

Analyzing spatial and spatio-temporal angular and linear data using Gaussian processes

Alan E. Gelfand and Fangpo Wang

Abstract It is valuable to have a better understanding of factors that influence sea motion in order to provide more accurate forecasts. Our application here is to data on wave heights and outgoing wave directions over a region in the Adriatic sea during the time of a storm, with the overarching goal of understanding the association between wave directions and wave heights and enabling improved prediction of wave behavior. Our contribution is to develop a fully model-based approach to capture joint structured spatial and temporal dependence between a linear and an angular variable. Model fitting is carried out using a suitable data augmented Markov chain Monte Carlo (MCMC) algorithm. Spatial interpolation and temporal forecasting is straightforward. We illustrate with data outputs from a deterministic wave model for region in the Adriatic Sea.

Key words: angular variable, Bayesian kriging, hierarchical model, latent variables, Markov chain Monte Carlo, projected Gaussian process, significant wave height

1 Introduction

Sea motion data are mostly available as outputs from deterministic models, especially if we are interested in exploring a relatively broad area of the sea. Recently, deterministic models based on the dynamics and thermodynamics of the atmosphere have been used to forecast the weather with increasing reliability. The outputs from such models are usually computed at several spatial and temporal resolutions. Build-

Alan E. Gelfand
Department of Statistical Science, Duke University, Durham, NC, USA e-mail: alan@stat.duke.edu

Fangpo Wang
Adobe Systems Inc., San Jose, CA. e-mail: wangfangpo@gmail.com

ing upon earlier work in [9] which focused solely upon wave directions, in this paper, we extend our interest to include wave heights along with outgoing wave directions and provide a general framework for joint modeling of these two variables. More generally, we offer a modeling approach for joint spatial and spatio-temporal analysis of an angular and a linear variable.

[9] have dealt with wave directions associated with spatial locations over time. In fact, our available data include both *significant* wave heights and *outgoing* wave directions (formally defined below) at the same spatial and temporal resolution. Figure 1 displays both the outgoing wave directions and wave heights over a region of the Adriatic sea, for a subset of available locations, at a particular time point during a storm. The wave heights are displayed in the image plot, as an additional layer over the arrow plots of the directions. There is strong visual evidence of spatial dependence for both heights and directions, although with quite different patterns. Rather than separately modeling the wave heights and wave directions, it is natural to think of building a joint model to accommodate the underlying association between them.

More broadly, the contribution of this paper is to develop a fully model-based approach to capture joint structured spatial dependence for modeling linear data and directional data at different spatial locations. We show that Bayesian model fitting under our model specification is straightforward using a suitable data augmented Markov chain Monte Carlo (MCMC) algorithm. This joint model framework allows natural extension to space-time data and can directly incorporate space-time covariate information, enabling both spatial interpolation and temporal forecasting.

Wave heights, like wind speed, are linear variables while wave directions are circular variables, measured in degrees relative to a fixed orientation. Some recent work to build space and space-time models for wave directions can be found in [5] and in [9]. Modeling linear and circular variables in a marine context has been considered in a likelihood framework by [6] and [1]. We often simultaneously observe realizations of a circular variable Θ and a linear variable X as pairs $(\theta_1, x_1), \dots, (\theta_n, x_n)$. In a spatio-temporal setting the pair might be the time of day of a robbery along with the amount stolen, indexed by a geo-coded location. Other pairs include the direction of departure and the distance to the destination or wind direction and humidity. In the literature to date, the discussion usually focuses on *conditional* modeling. For instance, if the angular variable Θ depends on linear variables, it is referred to as a *Linear-Circular regression* model. i.e. a regression specification to use the linear covariates to explain a directional response. Such modeling typically adopts the von Mises distribution and with a suitable link function which maps $\mathbb{R}^k$ to $(-\pi, \pi)$, e.g., the arctan function, where k is the dimension of covariates [4, 3]. Regression under the general projected normal was proposed in [8]. Very recent work by [2] considers nonparametric estimation in the regression setting, creating weighted mean direction estimators by minimizing expected angular distance.

Conditioning in the opposite direction, now using an angular covariate to explain a linear response, i.e., if the linear variable X depends on an independent angular variable, Θ , we have a regression model with a linear response and angular covariates, known as a *Circular-Linear regression* model. Again, a link function is employed [7, 4].

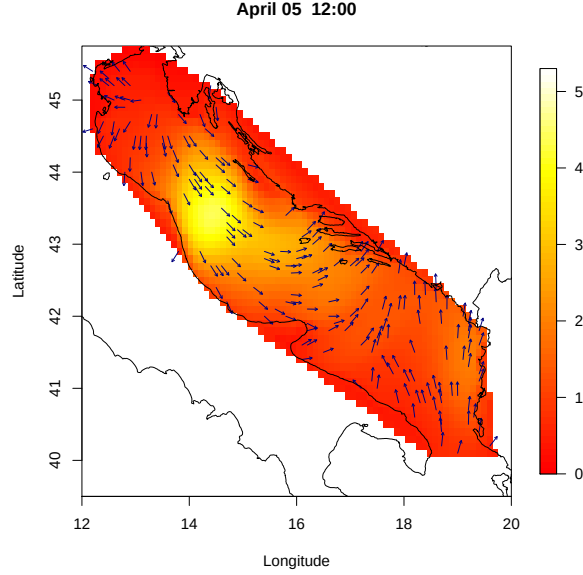


Fig. 1 Plot of wave heights and wave directions at 12:00 on April 5th, 2012

The circular-linear association between these two random variables is not as obvious as that between two linear variables. Since the angular variable Θ has the support on the circle, the pair (Θ, X) has the support on a cylinder. If their relationship can be written as $X = a + b \cos \Theta + c \sin \Theta$, [7] proposes a measure using the ordinary multiple correlation in the regression setting between X and $(\cos \Theta, \sin \Theta)$. Similarly, [4] discuss the dominant canonical correlation coefficient between X and $(\cos \Theta, \sin \Theta)$. The modeling that we offer below is in the spirit of this representation, of conditioning X on Θ .

