CrossMark

ORIGINAL PAPER

Spatio-temporal circular models with non-separable covariance structure

Gianluca Mastrantonio[1] ·
Giovanna Jona Lasinio[2] · **Alan E. Gelfand**[3]

Abstract Circular data arise in many areas of application. Recently, there has been interest in looking at circular data collected separately over time and over space. Here, we extend some of this work to the spatio-temporal setting, introducing space–time dependence. We accommodate covariates, implement full kriging and forecasting, and also allow for a nugget which can be time dependent. We work within a Bayesian framework, introducing suitable latent variables to facilitate Markov chain Monte Carlo model fitting. The Bayesian framework enables us to implement full inference, obtaining predictive distributions for kriging and forecasting. We offer comparison between the less flexible but more interpretable wrapped Gaussian process and the more flexible but less interpretable projected Gaussian process. We do this illustratively using both simulated data and data from computer model output for wave directions in the Adriatic Sea off the coast of Italy.

Keywords Average prediction error · Continuous ranked probability score · Kriging · Markov chain Monte Carlo · Projected distribution · Wrapped distribution

✉ Gianluca Mastrantonio
gianluca.mastrantonio@yahoo.it; gianluca.mastrantonio@uniroma3.it

1 Roma Tre University, Via Silvio D'Amico 77, 00145 Rome, Italy

2 Sapienza University of Rome, P.le Aldo Moro 5, 00185 Rome, Italy

3 Duke University, 223-A Old Chemistry Building, Box 90251, Durham, NC 27708-0251, USA

🌐 Springer

1 Introduction

Circular data, i.e., observations with support on the unit circle, arise in many contexts. Examples include natural directions, such as wind directions (meteorology), animal movement directions (biology) and rock fracture orientations (geology). Another type of circular data arises by wrapping periodic time data with period L (say, day or week) onto a circle with circumference L and then rescaling the circumference to 2π, that of the unit circle. Two-dimensional directional data may be observed in space and time, along with linear variables, as in marine studies where, for example wave heights and directions are jointly observed, or in atmospheric modeling where wind fields are represented by wind intensity as well as direction. Due to the restriction of the domain to the circle, analysis of circular data must be treated differently from linear data. Customary statistical summaries are replaced with their circular counterparts. For a discussion of inference with circular variables see, e.g., Fisher (1996), Mardia and Jupp (1999), Jammalamadaka and SenGupta (2001) or the recent paper by Lee (2010).

The contribution of this paper is to extend recent spatial and spatio-temporal circular data models. In particular, Jona Lasinio et al. (2012) consider the use of the wrapped normal approach by developing the wrapped Gaussian process while Wang and Gelfand (2014) consider the use of the projected normal approach by developing the projected Gaussian process. Here we: (i) extend both processes to the spatio-temporal setting, introducing space–time dependence; (ii) introduce space and time varying covariate information; (iii) show how to implement fully model-based kriging and forecasting; (iv) allow a nugget which can be time dependent, and (v) provide an extensive comparison between the more sparsely parametrized wrapped Gaussian process with the more flexible projected Gaussian process. We do this illustratively using simulation, as a proof of concept, as well as with data in the form of computer model output for wave directions in the Adriatic Sea off the coast of Italy. The models are fitted under a Bayesian framework, introducing suitable latent variables, enabling full inference.

Modeling of circular data has progressed well beyond the i.i.d. case. Examples include linear models (Harrison and Kanji 1988; Fisher 1996; Kato and Shimizu 2008), linear models in a Bayesian context (Guttorp and Lockhart 1988; Damien and Walker 1999), models for circular time series (Breckling 1989; Fisher and Lee 1992; Coles 1998; Holtzman et al. 2006; Ravindran and Ghosh 2011), and hidden Markov models to address classification issues (Lagona and Picone 2011; Bulla et al. 2012; Mastrantonio et al. 2015). In Kato (2010) a Markov process for circular variables is presented. Jona Lasinio et al. (2012) consider a spatial wrapped Gaussian process. Wang and Gelfand (2013) explore the general projected normal model while in Wang and Gelfand (2014) Bayesian analysis of space–time circular data is developed using projected Gaussian processes. In Wang et al. (2015), directional wave data is modeled jointly with linear wave height data.

The format of the remainder of the paper is as follows. In Sect. 2, we review the wrapping approach and offer a non-separable space–time model for circular data. In Sect. 3, an analogous model is presented using the projected normal process. Section 4 presents several simulation examples giving insight into the inferential performance

of the models, while Sect. 5 analyzes the behavior of the models for wave directions. Section 6 extends the modeling approach to enable space–time varying covariates reflecting sea state at a location and time. Some concluding remarks are provided in Sect. 7. Implementation details, further simulated examples and more details on the real data application are available in the Supplementary Online Material, Sections S1, S2 and S3.

2 A brief review of the wrapped modeling approach

Let $Y \in \mathbb{R}$ be a random variable on the real line and let $g(y)$ and $G(y)$ be, respectively, its probability density function and cumulative distribution function. The random variable

$$X = Y \bmod 2\pi, \quad 0 \le X < 2\pi$$

is the wrapped version of Y having period 2π. The probability density function of X, $f(x)$, is obtained by wrapping the probability density function of Y, $g(y)$, around a circle of unit radius via the transformation $Y = X + 2\pi K$, with $K \in \mathbb{Z} \equiv \{0, \pm 1, \pm 2, \ldots\}$, and takes the form

$$f(x) = \sum_{k=-\infty}^{\infty} g(x + 2\pi k), \tag{1}$$

that is, a doubly infinite sum.

Equation (1) shows that $g(x + 2\pi k)$ is the joint distribution of (X, K). Hence, the marginal distribution of K is $P(K = k) = \int_0^{2\pi} g(x + 2\pi k) dx$, the conditional distributions $P(K = k|X = x) = g(x + 2\pi k)/\sum_{j=-\infty}^{\infty} g(x + 2\pi j)$ and the distribution of $X|K = k$ is $g(x + 2\pi k)/\int_0^{2\pi} g(x + 2\pi k) dx$. The introduction of K as latent variable facilitates model fitting (Jona Lasinio et al. 2012).

Following Coles (1998), we can extend the wrapping approach to multivariate distributions. Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_p) \sim g(\cdot)$, with $g(\cdot)$ a p-variate distribution on $\mathbb{R}^p$ indexed by say $\boldsymbol{\theta}$ and let $\mathbf{K} = (K_1, K_2, \ldots, K_p)$ be such that $\mathbf{Y} = \mathbf{X} + 2\pi\mathbf{K}$. Then the distribution of $\mathbf{X}$ is

$$f(\mathbf{X}) = \sum_{k_1=-\infty}^{+\infty} \sum_{k_2=-\infty}^{+\infty} \cdots \sum_{k_p=-\infty}^{+\infty} g(\mathbf{X} + 2\pi\mathbf{K}). \tag{2}$$

From (2) we see, as in the univariate case, that the joint density of $(\mathbf{X}, \mathbf{K})$ is $g(\mathbf{X} + 2\pi\mathbf{K})$. If $g(\cdot; \boldsymbol{\theta})$ is a p-variate normal density, with $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{X}$ has a p-variate wrapped normal distribution with parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Here, we introduce the latent random vector of winding numbers $\mathbf{K}$ to facilitate model fitting. Mardia and Jupp (1999) point out that only a few values of K are needed to obtain a reasonable approximation of the wrapped distribution and Jona Lasinio et al. (2012) show, when $g(\cdot; \boldsymbol{\theta})$ is Gaussian, how to choose the set of values of K based on the variance of the associated conditional distribution.

Let $Y(\mathbf{s})$ be a Gaussian process (GP) with $\mathbf{s} \in \mathbb{R}^2$, mean function $\mu(\mathbf{s})$ and covariance function say $\sigma^2 \rho(||\mathbf{s}_i - \mathbf{s}_j||; \boldsymbol{\psi})$, where $\boldsymbol{\psi}$ is a set of parameters. For a set of locations $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$, $\mathbf{Y} = (Y(\mathbf{s}_1), Y(\mathbf{s}_2), \ldots, Y(\mathbf{s}_n)) \sim N(\boldsymbol{\mu}, \sigma^2 C(\boldsymbol{\psi}))$, where $\boldsymbol{\mu} = (\mu(\mathbf{s}_1), \ldots, \mu(\mathbf{s}_n))$ and $C(\boldsymbol{\psi})_{ij} = \rho(\mathbf{s}_i - \mathbf{s}_j; \boldsymbol{\psi})$. As a consequence $\mathbf{X} = (X(\mathbf{s}_1), X(\mathbf{s}_2), \ldots, X(\mathbf{s}_n)) \sim WrapN(\boldsymbol{\mu}, \sigma^2 \mathbf{C}(\boldsymbol{\psi}))$ (Jona Lasinio et al. 2012), where $WrapN(\cdot, \cdot)$ indicates the wrapped normal distribution.

