

Models for space-time directional data using Wrapped Gaussian processes

Modelli spazio temporali per dati direzionali basati su processi Gaussiani arrotondati

Giovanna Jona Lasinio, Gianluca Mastrantonio and Alan E. Gelfand

Abstract In the present work we review modeling strategies based on wrapped Gaussian processes defined to model directional spatio-temporal data. We first illustrate the model-based approach to handle spatial periodic data. The wrapped Gaussian spatial process is here induced by a customary linear Gaussian process. We formulate the model as a Bayesian hierarchical one and we show that the fitting of the model is possible using standard Markov chain Monte Carlo methods. Then we move to some spatio-temporal generalizations of the spatial model. In the spatio-temporal setting we present a simulation study of our proposal aiming at understanding its computational and statistical properties. We highlight the pros and cons of this model and the difficulties arising in the implementation of the MCMCs. Eventually we provide some general advice in the use of spatio-temporal wrapped Gaussian process and we provide a real data example.

Abstract In questo lavoro si discutono approcci per la modellistica di dati circolari spazio-temporali basati su processi gaussiani arrotondati. Introduciamo l'approccio arrotondato nel caso di processi gaussiani spaziali. Il processo gaussiano spaziale arrotondato è indotto da un processo lineare gaussiano, in questo contesto presentiamo il modello in una formulazione gerarchica bayesiana mostrando come la stima del modello sia possibile tramite metodi Markov chain Monte Carlo standard. Quindi procediamo alla generalizzazione spazio-temporale del modello e presentiamo uno studio su dati simulati con lo scopo di definire vantaggi e svantaggi di questo approccio. Infine proponiamo un esempio su dati reali

Giovanna Jona Lasinio

Sapienza University of Rome, P.le Aldo Moro 5, 00185 Roma, e-mail: giovanna.jonalasinio@uniroma1.it

Gianluca Mastrantonio

University of ROMATRE, Via Silvio D'Amico 77, 00145 Rome e-mail: gianluca.mastrantonio@uniroma3.it

Alan E. Gelfand

223-A Old Chemistry Building Box 90251 Duke University Durham, NC 27708-0251 USA
e-mail: alan@stat.duke.edu

Key words: Wrapped Gaussian processes, directional data, Bayesian hierarchical models.

1 Introduction

Circular data, i.e. observations recorded as direction, arise in many different context, examples of directional data include natural directions, such as wind directions (meteorology), animal movement directions (biology) and rock fracture orientations (geology). Another type of directional data arises by wrapping periodic time data with period L (say, daily or weekly) onto a circle with circumference L and then rescaling the circumference to 2π , that of a unit circle. As with data on the real line, directional data can also arise in the space and space-time setting. Analysis of circular data can be difficult due to the restriction of the domain on the circle. With such data, well know statistical indexes can loose their meaning and should be replaced with their circular counterpart, for example the mean is replaced by the circular mean [14]. Most of the general inferential tools have been adapted to work with circular variables, see for example the books by [20], [7], [14] or the paper by [19]. There is a growing number of papers on the modeling of directional data since the 1990's, examples include linear models [11, 7, 17], linear models in a Bayesian context [10, 5], models for circular time series [2, 4, 20, 21, 13, 8, 12] and hidden Markov models to address classification issues [18, 3]. In [16] a Markov process for circular variables is presented, while in [24] the Bayesian analysis of space-time directional data is developed using projected normal distributions. In this paper we discuss the modeling of space-time directional data under the wrapped Gaussian process developed in [15] in a purely spatial setting.

2 Sketching the wrapped modeling approach

To introduce the wrapping approach let us consider simple univariate representation, then the wrapping approach can be formalized as follows. Let Y be a real valued random variable on $\mathbb{R}$, henceforth referred to as a *linear* random variable, with probability density function $g(y)$ and distribution function $G(y)$. The induced wrapped variable (X) of period 2π , is given by

$$X = Y \bmod 2\pi. \quad (1)$$

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

$$f(x) = \sum_{k=-\infty}^{\infty} g(x + 2k\pi), \quad 0 \leq x < 2\pi \quad (2)$$

From (2), we see that the joint distribution of (X, K) is $g(x + 2k\pi)$ with $x \in [0, 2\pi)$ and $k \in \mathbb{Z} \equiv \{0, \pm 1, \pm 2, \dots\}$. Then, marginalization over k produces (2). Expressed differently, $g(\cdot)$ is the distribution of $Y = X + 2K\pi$, Y determines X and K through (1), and X is a wrapped version of Y . 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.