2 Model specification

We begin with a single linear variable and a single directional variable, e.g., wave height and wave direction. We build a joint parametric model for the wave height H and the wave direction Θ by introducing a latent variable R (which will be specified below), in the form

$$\begin{aligned} f(H, \Theta | \Psi_h, \Psi_\theta) &= \int f(H, \Theta, R | \Psi_h, \Psi_\theta) dR \\ &= \int f(H | \Theta, R, \Psi_h) f(\Theta, R | \Psi_\theta) dR, \end{aligned} \quad (1)$$

where Ψ_h and Ψ_θ are sets of parameters associated with the conditional model for height and the marginal model for direction, respectively.

More precisely, as stochastic processes, for each location $\mathbf{s}$ in the domain of interest $\mathcal{D}$, the linear variable (height) is denoted as $H(\mathbf{s})$ and the angular variable (direction) as $\Theta(\mathbf{s})$. For the marginal distribution $\Theta(\mathbf{s})|\Psi_\theta$, we propose a stationary projected Gaussian process [9] with a constant mean $\mu = (\mu_1, \mu_2)^T$ and separable cross-covariance $C_\theta(\mathbf{s}, \mathbf{s}') = \rho_\theta(\mathbf{s} - \mathbf{s}'; \phi_\theta) \cdot T$, where $\Psi_\theta = \{\mu, T, \phi_\theta\}$, ϕ_θ is the decay parameter associated with the correlation function $\rho_\theta(\cdot)$ and T is a 2×2 matrix defined as $T = \begin{pmatrix} \tau_\theta^2 & \rho\tau_\theta \\ \rho\tau_\theta & 1 \end{pmatrix}$.

As defined in [8], the projected Gaussian process is induced from an inline Gaussian process $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), Y_2(\mathbf{s}))^T$ with a constant mean μ and cross-covariance $C_\theta(\mathbf{s}, \mathbf{s}')$. Essentially, we can transform back and forth between the two random variable spaces $(\Theta(\mathbf{s}), R(\mathbf{s}))$ and $(Y_1(\mathbf{s}), Y_2(\mathbf{s}))$ through $Y_1(\mathbf{s}) = R(\mathbf{s}) \cos \Theta(\mathbf{s})$ and $Y_2(\mathbf{s}) = R(\mathbf{s}) \sin \Theta(\mathbf{s})$, thus defining the latent process, $R(\mathbf{s})$.

At the top or conditional level of the hierarchy in (1), $[H|\Theta, R, \Psi_h]$, we specify a univariate Bayesian spatial regression (a customary *geostatistical* model) for the wave height $H(\mathbf{s})|\Theta(\mathbf{s}), R(\mathbf{s}), \Psi_h$, with $\Theta(\mathbf{s})$ and $R(\mathbf{s})$ included in the mean through a link function $g(\cdot)$. That is,

$$H(\mathbf{s}) = g(\Theta(\mathbf{s}), R(\mathbf{s})) + w(\mathbf{s}) + \varepsilon(\mathbf{s}).$$

As usual, the residual is partitioned into the spatial effect term $w(\mathbf{s})$ and the non-spatial error term $\varepsilon(\mathbf{s})$, where $w(\mathbf{s})$ is assumed to follow a zero mean stationary Gaussian process with covariance function $C_h(\mathbf{s} - \mathbf{s}')$ and the $\varepsilon(\mathbf{s})$'s are uncorrelated pure errors. A natural choice for the link function $g(\cdot)$ will be the linear combination of the trigonometric terms of $\Theta(\mathbf{s})$ multiplied by the latent $R(\mathbf{s})$. In fact, this actually is the linear regression on $Y_1(\mathbf{s})$ and $Y_2(\mathbf{s})$ from the “unobserved” inline Gaussian process $\mathbf{Y}(\mathbf{s})$. Explicitly, with this choice of $g(\cdot)$, we have

$$H(\mathbf{s}) = \beta_0 + \beta_1 R(\mathbf{s}) \cos \Theta(\mathbf{s}) + \beta_2 R(\mathbf{s}) \sin \Theta(\mathbf{s}) + w(\mathbf{s}) + \varepsilon(\mathbf{s}) \quad (2)$$

$$= \beta_0 + \beta_1 Y_1(\mathbf{s}) + \beta_2 Y_2(\mathbf{s}) + w(\mathbf{s}) + \varepsilon(\mathbf{s})$$

$$= \mathbf{X}(\mathbf{s})^T \beta + w(\mathbf{s}) + \varepsilon(\mathbf{s}) \quad (3)$$