2.1 Space–time model specification and model fitting

Turning to space and time, suppose we seek $\{X(\mathbf{s}, t) \in [0, 2\pi), \mathbf{s} \in \mathcal{S} \subseteq \mathbb{R}^2, t \in \mathcal{T} \subseteq \mathbb{Z}^+\}$, a spatio-temporal process of angular variables. We can model $X(\mathbf{s}, t)$ as a spatio-temporal wrapped Gaussian process through its linear counterpart $Y(\mathbf{s}, t)$, extending the above approach. We assume that the linear process is a spatio-temporal Gaussian process having non-separable covariance structure with variance σ^2 and the stationary correlation function due to Gneiting [see equation (14) in Gneiting 2002]:

$$\text{Cor}(Y(\mathbf{s}, t), Y(\mathbf{s}', t')) \equiv \rho(\mathbf{h}, u) = \frac{1}{(a|u|^{2\alpha} + 1)^{\tau}} \exp\left(-\frac{c\|\mathbf{h}\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right), \quad (3)$$

where $(\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$, $\mathbf{h} = \mathbf{s} - \mathbf{s}'$ and $u = t - t'$. Here $d = 2$, a and c are non-negative scaling parameters for time and space, respectively. The smoothness parameters α and γ take values in $(0, 1]$, the space–time interaction parameter β is in $[0, 1]$, and $\tau \geq d/2 = 1$ is, in fact, fixed at 1 following Gneiting (2002). Attractively, as β decreases toward zero, we tend to separability in space and time.

We write the linear GP $Y(\mathbf{s}, t)$ as $Y(\mathbf{s}, t) = \mu_Y + \omega_Y(\mathbf{s}, t) + \tilde{\varepsilon}_Y(\mathbf{s}, t)$ where μ_Y is a constant mean function, $\omega_Y(\mathbf{s}, t)$ is a zero mean space–time GP with covariance function $\sigma^2 \rho(\mathbf{h}, u)$, and $\tilde{\varepsilon}(\mathbf{s}, t) \overset{iid}{\sim} N(0, \phi_Y^2)$, i.e., is pure error. It is convenient to work with the marginalized model where we integrate over all of the $\omega_Y(\mathbf{s}, t)$, see Banerjee et al. (2014). That is,

$$Y(\mathbf{s}, t) = \mu_Y + \varepsilon_Y(\mathbf{s}, t). \quad (4)$$

Then, $\varepsilon(\mathbf{s}, t)$ is a zero mean Gaussian process with covariance function

$$\text{Cov}(\varepsilon_Y(\mathbf{s}_i, t_j), \varepsilon_Y(\mathbf{s}_{i'}, t_{j'})) = \sigma_Y^2 \text{Cor}(\mathbf{h}_{i,i'}, u_{j,j'}) + \phi_Y^2 1_{(i=i')} 1_{(j=j')}.$$

To complete the model specification, we need to specify prior distributions. We suggest the following choices. Since a and c are positive, a and $c \sim G(\cdot, \cdot)$ where $G(\cdot, \cdot)$ denotes a gamma distribution. Since α, β, and γ are bounded between 0 and 1, we adopt a beta distribution ($B(\cdot, \cdot)$). Priors for the variances and the mean direction are given the usual normal-inverse gamma form, i.e., $\sigma_Y^2, \phi_Y^2 \sim IG(\cdot, \cdot)$, where $IG(\cdot, \cdot)$ denotes the inverse gamma, and $\mu_y \sim WrapN(\cdot, \cdot)$. In the sequel, this model will be denoted by WN.

2.2 Kriging and forecasting

We clarify prediction of the process at a new location and time, say $(\mathbf{s}_0, t_0)$, given what we have observed. We provide a full predictive distribution, extending Jona Lasinio et al. (2012) who only provide a posterior mean. Let $\mathcal{D} \subset \mathbb{R}^2 \times \mathbb{Z}^+$ be the set of n observed points. Let $\mathbf{X} = \{X(\mathbf{s}, t), (\mathbf{s}, t) \in \mathcal{D}\}$ be the vector of observed circular variables. Let $\mathbf{Y} = \{Y(\mathbf{s}, t), (\mathbf{s}, t) \in \mathcal{D}\}$ be the associated linear ones and let $\mathbf{K} = \{K(\mathbf{s}, t), (\mathbf{s}, t) \in \mathcal{D}\}$ be the associated vector of winding numbers. The predictive distribution we seek is $g(X(\mathbf{s}_0, t_0)|\mathbf{X})$. We use usual composition sampling within Markov chain Monte Carlo (MCMC) to obtain samples from it. Here, again we move from the circular process to the linear one, i.e., a sample from the distribution of $Y(\mathbf{s}_0, t_0)|\mathbf{X}$ can be considered as a sample from $X(\mathbf{s}_0, t_0), K(\mathbf{s}_0, t_0)|\mathbf{X}$. If we let $\boldsymbol{\Psi}_Y$ be the vector of all parameters, we can write

$$
\begin{aligned}
&g(X(\mathbf{s}_0, t_0), K(\mathbf{s}_0, t_0)|\mathbf{X}) \\
&= \sum_{\mathbf{K} \in \mathbb{Z}^n} \int_{\boldsymbol{\Psi}_Y} g(X(\mathbf{s}_0, t_0), K(\mathbf{s}_0, t_0)|\boldsymbol{\Psi}_Y, \mathbf{K}, \mathbf{X}) g(\boldsymbol{\Psi}_Y, \mathbf{K}|\mathbf{X}) d\boldsymbol{\Psi}_Y.
\end{aligned}
$$

So, suppose, for each posterior sample of $\mathbf{K}$ and $\boldsymbol{\Psi}_Y$ in $\{\mathbf{K}_l^*, \boldsymbol{\Psi}_{Y,l}^*, l = 1, 2, \ldots, L\}$ we generate a value from the distribution of $X(\mathbf{s}_0, t_0), K(\mathbf{s}_0, t_0)|\boldsymbol{\Psi}_Y, \mathbf{K}, \mathbf{X}$. Then, we will obtain the set of posterior samples $\{X_l^*(\mathbf{s}_0, t_0), K_l^*(\mathbf{s}_0, t_0), l = 1, 2, \ldots, L\}$ from $X(\mathbf{s}_0, t_0), K(\mathbf{s}_0, t_0)|\mathbf{X}$. If, we retain the set $\{X_l^*(\mathbf{s}_0, t_0), l = 1, 2, \ldots, L\}$, we will have samples from the desired predictive distribution.

Therefore, we need to sample from the distribution of $X(\mathbf{s}_0, t_0), K(\mathbf{s}_0, t_0)|\boldsymbol{\Psi}_Y, \mathbf{K}, \mathbf{X}$ or equivalently $Y(\mathbf{s}_0, t_0)|\mathbf{Y}, \boldsymbol{\Psi}_Y$. Let $\mathbf{1}_m$ be the $m \times 1$ vector of 1s, let $C_{\mathbf{Y}}$ be the correlation matrix of $\mathbf{Y}$, and let $C_{\mathbf{Y}, Y(\mathbf{s}_0, t_0)}$ be the correlation vector between $\mathbf{Y}$ and $Y(\mathbf{s}_0, t_0)$. Then, the joint distribution of $Y(\mathbf{s}_0, t_0), \mathbf{Y}|\boldsymbol{\Psi}_Y$ is

$$
\begin{pmatrix} Y(\mathbf{s}_0, t_0) \\ \mathbf{Y} \end{pmatrix} |\boldsymbol{\Psi}_Y \sim N\left(\begin{pmatrix} \mu_Y \\ \mu_Y \mathbf{1}_n \end{pmatrix}, \sigma_Y^2 \begin{pmatrix} 1 & C'_{\mathbf{Y}, Y(\mathbf{s}_0, t_0)} \\ C_{\mathbf{Y}, Y(\mathbf{s}_0, t_0)} & C_{\mathbf{Y}} \end{pmatrix} + \phi_Y^2 \mathbf{I}_{n+1} \right).
$$

As a result, the conditional distribution of $Y(\mathbf{s}_0, t_0)|\mathbf{Y}, \boldsymbol{\Psi}_Y$ is Gaussian with mean

$$
\mathrm{M}_{Y(\mathbf{s}_0, t_0)} = \mu_Y + \sigma_Y^2 C'_{\mathbf{Y}, Y(\mathbf{s}_0, t_0)} \left(\sigma_Y^2 C_{\mathbf{Y}} + \phi_Y^2 \mathbf{I}_n \right)^{-1} (\mathbf{Y} - \mu_Y \mathbf{1}_n)
$$

and variance

$$
\mathrm{V}_{Y(\mathbf{s}_0, t_0)} = \sigma_Y^2 + \phi_Y^2 - \sigma_Y^2 C'_{\mathbf{Y}, Y(\mathbf{s}_0, t_0)} \left(\sigma_Y^2 C_{\mathbf{Y}} + \phi_Y^2 \mathbf{I}_n \right)^{-1} \sigma_Y^2 C_{\mathbf{Y}, Y(\mathbf{s}_0, t_0)}.
$$

Finally, suppose, for each posterior sample, we simulate $Y_l^*(\mathbf{s}_0, t_0)$ from $N(\mathrm{M}_{Y(\mathbf{s}_0, t_0), l}^*, \mathrm{V}_{Y(\mathbf{s}_0, t_0), l}^*)$, where $\mathrm{M}_{Y(\mathbf{s}_0, t_0), l}^*$ and $\mathrm{V}_{Y(\mathbf{s}_0, t_0), l}^*$ are $\mathrm{M}_{Y(\mathbf{s}_0, t_0)}$ and $\mathrm{V}_{Y(\mathbf{s}_0, t_0)}$ computed with the lth sample. then, $X_l^*(\mathbf{s}_0, t_0) = Y_l^*(\mathbf{s}_0, t_0) \bmod 2\pi$ is a posterior sample from the predictive distribution.