Moving to the multivariate setting, as a general definition, it is perhaps easiest to obtain a multivariate wrapped distribution for say $\mathbf{X} = (X_1, X_2, \dots, X_p)$ starting with a multivariate linear distribution for $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$. In particular, suppose $\mathbf{Y} \sim g(\cdot)$ where $g(\cdot)$ is a p -variate distribution on $\mathbb{R}^p$. Usually, g is a family of distributions indexed by say, θ ; a convenient choice for $g(\cdot)$ is a p -variate normal distribution. Let $\mathbf{K} = (K_1, K_2, \dots, K_p)$ be such that $\mathbf{Y} = \mathbf{X} + 2\pi\mathbf{K}$, analogous to the univariate case developed above. Then, the joint distribution of $\mathbf{X}$ and $\mathbf{K}$ is $g(\mathbf{x} + 2\pi\mathbf{k})$ for $0 \leq x_j < 2\pi, j = 1, 2, \dots, p$ and $k_j \in \mathbb{Z}, j = 1, 2, \dots, p$. The marginal distribution of $\mathbf{X}$ is, directly

$$\sum_{k_1=-\infty}^{+\infty} \sum_{k_2=-\infty}^{+\infty} \dots \sum_{k_p=-\infty}^{+\infty} g(\mathbf{x} + 2\pi\mathbf{k}) \quad (3)$$

Again we introduce latent K_j 's to facilitate the model fitting. We say that $\mathbf{X}$ has a p -variate wrapped normal distribution (WN) when $g(\cdot; \theta)$ is a multivariate normal where $\theta = (\mu, \Sigma)$, with μ a $p \times 1$ vector of mean directions and Σ is a positive definite matrix. Using standard results, the conditional distribution of Y_j given $\{Y_l, l \neq j\}$ and θ is immediate, hence, as well, the distribution of X_j, K_j given $\{X_l, K_l, l \neq j\}$ and θ . The only serious issue that appears is that the forms in (2) and (3) are intractable to work with as they are. In [15] it is shown how to truncate the series when $g(\cdot)$ is Gaussian based on distribution variability.

It is now clear that in the wrapped approach introducing dependence structures among variables is straight forward starting from the linear variable definition. Indeed a Gaussian process on $\mathbb{R}^2$ induces a wrapped Gaussian process on $\mathbb{R}^2$. 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. In particular, if $Y(\mathbf{s})$ is a GP with mean $\mu(\mathbf{s})$ and covariance function, say $\sigma^2 \rho(s - s'; \psi)$ where ψ is a decay parameter, then, for locations $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$, $\mathbf{X} = (X(\mathbf{s}_1), X(\mathbf{s}_2), \dots, X(\mathbf{s}_n)) \sim WN(\mu, \sigma^2 \mathbf{R}(\psi))$ where $\mu = (\mu(\mathbf{s}_1), \dots, \mu(\mathbf{s}_n))$ and $R(\psi)_{ij} = \rho(\mathbf{s}_i - \mathbf{s}_j; \psi)$ [15]). In this framework if now we assume that $X(\mathbf{s}, t) \equiv \{x_t(\mathbf{s}) \in [0, 2\pi) : \mathbf{s} = (s_1, s_2) \in \mathcal{S} \subseteq \mathbb{R}^2, t \in \mathcal{T}\}$ is a spatio-temporal process of angular measures (in radiant) evolving in space and time. We can model $X(\mathbf{s}, t)$ as a spatio-temporal Wrapped Gaussian process through the relation between X and its