where the spatial random effect $w(\mathbf{s})$ follows a zero-centered GP with covariance function $C_h = \sigma_h^2 \rho_h(\mathbf{s} - \mathbf{s}'; \phi_h)$, the error term $\varepsilon(\mathbf{s}) \stackrel{\text{iid}}{\sim} N(0, \tau_h^2)$. In (3), we denote the regression coefficient vector by $\beta = (\beta_0, \beta_1, \beta_2)^T$ and the location specific covariate vector as $\mathbf{X}(\mathbf{s}) = (1, R(\mathbf{s}) \cos \Theta(\mathbf{s}), R(\mathbf{s}) \sin \Theta(\mathbf{s}))^T$. Under this model specification, the parameters associated with the heights are $\Psi_h = \{\beta, \phi_h, \sigma_h^2, \tau_h^2\}$. Altogether, $(H(\mathbf{s}), Y_1(\mathbf{s}), Y_2(\mathbf{s}))^T$ specifies a trivariate Gaussian process.

The coefficients β_1 and β_2 provide information regarding the association between the circular random variable Θ and the linear variable H under the specific specification in (2). When β_1 and β_2 are both 0, we are fitting the wave heights and the wave directions separately as a spatial regression model and a projected Gaussian process model. The square of the multiple correlation coefficient, $R_{H|Y}^2$, is a measure

of the strength of the conditional dependence (explanation), as is customary in linear regression. Notably, it is for a regression where the covariates are not observed. However, it can be obtained, as a parametric function, from the joint dependence structure and is free of location. In particular, if $\Delta = \beta_1^2 \tau_\theta^2 + \beta_2^2 + 2\beta_1 \beta_2 \tau_\theta \rho$,

$$R_{H|Y}^2 = \frac{\Delta}{\Delta + \sigma_h^2 + \tau_h^2}. \quad (4)$$

The posterior distribution of $R_{H|Y}^2$ can be obtained directly from posterior samples.

3 Implementing kriging

Bayesian kriging under the joint model framework is straightforward. For a new location $\mathbf{s}_0$, our goal is to obtain a joint distribution of $H(\mathbf{s}_0)$ and $\Theta(\mathbf{s}_0)$ given the data $(\theta, \mathbf{h})$, expressed as

$$f(\theta(\mathbf{s}_0), h(\mathbf{s}_0) | \theta, \mathbf{h}) = \int \int \int f(\theta(\mathbf{s}_0), h(\mathbf{s}_0), r(\mathbf{s}_0) | \mathbf{r}, \Psi, \theta, \mathbf{h}) f(\mathbf{r}, \Psi | \theta, \mathbf{h}) d\mathbf{r} dr_0 d\Psi,$$

where $R(\mathbf{s}_0)$ is the latent random variable associated with $\Theta(\mathbf{s}_0)$ and the parameters $\Psi = \{\Psi_\theta, \Psi_h\}$.

For kriging of the circular variable at the new location, $\Theta(\mathbf{s}_0)$, we simplify by starting from the lower level, the joint distribution of $\mathbf{Y}(\mathbf{s}_0) = (Y_1(\mathbf{s}_0), Y_2(\mathbf{s}_0))^T$ and $\mathbf{Y}^* = (Y_1(\mathbf{s}_1), Y_2(\mathbf{s}_1), \dots, Y_1(\mathbf{s}_n), Y_2(\mathbf{s}_n))^T$ is,

$$\begin{pmatrix} \mathbf{Y}(\mathbf{s}_0) \\ \mathbf{Y}^* \end{pmatrix} \sim MVN \left(\begin{pmatrix} \mu(\mathbf{s}_0) \\ \mu^* \end{pmatrix}, \begin{pmatrix} 1 & \rho_{0,Y}^T(\phi_\theta) \\ \rho_{0,Y}(\phi_\theta) & \Gamma_Y(\phi_\theta) \end{pmatrix} \otimes T \right),$$

where $\mu^* = (\mu^T(\mathbf{s}_1), \dots, \mu^T(\mathbf{s}_n))^T$, $\{\Gamma_Y(\phi_\theta)\}_{j,k} = \rho_\theta(\mathbf{s}_j - \mathbf{s}_k; \phi_\theta)$ and $\{\rho_{0,Y}(\phi_\theta)\}_j = \rho_\theta(\mathbf{s}_0 - \mathbf{s}_j; \phi_\theta)$, $j, k = 1, \dots, n$. Thus, the conditional distribution for $\mathbf{Y}(\mathbf{s}_0) | \mathbf{Y}^*$ is a bivariate normal with mean $E_{\mathbf{s}_0}$ and covariance matrix $\Sigma_{\mathbf{s}_0}$, where

$$\begin{aligned} E_{\mathbf{s}_0} &= \mu(\mathbf{s}_0) - \mu_{0,Y}^T(\phi_\theta) \otimes T \cdot \Gamma_Y^{-1}(\phi_\theta) \otimes T^{-1}(\mathbf{Y}^* - \mu^*), \\ \Sigma_{\mathbf{s}_0} &= (1 - \rho_{0,Y}^T(\phi_\theta) \Gamma_Y^{-1}(\phi_\theta) \rho_{0,Y}(\phi_\theta)) \otimes T. \end{aligned}$$