3 The spatio-temporal projected normal process

Let (Z_1, Z_2) be a bivariate vector normally distributed with mean $\boldsymbol{\mu}_Z = (\mu_{Z_1}, \mu_{Z_2})$ and covariance matrix

$$\tilde{\mathbf{V}} = \begin{pmatrix} \sigma_{Z_1}^2 & \sigma_{Z_1}\sigma_{Z_2}\rho_z \\ \sigma_{Z_1}\sigma_{Z_2}\rho_z & \sigma_{Z_2}^2 \end{pmatrix}.$$

The vector $\mathbf{Z}$ is mapped into an angular variable Θ by the transformation $\Theta = \text{atan}^*(Z_2/Z_1)$, where the function $\text{atan}^*(S/C)$ is defined as $\text{atan}(S/C)$ if $C > 0$ and $S \geq 0$, $\pi/2$ if $C = 0$ and $S > 0$, $\text{atan}(S/C) + \pi$ if $C < 0$, $\text{atan}(S/C) + 2\pi$ if $C \geq 0$ and $S < 0$, undefined if $C = S = 0$. Θ is referred to as a projected normal random variable (Mardia 1972, p. 52) with parameters $\boldsymbol{\mu}_Z$ and $\tilde{\mathbf{V}}$. Wang and Gelfand (2013) note that the distribution of Θ does not change if we multiply (Z_1, Z_2) by a positive constant, so, following their lead, to identify the distribution we set $\sigma_{Z_2}^2 = 1$ and the covariance matrix becomes

$$\mathbf{V} = \begin{pmatrix} \sigma_{Z_1}^2 & \sigma_{Z_1}\rho_z \\ \sigma_{Z_1}\rho_z & 1 \end{pmatrix}.$$

Again, it is convenient to introduce a latent variable. Here, it is $R = ||\mathbf{Z}||$, obtaining the joint density of (Θ, R):

$$(2\pi)^{-1}|\mathbf{V}|^{1/2} \exp\left(-\frac{(r(\cos\theta, \sin\theta)' - \boldsymbol{\mu}_Z)'\mathbf{V}^{-1}(r(\cos\theta, \sin\theta)' - \boldsymbol{\mu}_Z)}{2}\right) r.$$

We can move back and forth between the linear variables and the pair (Θ, R) using the transformation $Z_1 = R\cos\Theta$, $Z_2 = R\sin\Theta$ and the equation $\Theta = \text{atan}^*(Z_2/Z_1)$.

Consider a bivariate spatio-temporal process $\mathbf{Z}(\mathbf{s}, t) = (Z_1(\mathbf{s}, t), Z_2(\mathbf{s}, t))$ with constant mean $\boldsymbol{\mu}_Z$ and cross covariance function $C(\mathbf{Z}(\mathbf{s}_i, t_j), \mathbf{Z}(\mathbf{s}_{i'}, t_{j'})) = \text{Cor}(\mathbf{s}_i - \mathbf{s}_{i'}, t_j - t_{j'})\mathbf{V}$ where $\text{Cor}(\cdot, \cdot)$ is a given space–time correlation function and $\mathbf{V}$ is as above. Then the circular process $\Theta(\mathbf{s}, t)$ induced by $\mathbf{Z}(\mathbf{s}, t)$ with the atan^* transformation is a projected Gaussian process with mean $\boldsymbol{\mu}_Z$ and covariance function induced by $C(\mathbf{Z}(\mathbf{s}_i, t_j), \mathbf{Z}(\mathbf{s}_{i'}, t_{j'}))$. More details on the properties of the process can be found in Wang and Gelfand (2014). Now, latent $R(\mathbf{s}, t)$s are introduced to facilitate model fitting.

3.1 Model specification and model fitting

We define the bivariate linear process as

$$Z_\ell(\mathbf{s}, t) = \mu_{Z_\ell} + \omega_{Z_\ell}(\mathbf{s}, t) + \tilde{\varepsilon}_{Z_\ell}(\mathbf{s}, t), \quad \ell = 1, 2, \tag{5}$$

where $\boldsymbol{\mu}_Z = (\mu_{Z_1}, \mu_{Z_2})'$ is the mean level, $\boldsymbol{\omega}_Z(\mathbf{s}, t) = (\omega_{Z_1}(\mathbf{s}, t), \omega_{Z_2}(\mathbf{s}, t))'$ is a bivariate Gaussian process with zero mean and covariance $\text{Cov}(\boldsymbol{\omega}_Z(\mathbf{s}_i, t_j), \boldsymbol{\omega}_Z(\mathbf{s}_{i'}, t_{j'})) = \text{Cor}(\mathbf{h}_{i,i'}, u_{j,j'})\mathbf{V}$ where $\text{Cor}(\mathbf{h}_{i,i'}, u_{j,j'})$ is defined in (3). Finally,

$\tilde{\boldsymbol{\varepsilon}}_Z(\mathbf{s}, t) = (\tilde{\varepsilon}_{Z_1}(\mathbf{s}, t), \tilde{\varepsilon}_{Z_2}(\mathbf{s}, t))$ is bivariate pure error with zero mean, independent components, and variance ϕ_Z^2. Marginalizing over the ω process in (5) yields

$$Z_\ell(\mathbf{s}, t) = \mu_{Z_\ell} + \varepsilon_{Z_\ell}(\mathbf{s}, t), \quad \ell = 1, 2,$$

where $\boldsymbol{\varepsilon}_Z(\mathbf{s}, t)$ is a mean zero bivariate Gaussian process with covariance function $\mathrm{Cov}(\boldsymbol{\varepsilon}_Z(\mathbf{s}_i, t_j), \boldsymbol{\varepsilon}_Z(\mathbf{s}_{i'}, t_{j'})) = \mathrm{Cor}(\mathbf{h}_{i,i'}, u_{j,j'})\mathbf{V} + \phi_Z^2\mathbf{I}_2 1_{(i=i')}1_{(j=j')}$.

$\Theta(\mathbf{s}, t) = \mathrm{atan}^*(Z_2(\mathbf{s}, t)/Z_1(\mathbf{s}, t))$ is a circular process and, as in the WN setting, correlation between the circular variables is induced by the Gneiting spatio-temporal correlation function. To specify the prior distributions for $\mu_{Z_1}, \mu_{Z_2}, \sigma_{Z_1}^2$ and ϕ_Z^2, we adopt the customary normal-inverse gamma specification. That is, $\mu_{Z_1}, \mu_{Z_2} \sim N(\cdot, \cdot)$, $\sigma_{Z_1}^2, \phi_Z^2 \sim IG(\cdot, \cdot)$ while, since ρ_Z is a correlation parameter, we adopt a truncated normal: $\rho_Z \sim N(\cdot, \cdot)I(-1, 1)$. In the sequel, this model will be denoted by *PN*.

We seek the predictive distribution at an unobserved location and time, $(\mathbf{s}_0, t_0)$. Let $\boldsymbol{\Theta}$ be the vector of observed circular values and $\mathbf{Z} = \{\mathbf{Z}(\mathbf{s}, t), (\mathbf{s}, t) \in \mathcal{D}\}$ be the associated linear ones. Let $\mathbf{Z}(\mathbf{s}_0, t_0) = (Z_1(\mathbf{s}_0, t_0), Z_2(\mathbf{s}_0, t_0))'$, $\mathbf{R} = \{R(\mathbf{s}, t), (\mathbf{s}, t) \in \mathcal{D}\}$ and let $\boldsymbol{\Psi}_Z$ be all the parameters of the projected model.

Specifically, the predictive distribution we seek is $\Theta(\mathbf{s}_0, t_0)|\boldsymbol{\Theta}$. If we sample from the distribution of $\mathbf{Z}(\mathbf{s}_0, t_0)|\boldsymbol{\Theta}$ then $\Theta(\mathbf{s}_0, t_0) = \mathrm{atan}^*(Z_2(\mathbf{s}_0, t_0)/Z_1(\mathbf{s}_0, t_0))$ is a sample from the desired predictive distribution. We have that

$$g(\mathbf{Z}(\mathbf{s}_0, t_0)|\boldsymbol{\Theta}) = \int_{\mathbf{R}} \int_{\boldsymbol{\Psi}_Z} g(\mathbf{Z}(\mathbf{s}_0, t_0)|\boldsymbol{\Psi}_Z, \mathbf{R}, \boldsymbol{\Theta})g(\boldsymbol{\Psi}_Z, \mathbf{R}|\boldsymbol{\Theta})\mathrm{d}\boldsymbol{\Psi}_Z\mathrm{d}\mathbf{R}.$$

