAKA, S. R. and SENGUPTA, A. (2001). *Topics in Circular Statistics*. World Scientific, Singapore.

JONA LASINIO, G., ORASI, A., DIVINO, F. and CONTI, P. L. (2007). Statistical contributions to the analysis of environmental risks along the coastline. In *Società Italiana di Statistica—Rischio e Previsione. Venezia, 6–8 Giugno* 255–262. Società Italiana di Statistica, CLEUP, ISBN/ISSN: 978-88-6129-093-8.

KALNAY, E. (2002). *Athmospheric Modeling*, *Data Assimilation and Predictability*. Cambridge Univ. Press, Cambridge.

KATO, S. (2010). A Markov process for circular data. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 655–672. MR2758240

KATO, S., SHIMIZU, K. and SHIEH, G. S. (2008). A circular–circular regression model. *Statist. Sinica* **18** 633–645. MR2432283

KENT, J. (1978). Limiting behaviour of the von Mises–Fisher distribution. *Math. Proc. Cambridge Philos. Soc.* **84** 531–536. MR0501259

LEE, A. J. (2010). Circular data. *Wiley Interdisciplinary Reviews*: *Computational Statistics* **2** 477–486.

MARDIA, K. V. (1972). *Statistics of Directional Data*. Academic Press, London. MR0336854

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. Wiley, Chichester. MR1828667

MARDIA, K. V., TAYLOR, C. C. and SUBRAMANIAM, G. K. (2007). Protein bioinformatics and mixtures of bivariate von Mises distributions for angular data. *Biometrics* **63** 505–512. MR2370809

MARDIA, K. V., HUGHES, G., TAYLOR, C. C. and SINGH, H. (2008). A multivariate von Mises distribution with applications to bioinformatics. *Canad. J. Statist.* **36** 99–109. MR2432195

RAVINDRAN, P. (2002). Bayesian analysis of circular data using wrapped distributions. Ph.D. thesis, North Carolina State Univ.

STEPHENS, M. A. (1963). Random walk on a circle. *Biometrika* **50** 385–390. MR0165591

STEPHENS, M. A. (1970). Use of the Kolmogorov–Smirnov, Cramer–von Mises and related statistics without extensive tables. *J. Roy. Statist. Soc. Ser. B* **32** 115–122.

WATSON, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika* **48** 109–114. MR0131930

WILKS, D. S. (2006). Comparison of ensemble-MOS methods in the Lorenz'96 setting. *Meteorological Applications* **13** 243–256.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99** 250–261. MR2054303

ZHANG, Z., WU, G. and CHANG, E. (2007). Semiparametric regression using Student t-processes. *IEEE Transactions on Neural Networks* **18** 1572–1588.

G. JONA-LASINIO
DSS UNIVERSITY OF ROME SAPIENZA
P. LE ALDO MORO 5
00185 ROME
ITALY
E-MAIL: giovanna.jonalasinio@uniroma1.it

A. GELFAND
DUKE UNIVERSITY
223-A OLD CHEMISTRY BUILDING
BOX 90251
DURHAM, NORTH CAROLINA 27708-0251
USA
E-MAIL: alan@stat.duke.edu

M. JONA-LASINIO
ITP—LEIBNIZ UNIVERSITAT HANNOVER
APPELSTRASSE 2
30167 HANNOVER
GERMANY
E-MAIL: mattia.jonalasinio@itp.uni-hannover.de