In fact, the conditional distribution for $\Theta(\mathbf{s}_0) | \mathbf{Y}^*$ is a general projected normal $PN_2(E_{\mathbf{s}_0}, \Sigma_{\mathbf{s}_0})$. The latent variable $R(\mathbf{s}_i)$ associated with the i -th location is updated during the model fitting. At the g -th iteration, we gather the posterior samples of $\mathbf{Y}^*$ through $y_1(\mathbf{s}_i)^{(g)} = r(\mathbf{s}_i)^{(g)} \cos \theta(\mathbf{s}_i)$ and $y_2(\mathbf{s}_i)^{(g)} = r(\mathbf{s}_i)^{(g)} \sin \theta(\mathbf{s}_i)$, $i = 1, \dots, n$ and $g = 1, \dots, G$. Finally, we are able to draw samples of $\Theta(\mathbf{s}_0)$ from the predictive distribution $f(\theta(\mathbf{s}_0) | \theta, \mathbf{h})$, since there exists an explicit form for $\Theta(\mathbf{s}_0) | \mathbf{Y}^*$, equivalently, $\Theta(\mathbf{s}_0) | \theta, \mathbf{R}$. As a byproduct of sampling $\Theta(\mathbf{s}_0)$, we also obtain a sample of $R(\mathbf{s}_0)$ at the g -th iteration, $r(\mathbf{s}_0)^{(g)}$. Let $\mathbf{r}^{(g)}$ be the realization of $R = (R(\mathbf{s}_1), \dots, R(\mathbf{s}_n))^T$ at the g -th iteration. We can evaluate $f(\theta(\mathbf{s}_0) | \mathbf{r}^{(g)}, \Psi^{(g)}, \theta)$ using a fine grid of points

on $[0, 2\pi)$ and take the average of the density values on each grid. This resulting average is a usual Rao-Blackwellized estimate of the predictive density at the kriged location $\mathbf{s}_0$.

For kriging of the linear variable at the new location, $H(\mathbf{s}_0)$, again we simplify, starting from the joint distribution of $H(\mathbf{s}_0)$ and $\mathbf{H} = (H(\mathbf{s}_1), \dots, H(\mathbf{s}_n))^T$,

$$\begin{pmatrix} H(\mathbf{s}_0) \\ \mathbf{H} \end{pmatrix} \sim MVN \left(\begin{pmatrix} \mathbf{X}(\mathbf{s}_0)^T \boldsymbol{\beta} \\ \mathbf{X}\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \tau_h^2 + \sigma_h^2 & \tau_h^2 \rho_{0,\mathbf{H}}^T(\phi_h) \\ \tau_h^2 \rho_{0,\mathbf{H}}(\phi_h) & \tau_h^2 \Gamma_{\mathbf{H}}(\phi_h) + \sigma_h^2 I_n \end{pmatrix} \right),$$

where $\mathbf{X}(\mathbf{s}_0) = (1, R(\mathbf{s}_0) \cos \Theta(\mathbf{s}_0), R(\mathbf{s}_0) \sin \Theta(\mathbf{s}_0))^T$, $\{\Gamma_{\mathbf{H}}(\phi_h)\}_{j,k} = \rho_h(\mathbf{s}_j - \mathbf{s}_k; \phi_h)$, $\{\rho_{0,\mathbf{H}}(\phi_h)\}_j = \rho_h(\mathbf{s}_0 - \mathbf{s}_j; \phi_h)$, $j, k = 1, \dots, n$ and $\mathbf{X}$ is a $n \times 3$ matrix with the i -th row as $(1, R(\mathbf{s}_i) \cos \Theta(\mathbf{s}_i), R(\mathbf{s}_i) \sin \Theta(\mathbf{s}_i))$, $i = 1, \dots, n$. At the g -th iteration, the posterior samples, $\theta(\mathbf{s}_0)^{(g)}$ and $r(\mathbf{s}_0)^{(g)}$, are included in $\mathbf{x}(\mathbf{s}_0)^{(g)}$ as the covariates. The remaining kriging steps for the linear variable $H(\mathbf{s}_0)$ are standard.

We can easily infer about the mean direction and the mean resultant length of the predictive density of Θ as well as the mean of the predictive density of H at a new location $\mathbf{s}_0$, enabled by MCMC Bayesian model fitting using posterior samples. According to the definition, the mean direction is $\omega = \arctan^*(S/C)$ and the resultant length is $\gamma = \sqrt{C^2 + S^2}$, where $C = \mathbb{E} \cos \Theta$ and $S = \mathbb{E} \sin \Theta$. Thus, for each posterior draw, $\Psi^{(g)}$ and $\mathbf{r}^{(g)}$, we draw a posterior sample $\mathbf{y}(\mathbf{s}_0)^{(g)}$ of $\mathbf{Y}(\mathbf{s}_0)$ from $N_2(E_{\mathbf{s}_0}^{(g)}, \Sigma_{\mathbf{s}_0}^{(g)})$, and then convert to $\theta(\mathbf{s}_0)^{(g)}$. We compute $\bar{C}(\mathbf{s}_0) = \Sigma_g \cos \theta(\mathbf{s}_0)^{(g)} / G$ and $\bar{S}(\mathbf{s}_0) = \Sigma_g \sin \theta(\mathbf{s}_0)^{(g)} / G$. Then, $\hat{\omega}(\mathbf{s}_0) = \arctan^*(\bar{S}(\mathbf{s}_0) / \bar{C}(\mathbf{s}_0))$ and $\hat{\gamma}(\mathbf{s}_0) = \sqrt{\bar{C}^2(\mathbf{s}_0) + \bar{S}^2(\mathbf{s}_0)}$.