So, we need to obtain $g(\mathbf{Z}(\mathbf{s}_0, t_0)|\boldsymbol{\Psi}_Z, \mathbf{R}, \boldsymbol{\Theta})$ and be able to sample from it. We start from the joint distribution of $\mathbf{Z}(\mathbf{s}_0, t_0), \mathbf{Z}|\boldsymbol{\Psi}_Z$:

$$\begin{pmatrix} \mathbf{Z}(\mathbf{s}_0, t_0) \\ \mathbf{Z} \end{pmatrix} |\boldsymbol{\Psi}_Z$$

$$\sim N\left(\begin{pmatrix} \boldsymbol{\mu}_Z \\ \mathbf{1}_n \otimes \boldsymbol{\mu}_Z \end{pmatrix}, \begin{pmatrix} 1 & \boldsymbol{C}'_{\mathbf{Z}, \mathbf{Z}(\mathbf{s}_0, t_0)} \\ \boldsymbol{C}_{\mathbf{Z}, \mathbf{Z}(\mathbf{s}_0, t_0)} & \mathbf{C}_{\mathbf{Z}} \end{pmatrix} \otimes \mathbf{V} + \phi_Z^2 \mathbf{I}_{2n+2}\right),$$

where $\mathbf{C}_{\mathbf{Z}}$ and $\boldsymbol{C}_{\mathbf{Z}, \mathbf{Z}(\mathbf{s}_0, t_0)}$ are the analogous of $\mathbf{C}_{\mathbf{Y}}$ and $\boldsymbol{C}_{\mathbf{Y}, Y(\mathbf{s}_0, t_0)}$ for the process $\mathbf{Z}(\mathbf{s}, t)$. The conditional distribution of $\mathbf{Z}(\mathbf{s}_0, t_0)|\mathbf{Z}, \boldsymbol{\Psi}_Z$ (equivalently $\mathbf{Z}(\mathbf{s}_0, t_0)|\boldsymbol{\Theta}, \mathbf{R}, \boldsymbol{\Psi}_Z$) is bivariate normal with mean

$$\mathbf{M}_{\mathbf{Z}(\mathbf{s}_0, t_0)} = \boldsymbol{\mu}_Z + \boldsymbol{C}'_{\mathbf{Z}, \mathbf{Z}(\mathbf{s}_0, t_0)} \otimes \mathbf{V}\left(\mathbf{C}_{\mathbf{Z}} \otimes \mathbf{V} + \phi_Z^2\mathbf{I}_{2n}\right)^{-1}(\mathbf{Z} - \mathbf{1}_n \otimes \boldsymbol{\mu}_Z)$$

and variance

$$\mathbf{V}_{\mathbf{Z}(\mathbf{s}_0, t_0)} = \mathbf{V} - \boldsymbol{C}'_{\mathbf{Z}, \mathbf{Z}(\mathbf{s}_0, t_0)} \otimes \mathbf{V}\left(\mathbf{C}_{\mathbf{Z}} \otimes \mathbf{V} + \phi_Z^2\mathbf{I}_{2n}\right)^{-1}\boldsymbol{C}_{\mathbf{Z}, \mathbf{Z}(\mathbf{s}_0, t_0)} \otimes \mathbf{V}.$$

Using the posterior samples $\{\mathbf{R}_l^*, \boldsymbol{\Psi}_{Z,l}^*, l = 1, 2, \ldots, L\}$ we can collect samples of $\Theta_l^*(\mathbf{s}_0, t_0)$ from its posterior predictive distribution.

4 Simulated examples

The Gneiting correlation function (3) has not been widely investigated within a Bayesian framework. The aim of this simulation study is essentially to provide a proof of concept. If space–time dependence, captured through the Gneiting correlation function, is driving an observed spatio-temporal circular dataset, can we learn about this dependence and can we demonstrate improved predictive performance by incorporating it in our modeling? We explore several different choices of parameters in (3).

For each proposed model, we simulated 48 datasets with $n = 240$ (20 locations and 12 time points) with spatial coordinates uniformly generated in $[0, 10] \times [0, 10]$. 24 datasets for the WN model were simulated from all possible combinations of $(a, c) = \{(1, 0.2), (0.2, 1)\}$, $\beta = \{0, 0.5, 1\}$, $\alpha = \{0.5, 0.8\}$, $\gamma = \{0.5, 0.8\}$ and $(\mu_Y, \sigma_Y^2, \phi_y^2) = (\pi, 0.1, 0.01)$. In the other 24 datasets we used the same combinations of correlation parameters but with $(\mu_Y, \sigma_Y^2, \phi_y^2) = (\pi, 1, 0.1)$. The datasets cover a wide range of situations in terms of spatio-temporal correlation: strong spatial correlation with weak temporal correlation $((a, c) = (1, 0.2))$, weak spatial correlation with strong temporal correlation $((a, c) = (0.2, 1))$, fully separable spatio-temporal correlation $(\beta = 0)$, non-separable $(\beta = \{0.5, 0.9\})$ and two levels for the smoothing parameters. The difference between the two collections of 24 datasets is that the first 24 have smaller circular variance than the remaining ones, where the circular variance was computed as one minus the mean resultant length divided by the sample size (Jammalamadaka and SenGupta 2001, p. 15).

The projected normal datasets were built according to the same rationale adopted for the wrapped normal, i.e., we built 24 datasets with small circular variance and 24 datasets with large circular variance. We simulated from unimodal projected distributions adopting the following sets of parameters:

– all possible combinations of $(a, c) = \{(1, 0.2), (0.2, 1)\}$, $\beta = \{0, 0.5, 1\}$, $\alpha = \{0.5, 0.8\}$, $\gamma = \{0.5, 0.8\}$ with $(\mu_{Z_1}, \mu_{Z_2}, \sigma_{Z_1}^2, \rho_Z, \phi_Z^2) = (2.5, 2.5, 1, 0, 0.01)$ which yields a circular variance close to the WN examples with $\sigma_Y^2 = 0.1$.
– all possible combinations of $(a, c) = \{(1, 0.2), (0.2, 1)\}$, $\beta = \{0, 0.5, 1\}$, $\alpha = \{0.5, 0.8\}$, $\gamma = \{0.5, 0.8\}$ with $(\mu_{Z_1}, \mu_{Z_2}, \sigma_{Z_1}^2, \rho_Z, \phi_Z^2) = (0.85, 0.85, 1, 0, 0.1)$ which, again, yields a circular variance close to the WN examples with $\sigma_Y^2 = 1$.

The parameters for the prior distributions were chosen so that the priors were centered on the "true" values used to simulate each dataset:

– correlation parameters: $a = 0.2 \Rightarrow a \sim G(2, 5)$, $a = 1 \Rightarrow a \sim G(5, 4)$, $c = 0.2 \Rightarrow c \sim G(2, 5)$, $c = 1 \Rightarrow c \sim G(5, 4)$, $\alpha = 0.5 \Rightarrow \alpha \sim B(5, 5)$, $\alpha = 0.8 \Rightarrow \alpha \sim B(6, 1.5)$, $\beta = 0 \Rightarrow \beta \sim B(1, 4)$, $\beta = 0.5 \Rightarrow \beta \sim B(5, 5)$, $\beta = 0.9 \Rightarrow \beta \sim B(6, 1.5)$, $\gamma = 0.5 \Rightarrow \gamma \sim B(5, 5)$, $\gamma = 0.8 \Rightarrow \gamma \sim B(6, 1.5)$;
– parameters of the WN model: $\mu_Y = 5 \Rightarrow \mu_Y \sim WN(\pi, 5)$, $\sigma_Y^2 = 0.1 \Rightarrow \sigma_Y^2 \sim IG(4.5, 0.55)$, $\sigma_Y^2 = 1 \Rightarrow \sigma_Y^2 \sim IG(2.01, 4.01)$, $\phi_Y^2 = 0.01 \Rightarrow \phi_Y^2 \sim IG(2.001, 0.03)$, $\phi_Y^2 = 0.1 \Rightarrow \phi_Y^2 \sim IG(4.5, 0.55)$;
– parameters of the PN model: $\mu_{Z_1} = 2.5 \Rightarrow \mu_{Z_1} \sim N(2.5, 5)$, $\mu_{Z_1} = 0.85 \Rightarrow \mu_{Z_1} \sim N(0.85, 5)$ $\mu_{Z_2} = 2.5 \Rightarrow \mu_{Z_2} \sim N(2.5, 5)$, $\mu_{Z_2} = 0.85 \Rightarrow$

$\mu_{Z_2} \sim N(0.85, 5)$, $\sigma^2_{Z_1} = 1 \Rightarrow \sigma^2_{Z_1} \sim IG(2.01, 4.01)$, $\rho_Z = 0 \Rightarrow \rho_Z \sim N(0, 1)I(-1, 1)$, $\phi^2_Z = 0.01 \Rightarrow \phi^2_Z \sim IG(2.001, 0.03)$, $\phi^2_Z = 0.1 \Rightarrow \phi^2_Z \sim IG(4.5, 0.55)$.