linear counterpart $Y(\mathbf{s}, t)$ extending the above approach to the spatio-temporal setting. In general we can say that given a linear spatio-temporal Gaussian process with mean μ_Y and covariance $\Sigma_Y = \sigma^2 \mathbf{R}$, with $\mathbf{R}(\phi)$ a given space-time correlation function parametrized by ϕ , it induces a Wrapped spatio-temporal Gaussian process $\mathbf{X} = \{X(\mathbf{s}, t), t \in \mathcal{T}, \mathbf{s} \in \mathcal{S}\} \sim WN(\mu_X, \sigma^2 \mathbf{R}(\phi))$, where $\mu_X = \ell(\mu_Y)$ with ℓ an appropriate link function such as the modulus or $\arctan^*$ ¹ [8] and the induced covariance structure is obtained using one of the many proposal that can be found in the literature, for example the $\frac{\sinh \Sigma_Y^{ij}}{\sinh(\sqrt{\Sigma_Y^{ii}} \sqrt{\Sigma_Y^{jj}})}$ function (see [15] and references therein)

that allows us to obtain a valid covariance function. Once prior distributions are chosen, estimation in the Bayesian setting, can be carried out using standard MCMC tools. To perform prediction with a space-time wrapped GP we have the same difficulty highlighted in the purely spatial context: to compute the predictive distribution we have to sum over the set of winding numbers K and this is unfeasible even with a problem of small dimensions. However if $(\mathbf{s}_0, t_0)$ is a new point in $\mathbb{R}^d \times \mathbb{R}$ and we want to assess information on $X(\mathbf{s}_0, t_0)$ within the Bayesian modeling framework, we seek the average of the conditional distribution of $X(\mathbf{s}_0, t_0)$ given the observed values. Fitting the space-time wrapped Gaussian process will yield posterior samples of parameters of the model θ , say $(\theta_b^*, \mathbf{K}_b^*), b = 1, 2, \dots, B$. Then as in the spatial setting we can compute Monte Carlo approximations of the desired mean.

3 Model details and model fitting

We aim at getting some insight on the behaviour of space-time wrapped models and on the MCMC implementation necessary to obtain parameters posterior estimates. To illustrate these developments we use a relatively simple model, we assume that the linear variable is a spatio-temporal Gaussian process (GP) with constant mean and non-separable covariance structure, in particular we adopt the following covariance function (see equation (14) in [9]):

$$\text{Cov}(\mathbf{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^\tau} \exp\left(-\frac{c\|\mathbf{h}\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right), (\mathbf{h}; u) \in \mathbb{R}^d \times \mathbb{R} \quad (4)$$

where $\|\mathbf{h}\|$ is the distance between two locations in space, $|u|$ is the time lag, here $d = 2$, a and c are non negative scaling parameters of time and space, respectively and the smoothness parameters α and γ take values in $(0, 1]$ and the space-time interaction parameter β in $[0, 1]$, while $\tau \geq d/2$ is here fixed to 1 following [9]. We then write the linear GP $Y(\mathbf{s}, t) = X(\mathbf{s}, t) + K(\mathbf{s}, t)$ as:

$$Y(\mathbf{s}, t) = \mu_Y + w(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t) \quad (5)$$

¹ We use the $\arctan$ definition of [?, page 13]Jammalamadaka2001

where μ_Y is a constant mean, $w(\mathbf{s}, t)$ is a space-time GP with zero mean and covariance function (4) and $\varepsilon(\mathbf{s}, t) \sim N(0, \sigma_\varepsilon^2)$ is an independent random error (nugget effect or measurement error). We justify our choice as it is a “reasonable” compromise between a realistic model (non-separable covariance structure) and complexity. In fact in several applications it is reasonable to assume a constant mean direction and a complex variability structure, for example in modeling wave directions or wind directions when a storm is moving across a large region, the mean direction is constant in subregions. To complete the model specification we need to choose a set of prior distributions. We suggest the following choices:

$$\begin{aligned} a &\sim \text{Gamma}(a_a, b_a); c \sim \text{Gamma}(a_c, b_c) \\ \alpha &\sim \text{Beta}(v_{1,\alpha}, v_{2,\alpha}); \beta \sim \text{Beta}(v_{1,\beta}, v_{2,\beta}); \gamma \sim \text{Beta}(v_{1,\gamma}, v_{2,\gamma}); \\ \sigma^2 &\sim \text{InGamma}(a_\sigma, b_\sigma); \sigma_\varepsilon^2 \sim \text{InGamma}(a_\varepsilon, b_\varepsilon) \\ \mu_Y &\sim N(\eta, \tau^2)I(0, 2\pi) \end{aligned}$$