4 A data example

We study wave heights and directions in the Adriatic Sea, a semi-enclosed basin. The orography of the Adriatic Sea plays a key role in associated wave behavior as most of the waves tend to travel from north-northwest in a south-easterly direction along the major axis of the basin. Monitored buoy data, supplying wave height and wave direction measurements would seem to be attractive for such analysis. However, at present, buoy networks are too sparse to be used as a data source for spatial analysis; currently, they also yield unreliable angular measurements. Instead, we employ output from a deterministic model providing climatic forecasts at several *spatial* and *temporal* resolutions. These models are increasingly accurate and may be eventually calibrated using buoy data, a data fusion problem.

In particular, we use data outputs from a deterministic wave model implemented by ISPRA (Istituto Superiore per la Protezione). Details about this deterministic model can be found on their web page (http://www.isprambiente.gov.it/pre_mare/wam.htm). ISPRA outputs are given in deep waters with a spatial resolution of 0.1 degree longitude, approximately on a grid with about 12.5×12.5 km cells. The ISPRA dataset has observations on a total of 1494 locations. We note

that, though these data are not *random*, we model them stochastically in order to enable kriging and forecasting with uncertainty given them.

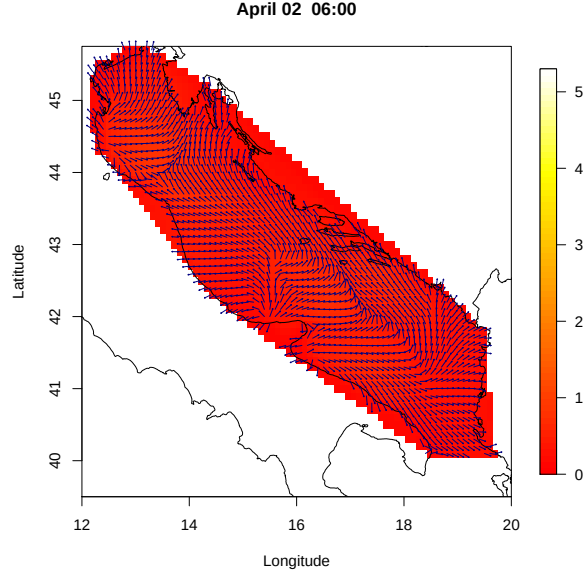


Fig. 2 Plot of wave heights and wave directions under *calm* sea motion status

According to meteorologists, the definition of “calm” in the Adriatic Sea refers to a time period with wave height lower than 1 meter. When the wave height is between 1 and 2 meters, a pre-storm warning is triggered and above 2 meters there is a storm warning. We plot the data, wave directions and wave heights, for an illustrative hour under each sea state, in Figures 2-4, with arrow plots for the wave directions and image plot (color) for the wave heights. We apply the joint modeling framework proposed in Section 2 separately to each of the three datasets and we employ the joint model in the simplest case. There are 1494 locations and we randomly select 300 (approximately 20%) locations for the purpose of validation. For the fitting, we run the MCMC algorithm for 25000 iterations with a burn-in of 5000 and thinning by every fifth sample.

The prior setting for Ψ_θ is as follows: $\mu \sim N(\mathbf{0}, 4\mathbf{I})$, $\tau_\theta^2 \sim IG(2, 1)$ and $\rho \sim \text{Unif}(-1, 1)$. A continuous uniform prior $\text{Unif}(0.0004, 0.1)$ is used for ϕ_θ , corresponding to a maximum and minimum range of 7500 km and 30 km respectively. In fact, the strong spatial dependence for the directions under storm conditions suggest that we may require a range beyond the largest pairwise distance in the dataset when there is little variation in wave directions. For the parameters associated with the heights $\Psi_h = \{\beta, \tau_h^2, \sigma_h^2, \phi_h\}$, a continuous uniform prior $\text{Unif}(0.005, 0.5)$ is used for ϕ_h , corresponding to a maximum and minimum range of 600 km and 60 km

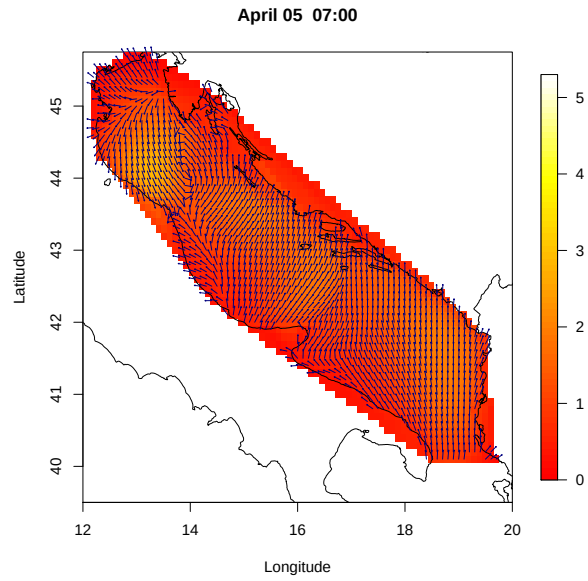


Fig. 3 Plot of wave heights and wave directions under *transition* sea motion status