Among the 240 simulated observations in each dataset, 170 points, chosen between the first and tenth time points, were used for estimation and the remaining 70 points were set aside for validation purposes. The predictive performance was evaluated using two criteria. We computed an *average prediction error* (APE), defined as the average circular distance between a validation dataset and model predicted values, where we adopted as circular distance $d(\alpha, \beta) = 1 - \cos(\alpha - \beta)$ (Jammalamadaka and SenGupta 2001, p. 15). In particular, suppose the validation set has n^* observations, the APE for the models based on the wrapped normal is $\frac{1}{n^*} \sum_{(s_0, t_0)} d(\mu(s_0, t_0|\mathbf{X}), x(s_0, t_0))$ and $\frac{1}{n^*} \sum_{(s_0, t_0)} d(\mu(s_0, t_0|\boldsymbol{\Theta}), \theta(s_0, t_0))$ for the projected normal ones. Here, $x(s_0, t_0)$ and $\theta(s_0, t_0)$ are the realizations of the processes at (s_0, t_0) and $\mu(s_0, t_0|\mathbf{X})$ and $\mu(s_0, t_0|\boldsymbol{\Theta})$ are the posterior mean directions.

We also computed the *continuous ranked probability score* (CRPS) for circular variables as defined in Grimit et al. (2006):

$$\text{CRPS}(F, \delta) = E(d(\Delta, \delta)) - \frac{1}{2}E(d(\Delta, \Delta^*)), \tag{6}$$

where F is a predictive distribution, δ is a holdout value, and Δ and Δ^* are independent copies of a circular variable with distribution F. In this form, small values of CRPS are preferred.

For both models we do not know F in closed form, but we can compute a Monte Carlo approximation of (6). For the wrapped model, the CRPS for a held-out space–time point (s_0, t_0) is

$$\frac{1}{L}\sum_{l=1}^{L} d(x_l^*(s_0, t_0), x(s_0, t_0)) - \frac{1}{2L^2}\sum_{l=1}^{L}\sum_{j=1}^{L} d(x_l^*(s_0, t_0), x_j^*(s_0, t_0))$$

and for the projected model it is

$$\frac{1}{L}\sum_{l=1}^{L} d(\theta_l^*(s_0, t_0), \theta(s_0, t_0)) - \frac{1}{2L^2}\sum_{l=1}^{L}\sum_{j=1}^{L} d(\theta_l^*(s_0, t_0), \theta_j^*(s_0, t_0)).$$

For each of the 48 simulated datasets, the values of the mean CRPS under the two models, computed over the set of points used for model validation, are shown in Fig. 1. For both models we see that the CRPS depends heavily on the variance of the process, but seems unaffected by changes in the other parameters.

A potentially important difference between the two models is the computational time required to fit them. The WN model is computationally more efficient than the PN model; the main issue is computational complexity (see Supplementary Online Material, Section S1). The PN requires, at each MCMC iteration, roughly 8 times as
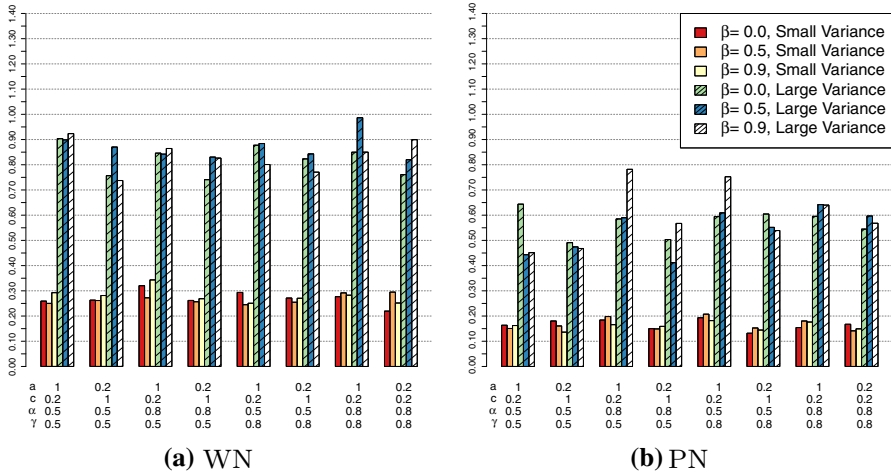
Fig. 1 Simulation study: CRPS comparing performances of the two proposed models. **a** WN. **b** PN

many operations as the WN to be fitted. If computational time is a relevant issue, then the WN may be more attractive.

5 Real data

We model wave directions obtained as outputs from a deterministic computer model implemented by Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA). The computer model starts from a wind forecast model predicting the surface wind over the entire Mediterranean. The hourly evolution of sea wave spectra is obtained by solving energy transport equations using the wind forecast as input. Wave spectra are locally modified using a source function describing the wind energy, the energy redistribution due to nonlinear wave interactions, and energy dissipation due to wave fracture. The model produces estimates every hour on a grid with 10×10 km cells (Speranza et al. 2004, 2007). The ISPRA dataset has forecasts for a total of 4941 grid points over the Italian Mediterranean. Over the Adriatic Sea area, there are 1494 points.

Our aim is to compare the performance of the WN and PN models. From a phenomenological perspective, the PN model is arguably the more natural choice since we are not wrapping a linear scale to obtain the directions. However, the WN model does provide a suitable model and, as suggested above, it may be attractive in terms of computational efficiency and interpretability of parameters. In the selected dataset, the three sea states, *calm, transition* and *storm* are present. The sea state is defined through the wave height (which is also supplied by the computer model output): when this height is below 1 m, we have *calm*, when it is between 1 and 2 m we have *transition* (between calm and storm) and when it is greater than 2 m we have a *storm*. Wave directions vary more in calm than in storm. Here, we seek to learn about the spatio-temporal structure of the data relying only on the specification of the correla-
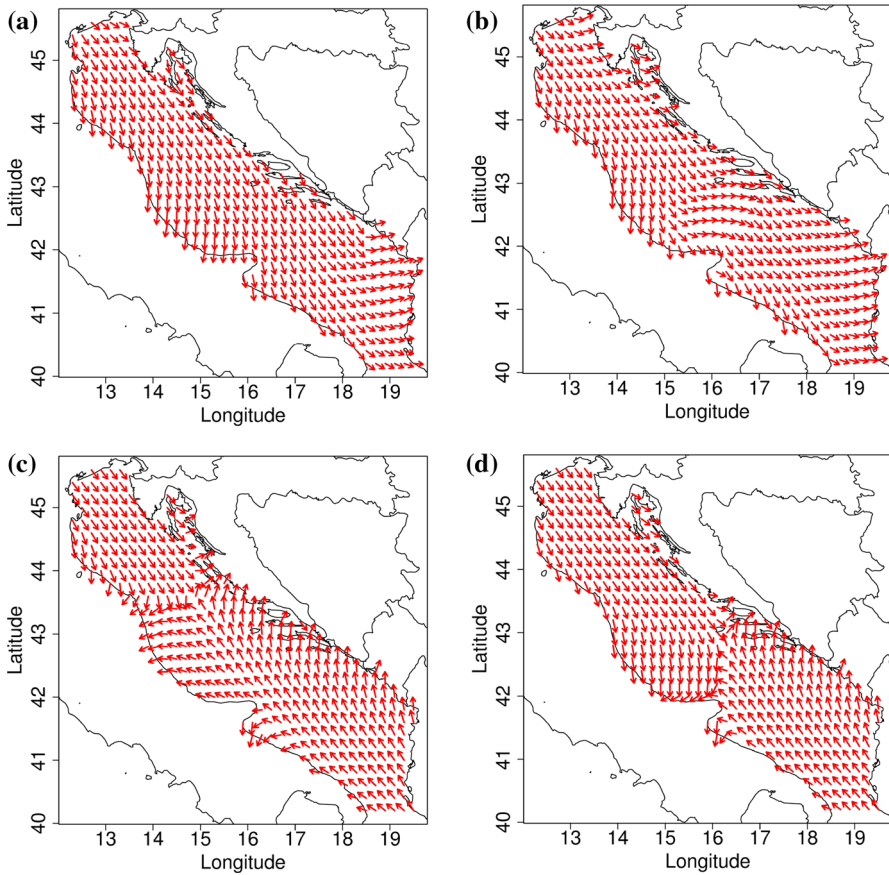
Fig. 2 Time windows for different sea states used for validation. The *four panels* represent the observed wave direction over the entire area at: **a** 12:00 on 5/5/2010 (storm); **b** 00:00 on 6/5/2010 (transition between storm and calm); **c** 00:00 on 7/5/2010 (calm); **d** 12:00 on 7/5/2010 (one-step prediction, calm)

tion function. We will use the information given by the wave heights in the models proposed in Sect. 6.1.

We fitted the model using 100 spatial points × 10 time points 6 h apart (1000 observations in total) in order to have a dataset including all sea states. Notice that spatial distances are evaluated in kilometers. Then, we developed four validation datasets, each with 350 spatial points and 1 time point. Specifically, we have one dataset for each sea state plus one for a one-step forward prediction. Finally, we used the model fitted over the 1000 points to predict each validation dataset. Three of the datasets are inside the time window used for model estimation, one in calm sea, one in transition and one during a storm. The fourth validation set is at 12:00 on May 7, 2010, 6 h after the last time used for model fitting. The observed circular process in each of these four time windows can be seen in Fig. 2. For each time window and model we computed the mean CRPS and APE, see Table 1. Furthermore, we computed the mean CRPS and APE over the four time windows.

		WN	PN
Table 1 Real data example: CRPS and APE for the WN and PN models computed on each validation dataset	Average		
	CRPS	0.655	0.629
	APE	0.437	0.421
	Calm		
	CRPS	1.450	1.398
	APE	0.995	0.973
	Transition		
	CRPS	0.082	0.074
	APE	0.033	0.028
	Storm		
	CRPS	0.063	0.042
	APE	0.026	0.009
	One-step prediction		
	CRPS	1.024	1.001
	APE	0.693	0.674

Following our discussion in Sects. 2.1 and 3, we used the following priors: $a \sim G(1.5, 1)$, $c \sim G(1.5, 1)$, $\alpha \sim B(2, 2.5)$, $\beta \sim B(1.1, 2)$, $\gamma \sim B(2, 2.5)$, $\sigma_Y^2 \sim IG(2, 2)$, $\phi_Y^2 \sim IG(1, 0.25)$, $\mu_Y \sim WrapN(\pi, 10)$, $\mu_{Z_1} \sim N(0, 10)$, $\mu_{Z_2} \sim N(0, 10)$, $\rho_Z \sim N(0, 5)I(-1, 1)$, $\sigma_Z^2 \sim IG(2, 2)$ and $\phi_Z^2 \sim IG(1, 0.25)$. Notice that all distributions are weakly informative. Also, the prior for β is centered near 0.1, i.e., close to the separable model. Decay parameters in space and time are related to the minimum and maximum distances in space and time, chosen to ensure that they concentrate the probability mass over such intervals.

As we expected, the predictive capability of the two models, in terms of both CRPS and APE, is poorest in a calm state, the variance being larger than in other states. On the other hand, it is very accurate during a storm or a transition for both models as we can see in Table 1. The PN always performs better that the WN. The largest difference between the APE values of the two models (0.022) is observed during the calm sea time window.

In Table 2 we give credible intervals and posterior mean estimates for the value of the parameters of the correlation function. For both models non-separable correlation structure is strongly supported. The point estimates of the spatial (c) and temporal (a) decay are smaller in the PN model. Notice that data are bimodal whenever the wave directions look like those in Fig. 2c, d, i.e., when over a large region at a given time a storm is rotating or two different weather systems are meeting. Then, scalar statistics, such as the overall mean direction or the overall concentration, may not be informative regarding this behaviour.

In the Supplementary Online Material, we provide the parameter estimates for the wrapped and projected distributions with associated 95 % credible intervals (Table S1). Since μ_Y is defined on a circular domain (recall that the prior on μ_Y is $WrapN(\cdot, \cdot)$), following Jona Lasinio et al. (2012), we can compute a 95 % credible interval as the arc that contains the central 95 % of the posterior samples.

Table 2 Real data example: mean point estimate (PE) and 95 % credible interval (CI) for the correlation parameters for the WN and PN models

	WN	PN
a		
PE	0.076	0.009
(CI)	(0.019, 0.200)	(0.005, 0.019)
c		
PE	3.2×10^{-4}	1.4×10^{-4}
(CI)	$(1.3 \times 10^{-4}, 7.1 \times 10^{-4})$	$(7.0 \times 10^{-4}, 2.9 \times 10^{-4})$
α		
PE	0.495	0.693
(CI)	(0.288, 0.744)	(0.562, 0.819)
β		
PE	0.592	0.430
(CI)	(0.158, 0.915)	(0.101, 0.774)
γ		
(PE)	0.797	0.872
(CI)	(0.697, 0.897)	(0.779, 0.939)

6 Extending the models

In the framework of the wrapped and projected normal models, introducing covariate information to explain the angular response is straightforward. For the wrapped approach we revise the linear version (4) to $Y(\mathbf{s}, t) = \mu_Y(\mathbf{s}, t) + \varepsilon_Y(\mathbf{s}, t)$.

The external variables can be introduced by modeling the mean of the circular process. Linear specification induces a circular likelihood for the regression coefficients that has infinitely many maxima of comparable size since this model wraps the line infinitely many times around the circle, (see for example Johnson and Wehrly 1978; Fisher and Lee 1992). To address this problem it is customary to limit the domain of $\mu_Y(\mathbf{s}, t)$ using a link function, i.e., $\mu_Y(\mathbf{s}, t) = \mathcal{L}(\mathbf{H}(\mathbf{s}, t)\,\boldsymbol{\eta})$, where $\mathcal{L}(\cdot) : \mathbb{R} \to I$ is the *link function* and I is some interval of $\mathbb{R}$ of length equal to the circular variable period, in our case 2π. We employ the *inverse tan link* (Fisher and Lee, 1992).

If only categorical covariates are available we do not need a link function; we can adopt an ANOVA representation for the relation between circular response and discrete covariates. This is computationally more efficient (see Supplementary Online Material, Section S1). Illustratively, suppose we have two predictors, with m_1 and m_2 levels, respectively, say $\mathbf{H}_1 = (H_{1,1}, \ldots, H_{1,m_1})$ and $\mathbf{H}_2 = (H_{2,1}, \ldots, H_{2,m_2})$. Then, to simplify the condition ensuring $\mu_Y(\mathbf{s}, t) \in I$, we use the following parametrization:

$$\mu_Y(\mathbf{s}, t) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \mu_{Y,im_2+j} \mathbf{1}_{(H_1(\mathbf{s},t)=H_{1,i})} \mathbf{1}_{(H_2(\mathbf{s},t)=H_{2,j})}.$$

Table 3 Real data example: CRPS and APE for WNR, WNA, PNR and PNA models computed on each validation dataset

	WNR	WNA	PNR	PNA
Average				
CRPS	0.668	0.644	0.507	0.588
APE	0.502	0.431	0.496	0.450
Calm				
CRPS	1.548	1.409	1.129	1.342
APE	1.158	0.997	0.985	0.984
Transition				
CRPS	0.095	0.094	0.092	0.093
APE	0.033	0.030	0.046	0.038
Storm				
CRPS	0.057	0.054	0.118	0.053
APE	0.016	0.013	0.110	0.012
One-step prediction				
CRPS	0.971	1.018	0.689	0.866
APE	0.802	0.685	0.841	0.765

We can also introduce the covariates into the specifications for the variances, creating $\sigma_Y^2(\mathbf{s}, t)$ and $\phi_Y^2(\mathbf{s}, t)$. Again, we consider ANOVA-type models, e.g., $\sigma_Y^2(\mathbf{s}, t) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \sigma_{Y,im_2+j}^2 \mathbf{1}_{(H_1(\mathbf{s},t)=H_{1,i})} \mathbf{1}_{(H_2(\mathbf{s},t)=H_{2,j})}$ and $\phi_Y^2(\mathbf{s}, t) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \phi_{Y,im_2+j}^2 \mathbf{1}_{(H_1(\mathbf{s},t)=H_{1,i})} \mathbf{1}_{(H_2(\mathbf{s},t)=H_{2,j})}$.

We investigate two models, both with an ANOVA parametrization for $\sigma_Y^2(\mathbf{s}, t)$ and $\phi_Y^2(\mathbf{s}, t)$ while for the mean, one has an ANOVA parametrization (*WNA*) and the other has a regression form (*WNR*). Below, we obtain an ANOVA form if we work with sea state and a regression form if we work with wave height. As prior distributions we propose: $N(\cdot, \cdot)$ for $\eta_{Y,i}, i = 1, 2, \ldots$, that is, a customary prior for a regression coefficient; $WrapN(\cdot, \cdot)$ for $\mu_{Y,i}, i = 1, 2, \ldots$, the circular equivalent of a normal prior over mean level; and $IG(\cdot, \cdot)$ for $\sigma_{Y,i}^2$ and $\phi_{Y,i}^2, i = 1, 2, \ldots$, that is, the customary prior for a variance. To sample from the predictive distribution, we adopt the same procedure used above for the WN model.