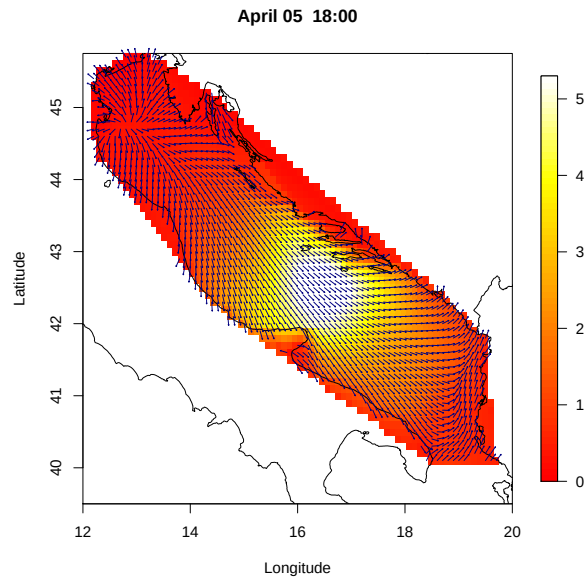


Fig. 4 Plot of wave heights and wave directions under *storm* sea motion status

respectively. Since ϕ_h is the decay parameter in the regular linear spatial regression model, we can adopt customary priors for ranges with linear variables. We employ a non-informative prior for β as $\beta \sim N_3(\mathbf{0}, 10\mathbf{I})$. Inverse gamma priors are used for τ_h^2 and σ_h^2 , respectively, $IG(2, 0.3)$ and $IG(2, 0.75)$.

We compare our results with those for independently modeling the spatial wave height and the spatial wave direction by setting β_1 and β_2 to be zero. In the independence case, the marginal model for wave heights is a spatial regression with constant mean β_0 , written as

$$H(\mathbf{s}) = \beta_0 + w(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad w(\mathbf{s}) \sim GP(\mathbf{0}, \sigma_h^2 \rho_h(\mathbf{s} - \mathbf{s}'; \phi_h)), \quad \varepsilon(\mathbf{s}) \stackrel{\text{iid}}{\sim} N(0, \tau_h^2),$$

where the parameters set is $\Psi_{hm} = \{\beta_0, \phi_h, \sigma_h^2, \tau_h^2\}$. The marginal model for wave directions becomes a constant mean projected Gaussian process model.

The posterior summaries of parameters are shown in Table 1, for both the joint model and the independence model above. Model comparison between the joint model and the independence model is provided in Table 2. Regarding the model comparison criterion for circular variables, we utilize the continuous ranked probability score (CRPS) and the predictive log scoring loss (PLSL), following the discussion [8]. For the linear variable, predictive mean square error and average length of 95% credible intervals are evaluated for each model.

During a calm day, we observe small values of wave heights (less than 1 meter) and large variation in wave directions over the region of interest. Conversely, at a time point during the storm, wave directions are fairly homogeneous while there is a large variation in wave heights over the region (ranging from 1 meter up to 6 meters). During the transition period, there is considerable variation both in wave heights and wave directions. From Table 1, the posterior means of the spatial variance component for wave heights under the joint model (σ_h^2) are 0.0051, 0.0344 and 0.0702 for the three different datasets. Since the contribution of τ_h^2 is relatively negligible, these differences across time points agree with our remark regarding increasing variation of wave heights from calm to storm. For the wave directions, perhaps there is no single parameter which can capture the variation across the region. Specifically, it is hard to glean much from the posterior summaries of the parameters of the marginal projected normal distribution since the four parameters of a general projected normal altogether determine the shapes in a complex fashion. However, the decay parameter ϕ_θ associated with the spatial correlation kernel of the projected Gaussian process provide some indication of the smoothness of the surface of directions. For the calm day, ϕ_θ is much larger than in the other two cases, indicating less smoothness in this surface, as expected.

Finally, we discuss the model comparison results in Table 2. During the calm days, the model comparison criteria find the performance of the two models to be essentially equivalent. For the transition case, the joint model emerges as slightly better, with shorter average length of 95% credible intervals and slightly smaller average CPRS and PLSL. For the time point chosen during a storm, the biggest gain is achieved. We have a substantially narrower average length of credible intervals and a roughly 5% smaller value of CRPS. As mentioned above, the independent

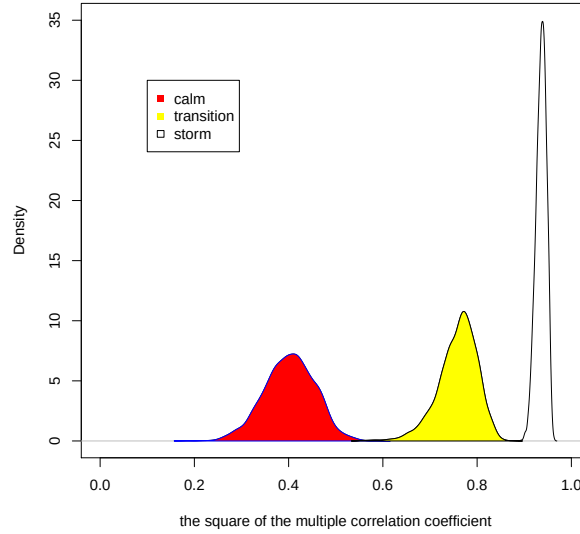
Table 1 Posterior summaries of the parameters: joint model (H, Θ) and independent models H, Θ

parameter	mean	2.5%	97.5%	mean	2.5%	97.5%
data (Calm)	(H, Θ)			H, Θ		
β_0	0.2446	0.1925	0.2918	0.2331	0.1836	0.2800
β_1	0.0125	-0.0242	0.0446	-	-	-
β_2	0.0653	0.0502	0.0800	-	-	-
σ_h^2	0.0051	0.0046	0.0057	0.0055	0.0051	0.0060
τ_h^2	0.0011	0.0011	0.0013	0.0011	0.0010	0.0013
ϕ_h	0.0101	0.0084	0.0126	0.0126	0.0105	0.0116
ϕ_θ	0.0020	0.0019	0.0022	0.0019	0.0017	0.0022
μ_1	0.2856	-0.4034	0.9887	0.3039	-0.4285	1.0073
μ_2	-0.5055	-1.8380	0.7821	-0.4141	-1.7264	0.9531
τ_θ^2	0.2717	0.2314	0.3242	0.2747	0.2392	0.3205
ρ	-0.4874	-0.5609	-0.4107	-0.4790	-0.5537	-0.4110
data (Transition)	(H, Θ)			H, Θ		
β_0	0.4660	0.3251	0.6138	0.5394	0.3651	0.6882
β_1	0.1887	0.1246	0.2432	-	-	-
β_2	0.3062	0.2448	0.3688	-	-	-
σ_h^2	0.0344	0.0295	0.0381	0.0413	0.0379	0.0487
τ_h^2	0.0021	0.0019	0.0024	0.0021	0.0019	0.0024
ϕ_h	0.0101	0.0084	0.0126	0.0084	0.0072	0.0090
ϕ_θ	0.0006	0.0005	0.0007	0.0006	0.0005	0.0007
μ_1	-0.1066	-2.3993	2.1268	0.0617	-2.1378	2.1864
μ_2	-0.2935	-1.8709	1.3000	-0.2152	-1.8383	1.4209
τ_θ^2	2.4400	2.1205	2.9033	2.2318	1.9467	2.5801
ρ	-0.2264	-0.3069	-0.1467	-0.2257	-0.3084	-0.1454
data (Storm)	(H, Θ)			H, Θ		
β_0	0.3795	0.1754	0.5712	0.4737	0.0414	0.8596
β_1	-0.3052	-0.4035	-0.2046	-	-	-
β_2	0.8794	0.7580	0.9876	-	-	-
σ_h^2	0.0702	0.0551	0.0835	0.1930	0.1687	0.2302
τ_h^2	0.0033	0.0028	0.0038	0.0034	0.0029	0.0040
ϕ_h	0.0105	0.0087	0.0122	0.0070	0.0063	0.0076
ϕ_θ	0.0006	0.0005	0.0007	0.0006	0.0005	0.0007
μ_1	0.5853	-0.8101	1.9415	0.6990	-0.6806	1.9630
μ_2	-0.6064	-2.0839	0.9323	-0.5152	-2.0437	0.9872
τ_θ^2	0.7216	0.6562	0.8090	0.7312	0.6536	0.8054
ρ	-0.6433	-0.6873	-0.6039	-0.6631	-0.7107	-0.6095

Table 2 Model comparison: joint model (H, Θ) and independent models H, Θ

	(a) calm		(b) transition		(c) storm	
Feature	(H, Θ)	H, Θ	(H, Θ)	H, Θ	(H, Θ)	H, Θ
Predictive Mean Square Error (height)	0.0006	0.0006	0.0031	0.0030	0.0109	0.0108
Average Length of 95% Credible Interval (height)	0.1821	0.1788	0.3405	0.3586	0.5163	0.6345
mean CRPS for wave direction	0.0407	0.0408	0.0276	0.0279	0.0213	0.0223
PLSL for wave direction	-977	-974	-1098	-1104	-1321	-1318

model is nested in the joint model framework. The posterior summaries of β_1 and β_2 in Table 1 concur with these findings. For the time point during a calm day, β_1 and β_2 are nearly zero, which supports the independence of wave heights and wave directions. As we move from calm to transition to storm, the β_1 and β_2 become increasingly different from 0. The multiple correlation coefficient is computed, as in (2) and we show the posterior densities of the multiple correlation coefficient for all three scenarios in Figure 5. The figure concisely supports Table 1, revealing stronger association between height and direction as we move from calm to storm. Finally, we note that the negative β_1 and positive β_2 are not unexpected. With storms tending to travel from the north-northwest to the southeast, and with the $\sin$ increasing and $\cos$ decreasing in these directions, this supplies an expectation of an increase in wave height as a function of θ as the storm develops.