To introduce dependence on covariates in the projected normal model, following Wang and Gelfand (2013), we revise Eq. (5) to $Z_\ell(\mathbf{s}, t) = \mu_{Z_\ell}(\mathbf{s}, t) + \omega_{Z_\ell}(\mathbf{s}, t) + \tilde{\varepsilon}_{Z_\ell}(\mathbf{s}, t), \ell = 1, 2$ where the mean of the linear bivariate process is a function of space and/or time and $\tilde{\varepsilon}_{Z_\ell}(\mathbf{s}, t) \overset{iid}{\sim} N(0, \phi_Z^2(\mathbf{s}, t))$. Then we marginalize over $\omega_Z(\mathbf{s}, t)$ to obtain $Z_\ell(\mathbf{s}, t) = \mu_{Z_\ell}(\mathbf{s}, t) + \varepsilon_{Z_\ell}(\mathbf{s}, t), \ell = 1, 2$. We write $\mu_{Z_\ell}(\mathbf{s}, t) = \mathbf{H}(\mathbf{s}, t)\boldsymbol{\eta}_{Z_\ell}, \ell = 1, 2$ and $\phi_Z^2(\mathbf{s}, t) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \phi_{Z,im_2+j}^2 \mathbf{1}_{(H_1(\mathbf{s},t)=H_{1,i})} \mathbf{1}_{(H_2(\mathbf{s},t)=H_{2,j})}$, where $\boldsymbol{\eta}_{Z_\ell} = (\eta_{Z_\ell,1}, \eta_{Z_\ell,2}, \ldots)'$. Note that, depending on the types of variables in $\mathbf{H}(\mathbf{s}, t)$, continuous or categorical, we can specify a (projected normal) regression (*PNR*) or (projected normal) ANOVA (*PNA*). As noted in Wang and Gelfand (2014), there is complex interaction among the parameters in the general projected normal, complicating interpretation of the behavior of the resulting projected normal distributions as we vary
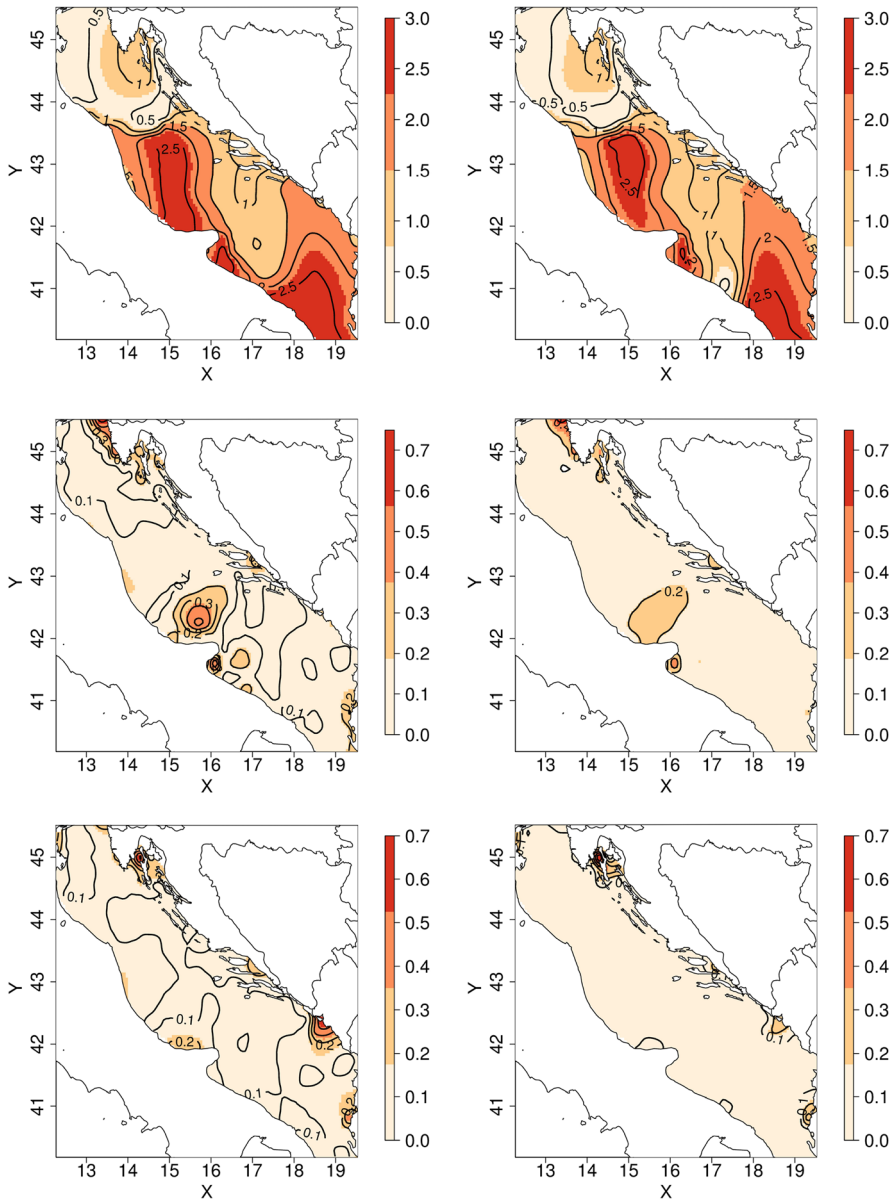
Fig. 3 Real data example: CRPS surfaces for the WN (*first column*) and PN (*second column*) models, under calm (*first row*), transition (*second row*) and storm (*third row*) states. Scales differ across states

them. With the same rationale used for the priors of the WNA and WNR models, we propose $\eta_{Z_\ell,i} \sim N(\cdot,\cdot)$, $l = 1, 2$, $i = 1, 2, \ldots$ and $\phi_{Z,i} \sim IG(\cdot,\cdot)$, $i = 1, 2, \ldots$. Here, again, we can sample from the predictive distribution adopting the same procedure as illustrated in Sect. 3.1.

Table 4 Real data example: mean point estimate (PE) and 95 % credible interval (CI) for the correlation parameters of the WNA, WNR, PNA and PNR models

	WNR	WNA
a		
PE	0.015	0.008
(CI)	(0.005, 0.035)	(0.003, 0.020)
c		
PE	6.1×10^{-5}	4.0×10^{-5}
(CI)	$(2.0 \times 10^{-5}, 1.4 \times 10^{-4})$	$(2.0 \times 10^{-5}, 7.0 \times 10^{-5})$
α		
PE	0.620	0.611
(CI)	(0.445, 0.786)	(0.434, 0.765)
β		
PE	0.396	0.539
(CI)	(0.070, 0.830)	(0.181, 0.868)
γ		
(PE)	0.705	0.936
(CI)	(0.620, 0.794)	(0.880, 0.976)
	PNR	PNA
a		
PE	0.119	0.108
(CI)	(0.042, 0.267)	(0.042, 0.225)
c		
PE	3.0×10^{-3}	1.0×10^{-3}
(CI)	$(1.01 \times 10^{-3}, 1.35 \times 10^{-3})$	$(4.60 \times 10^{-4}, 3.46 \times 10^{-3})$
α		
PE	0.575	0.506
(CI)	(0.345, 0.763)	(0.340, 0.706)
β		
PE	0.082	0.063
(CI)	(0.000, 0.402)	(0.000, 0.300)
γ		
(PE)	0.561	0.541
(CI)	(0.435, 0.677)	(0.441, 0.645)

6.1 Application to the wave data

We fitted the new models using the same dataset as in Sect. 5. For the ANOVA representation we used, as a categorical variable, the state of the sea while for the regression setting we used the significant wave height. Adopting the same rationale as in Sect. 5, the prior distributions for the regression coefficients ($\eta_{Y,j,i}$ and

Table 5 Real data example: mean point estimate (PE) and 95 % credible interval (CI) of the parameters of the WNA and WNR models

	WNA			
	$\mu_{Y,\mathrm{calm}}$	$\sigma^2_{Y,\mathrm{calm}}$	$\phi^2_{Y,\mathrm{calm}}$	
PE	0.095	1.524	0.051	
(CI)	(5.232, 1.328)	(0.959, 2.387)	(0.039, 0.068)	
	$\mu_{Y,\mathrm{tran}}$	$\sigma^2_{Y,\mathrm{tran}}$	$\phi^2_{Y,\mathrm{tran}}$	
PE	5.998	0.541	0.018	
(CI)	(5.278, 0.490)	(0.332, 0.876)	(0.013, 0.026)	
	$\mu_{Y,\mathrm{storm}}$	$\sigma^2_{Y,\mathrm{storm}}$	$\phi^2_{Y,\mathrm{storm}}$	
PE	5.860	0.385	0.009	
(CI)	(5.254, 0.281)	(0.246, 0.582)	(0.007, 0.012)	
	WNR			
	$\eta_{Y,0,\mathrm{calm}}$	$\eta_{Y,1,\mathrm{calm}}$	$\sigma^2_{Y,\mathrm{calm}}$	$\phi^2_{Y,\mathrm{calm}}$
PE	0.997	4.918	5.000	0.041
(CI)	(0.360, 1.901)	(2.433, 7.619)	(2.313, 9.494)	(0.027, 0.058)
	$\eta_{Y,0,\mathrm{tran}}$	$\eta_{Y,1,\mathrm{tran}}$	$\sigma^2_{Y,\mathrm{tran}}$	$\phi^2_{Y,\mathrm{tran}}$
PE	3.166	2.526	1.825	0.018
(CI)	(0.763, 5.894)	(0.174, 6.844)	(1.013, 3.046)	(0.012, 0.025)
	$\eta_{Y,0,\mathrm{storm}}$	$\eta_{Y,1,\mathrm{storm}}$	$\sigma^2_{Y,\mathrm{storm}}$	$\phi^2_{Y,\mathrm{storm}}$
PE	3.470	1.933	1.322	0.010
(CI)	(0.666, 6.445)	(0.064, 5.870)	(0.660, 2.167)	(0.007, 0.013)

$\eta_{Z_\ell,j,i}$, $j = 0, 1$, $i = calm, trans, storm$) were all $N(0, 10)$. For the ANOVA coefficients, $\mu_{Y,i}$ and $\mu_{Z_\ell,i}$, they were all $WrapN(\pi, 10)$. For the $\sigma^2_{Y,i}$, they were all $IG(2, 2)$ and for the $\phi_{Y,i}$ and $\phi_{Z,i}$ they were all $IG(1, 0.25)$. The prior distributions for the other parameters were the same as those used in Sect. 5.

From Table 3 we see that the WNA model is generally preferred to the WNR. For the projected models, APE and CRPS are almost indistinguishable between PNA and PNR during transition. With one-step ahead predictions, the two criteria return contradicting choices; PNR is preferred with CRPS, PNA with the APE. With the calm sea state, the CRPS chooses PNR while APE does not yield a clear decision. With the storm state, both criteria are lower with the PNA model. Overall, our selection would be the PNA model but, more importantly, we value the informative comparison our approach enables. In fact, the remarkable improvement of PNA over PNR in storm is likely due to the very high predictability of direction during a storm period. In this regard, the PN models are generally preferred to the WN models except in storm where WNR, WNA, and PNA are essentially equivalent.