**Fig. 5** The posterior densities of the square of the multiple correlation coefficient $R^2_{H|Y}$

5 Joint space-time models

In the previous section, we modeled spatial wave directions and wave heights at a static time slice. It is natural to envision an underlying process for heights and directions in continuous space and time; fitting such a model would enable both interpolation and forecasting at future time points. In fact, such data are also available across time at hourly resolution using the ISPRA output; here we suggest a natural extension to accommodate such data. Now, wave heights and wave directions at location $\mathbf{s}$ and time t are denoted as $H(\mathbf{s}, t)$ and $\Theta(\mathbf{s}, t)$ both over both the domain, $\mathbf{s} \in \mathcal{D}$ and $t \in (0, K)$. We adopt the same joint framework shown in expression (1), where Ψ_h and Ψ_θ are sets of parameters associated with the conditional model for height and the marginal model for direction, respectively.

Again, this joint space-time model is provided conditionally. For the marginal distribution of the circular variables $\Theta(\mathbf{s}, t) | \Psi_\theta$, we propose a stationary spatio-temporal projected Gaussian process with a constant mean $\mu = (\mu_1, \mu_2)^T$ and separable cross-covariance form,

$$C_\theta((\mathbf{s}, t), (\mathbf{s}', t')) = \rho_{\theta, \mathbf{s}}(\mathbf{s} - \mathbf{s}'; \phi_{\theta, \mathbf{s}}) \rho_{\theta, t}(t - t'; \phi_{\theta, t}) \cdot T, \quad (5)$$

where $\rho_{\theta, \mathbf{s}}$ is the spatial correlation and $\rho_{\theta, t}$ is the temporal correlation. The parameters associated with the directions $\Psi_\theta = (\mu, T, \phi_{\theta, \mathbf{s}}, \phi_{\theta, t})$, $\phi_{\theta, \mathbf{s}}$ and $\phi_{\theta, t}$ are the decay parameters associated with their corresponding correlation functions $\rho_{\theta, \mathbf{s}}$ and $\rho_{\theta, t}$ and T is still defined as in the previous section, $T = \begin{pmatrix} \tau_\theta^2 & \rho \tau_\theta \\ \rho \tau_\theta & 1 \end{pmatrix}$. This model has been considered in [9].

Again, we only need one space-time covariance function. Under this model specification, the projected Gaussian process is induced from an inline Gaussian process $\mathbf{Y}(\mathbf{s}, t)$ with a constant mean μ and cross-covariance $C_\theta((\mathbf{s}, t), (\mathbf{s}', t'))$. So, we can transform back and forth between the spaces for the pairs of random variables, $(\Theta(\mathbf{s}, t), R(\mathbf{s}, t))$ and $(Y_1(\mathbf{s}, t), Y_2(\mathbf{s}, t))$ through $Y_1(\mathbf{s}, t) = R(\mathbf{s}, t) \cos \Theta(\mathbf{s}, t)$ and $Y_2(\mathbf{s}, t) = R(\mathbf{s}, t) \sin \Theta(\mathbf{s}, t)$.

At the conditional level, we introduce

$$H(\mathbf{s}, t) = \mathbf{X}(\mathbf{s}, t)^T \beta + w(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t) = \beta_0 + \beta_1 Y_1(\mathbf{s}, t) + \beta_2 Y_2(\mathbf{s}, t) + w(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t). \quad (6)$$

Suppose we simplify $w(\mathbf{s}, t)$ to $w(\mathbf{s})$, a zero-centered GP with covariance function $C_h = \sigma_h^2 \rho_h(\mathbf{s} - \mathbf{s}'; \phi_h)$ and the random the error term $\varepsilon(\mathbf{s}, t) \stackrel{\text{iid}}{\sim} N(0, \tau_h^2)$. Then, the model asserts that the temporal dependence between $H(\mathbf{s}, t)$ is fully contributed by the latent components $Y_1(\mathbf{s}, t)$ and $Y_2(\mathbf{s}, t)$.

References

1. Bulla, J., Lagona, F., Maruotti, A., Picone, M.: A Multivariate Hidden Markov Model for the Identification of Sea Regimes from Incomplete Skewed and Circular Time Series. *J. Agric. Biol. Envir. S.* **17**, 544–567 (2012)
2. Di Marzio, M., Panzera, A., Taylor, C.C.: Non-parametric Regression for Circular Responses. *Scand J. Stat.* **40**, 238–255 (2013)
3. Fisher, N.I., Lee, A.J.: Regression Models for an Angular Response. *Biometrics.* **48**, 665–677 (1992).
4. Johnson, R.A., Wehrly, T.E.: Measures and Models for Angular Correlation and Angular-Linear Correlation. *J. Roy. Stat. Soc. B. Met.* **39**, 222–229 (1977)
5. Jona Lasinio, G., Gelfand, A.E., Jona Lasinio, M.: Spatial Analysis of Wave Direction Data Using Wrapped Gaussian Processes. *Ann. Appl. Stat.* **6**, 1478–1498 (2012)
6. Lagona, F., Picone, M.: Maximum Likelihood Estimation of Bivariate Circular Hidden Markov Models from Incomplete Data. *J. Stat. Comput. Sim.* **83**, 1223–1237 (2013)
7. Mardia, K.V.: Linear-Circular Correlation Coefficients and Rhythmometry. *Biometrika.* **63**, 403–405 (1976)
8. Wang, F., Gelfand, A.E.: Directional Data Analysis under the General Projected Normal Distribution. *Stat. Methodol.* **10**, 113–127 (2013a)
9. Wang, F., Gelfand, A.E.: Modeling Space and Space-time Directional Data Using Projected Gaussian Processes. Tech. rep. Duke University (2013b)