To analyze the local behavior of model fitting, in Fig. 3 we report CRPS surfaces, evaluated in calm, transition and storm for the two "best average APE" models, the WNA (see Table 3) and PN (see Table 1). We see that the local behavior of the

Table 6 Real data example: mean point estimate (PE) and 95 % credible interval (CI) for the parameters of the PNA and PNR models

	PNA				
	$\mu_{Z_1,\text{calm}}$	$\mu_{Z_2,\text{calm}}$	$\phi^2_{Z,\text{calm}}$		
PE	0.841	−0.404	0.027		
(CI)	(−1.112, 2.706)	(−2.408, 1.427)	(0.014, 0.051)		
	$\mu_{Z_1,\text{tran}}$	$\mu_{Z_2,\text{tran}}$	$\phi^2_{Z,\text{tran}}$		
PE	0.697	−0.724	0.047		
(CI)	(−1.281, 2.640)	(−2.600, 1.173)	(0.018, 0.099)		
	$\mu_{Z_1,\text{storm}}$	$\mu_{Z_2,\text{storm}}$	$\phi^2_{Z,\text{storm}}$		
PE	0.615	−0.615	0.037		
(CI)	(−1.376, 2.615)	(−2.543, 1.289)	(0.016, 0.076)		
	$\sigma^2_{Z,1}$	ρ_Z			
(PE)	2.072	−0.161			
(CI)	(1.425, 2.938)	(−0.320, 0.003)			
	PNR				
	$\eta_{Z_1,0,\text{calm}}$	$\eta_{Z_1,1,\text{calm}}$	$\eta_{Z_2,0,\text{calm}}$	$\eta_{Z_2,1,\text{calm}}$	$\phi^2_{Z,\text{calm}}$
PE	0.997	0.875	−0.925	0.840	0.110
(CI)	(−0.989, 3.026)	(−1.160, 2.927)	(−2.878, 1.091)	(−1.162, 2.798)	(0.033, 0.250)
	$\eta_{Z_1,0,\text{tran}}$	$\eta_{Z_1,1,\text{tran}}$	$\eta_{Z_2,0,\text{tran}}$	$\eta_{Z_2,1,\text{tran}}$	$\phi^2_{Z,\text{tran}}$
PE	0.916	0.976	−1.117	−0.554	0.127
(CI)	(−1.195, 3.015)	(−1.258, 3.117)	(−3.322, 0.893)	(−2.601, 1.649)	(0.037, 0.322)
	$\eta_{Z_1,0,\text{storm}}$	$\eta_{Z_1,1,\text{storm}}$	$\eta_{Z_2,0,\text{storm}}$	$\eta_{Z_2,1,\text{storm}}$	$\phi^2_{Z,\text{storm}}$
PE	0.768	1.088	−0.974	−1.190	0.091
(CI)	(−1.424, 2.899)	(−1.083, 3.235)	(−3.146, 1.177)	(−3.281, 0.955)	(0.031, 0.201)
	$\sigma^2_{Z,1}$	ρ_Z			
PE	2.293	−0.191			
(CI)	(1.602, 3.212)	(−0.358, −0.013)			

models is very similar. The worst predictions are found around the Gargano peninsula during calm. This is consistent with the physics of wave movement since, around the peninsula, local winds play a more relevant role, inducing very high variability in wave directions. The same behavior is shown with the other models. In terms of parameter estimation the WNA and PN models suggest a non-separable model (Tables 2, 4) with very strong spatial (*c*) and temporal (*a*) dependence. WNA suggests that a different nugget is necessary for each sea state. In fact analyzing the credible intervals of these parameters we observe that, for each sea state, nuggets are significantly different among them as their credible intervals do not overlap (Table 5). For the projected normal models (Table 6), all nugget credible intervals are substantially overlapping, suggesting that one nugget should be enough to model all sea states.

7 Conclusions

We have presented a range of models for spatio-temporal circular data based on the wrapped and projected normal distributions, incorporating space–time dependence, allowing explanatory variables, introducing a nugget, implementing kriging and forecasting. The models based on the projected normal are more flexible since they allow bimodal and asymmetric distributions while the wrapped normal is unimodal and symmetric. On the other hand, the wrapped normal models are easy to interpret and are computationally better behaved and more efficient. Predictions obtained under the two models are very close and almost indistinguishable when data are roughly unimodal and symmetric (see Supplementary Online Material, Section S2). Then, if fast computation is sought, WN models become attractive.

The projected normal process can be straightforwardly extended to general directional fields on the sphere since the projected normal distribution is well defined in this case, see Mardia and Jupp (1999). The wrapped Gaussian process is not easily extended to a sphere. In fact, we are unaware of any approach to wrap multivariate linear data onto spheres. Conceptually, such wrapping would *not* appear to be well defined.

Future work will find us enriching wrapped modeling to allow asymmetry through the use of skewed distributions. Skewness is easy to introduce by wrapping skew normal distributions. In a completely different direction, we are also extending the modeling to explore spatio-temporal data consisting of geo-coded locations with periodic (in time) behaviour that can be represented as a circular variable. There, we work with trivariate GPs in space and time, incorporating temporal projection.

References

Banerjee S, Gelfand AE, Carlin BP (2014) Hierarchical modeling and analysis for spatial data, 2nd edn. Chapman and Hall/CRC, New York

Breckling J (1989) The analysis of directional time series: applications to wind speed and directions. Lecture notes in statistics. Springer-Verlag, Berlin

Bulla J, Lagona F, Maruotti A, Picone M (2012) A multivariate hidden Markov model for the identification of sea regimes from incomplete skewed and circular time series. J Agric Biol Environ Stat 17:544–567

Coles S (1998) Inference for circular distributions and processes. Stat Comput 8:105–113

Damien P, Walker S (1999) A full Bayesian analysis of circular data using the von Mises distribution. Can J Stat 27:291–298

Fisher NI (1996) Statistical analysis of circular data. Cambridge University Press, Cambridge

Fisher NI, Lee AJ (1992) Regression models for an angular response. Biometrics 48:665–677

Gneiting T (2002) Nonseparable, stationary covariance functions for space-time data. J Am Stat Assoc 97:590–600

Grimit EP, Gneiting T, Berrocal VJ, Johnson NA (2006) The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. Q J R Meteorol Soc 132:2925–2942

Guttorp P, Lockhart RA (1988) Finding the location of a signal: a Bayesian analysis. J Am Stat Assoc 83:322–330

Harrison D, Kanji GK (1988) The development of analysis of variance for circular data. J Appl Stat 15:197–224

Holtzman H, Munk A, Suster M, Zucchini W (2006) Hidden Markov models for circular and linear-circular time series. Environ Ecol Stat 13:325–347

Jammalamadaka SR, SenGupta A (2001) Topics in circular statistics. World Scientific, Singapore

Johnson RA, Wehrly TE (1978) Some angular-linear distributions and related regression models. J Am Stat Assoc 73:602–606

Jona Lasinio G, Gelfand AE, Jona Lasinio M (2012) Spatial analysis of wave direction data using wrapped Gaussian processes. Ann Appl Stat 6:1478–1498

Kato S (2010) A Markov process for circular data. J R Stat Soc Ser B 72:655–672

Kato S, Shimizu K (2008) Dependent models for observations which include angular ones. J Stat Plan Infer 138:3538–3549

Lagona F, Picone M (2011) A latent-class model for clustering incomplete linear and circular data in marine studies. J Data Sci 9:585–605

Lee A (2010) Circular data. Wiley Interdiscip Rev Comput Stat 2:477–486

Mardia KV (1972) Statistics of directional data. Academic Press, London

Mardia KV, Jupp PE (1999) Directional statistics. Wiley, Chichester

Mastrantonio G, Maruotti A, Jona Lasinio G (2015) Bayesian hidden Markov modelling using circular-linear general projected normal distribution. Environmetrics 26:145–158

Ravindran P, Ghosh SK (2011) Bayesian analysis of circular data using wrapped distributions. Stat Sin 5:547–561

Speranza A, Accadia C, Casaioli M, Mariani S, Monacelli G, Inghilesi R, Tartaglione N, Ruti PM, Carillo A, Bargagli A, Pisacane G, Valentinotti F, Lavagnini A (2004) Poseidon: an integrated system for analysis and forecast of hydrological, meteorological and surface marine fields in the Mediterranean area. Nuovo Cimento 27:329–345

Speranza A, Accadia C, Mariani S, Casaioli M, Tartaglione N, Monacelli G, Ruti PM, Lavagnini A (2007) Simm: an integrated forecasting system for the Mediterranean area. Meteorol Appl 14:337–350

Wang F, Gelfand AE (2013) Directional data analysis under the general projected normal distribution. Stat Methodol 10:113–127

Wang F, Gelfand AE (2014) Modeling space and space–time directional data using projected Gaussian processes. J Am Stat Assoc 109:1565–1580

Wang F, Gelfand AE, 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. Stat Sin 25:25–39