hyper parameters are chosen to ensure weakly informative distributions,

Let $\phi = (a, c, \alpha, \beta, \gamma)$ be the correlation function parameters vector and $\theta = (\mu_Y, \phi, \sigma_\varepsilon^2, \sigma^2)$ the vector of all model parameters. To perform prediction we need the conditional distribution of new points given the observed once. For the linear process this is a standard task as given a new point $(\mathbf{s}_0, t_0) \in \mathbb{R}^d \times \mathbb{R}$

$$\begin{pmatrix} \mathbf{Y} \\ Y(\mathbf{s}_0, t_0) \end{pmatrix} = N \left(\begin{pmatrix} \mu \\ \mu(\mathbf{s}_0, t_0) \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{R}_Y(\phi) & \rho_{0,Y}(\phi) \\ \rho_{0,Y}^T(\phi) & 1 \end{pmatrix} + \sigma_\varepsilon^2 \mathbf{I} \right). \quad (6)$$

and then $Y(\mathbf{s}_0, t_0) | \mathbf{Y} \sim N(\mu_0(\mathbf{Y}, \theta), \sigma_0^2(\theta))$ where $\mu_0(\mathbf{Y}, \theta)$ and $\sigma_0^2(\theta)$ are the conditional mean ($E(Y(\mathbf{s}_0, t_0) | \mathbf{Y})$) and variance ($\text{Var}(Y(\mathbf{s}_0, t_0) | \mathbf{Y})$). The induced circular variable is such that $X(\mathbf{s}_0, t_0) | \mathbf{X}, \mathbf{K}, \theta$ is wrapped normal. To compute the above mentioned Monte Carlo approximation of the conditional circular mean we can adopt the same approach as in [15] using the complex representation of the circular variable [?, see]]Jammalamadaka2001), $E(e^{iX(\mathbf{s}_0, t_0)} | \mathbf{X}) = E_{\mathbf{K}, \theta | \mathbf{X}}(e^{iX(\mathbf{s}_0, t_0)} | \mathbf{X}, \mathbf{K}, \theta)$ and so a Monte Carlo integration yields

$$E(e^{iX(\mathbf{s}_0, t_0)} | \mathbf{X}) \approx \frac{1}{B} \sum_b \exp(-\sigma_0^2(\theta_b^*)/2 + i\tilde{\mu}_0(\mathbf{X} + 2\pi\mathbf{K}_b^*; \theta_b^*)) \quad (7)$$

where $\tilde{\mu}_0(\mathbf{Y}; \theta) = \mu_0(\mathbf{Y}, \theta) \bmod 2\pi$. Then if $g_{0,c}(\mathbf{X}) = \frac{1}{B} \sum_b \exp(-\sigma_0^2(\theta_b^*)/2) \cos(\tilde{\mu}_0(\mathbf{X} + 2\pi\mathbf{K}_b^*; \theta_b^*))$ and $g_{0,s}(\mathbf{X}) = \frac{1}{B} \sum_b \exp(-\sigma_0^2(\theta_b^*)/2) \sin(\tilde{\mu}_0(\mathbf{X} + 2\pi\mathbf{K}_b^*; \theta_b^*))$, the posterior kriged mean direction and concentration are $\mu(\mathbf{s}_0, t_0, \mathbf{X}) = \arctan^*(g_{0,s}(\mathbf{X}), g_{0,c}(\mathbf{X}))$ and $c(\mathbf{s}_0, t_0, \mathbf{X}) = \sqrt{(g_{0,c}(\mathbf{X}))^2 + (g_{0,s}(\mathbf{X}))^2}$.

3.1 Implementation and computational issues

To obtain posterior samples of the model parameters, we use a Metropolis within Gibbs [22] MCMC algorithm. In the Gibbs step the mean level μ_Y , the random effect $w(s, t)$ and the variance of the nugget, σ_ϵ^2 are sampled from their full conditional distributions. All the parameters of the correlation function, together with σ^2 , are simulated in block in a Metropolis step with a Multivariate Normal proposal. The covariance matrix of the proposal is tuned using the past samples of the chains, as proposed by [1] (algorithm 4). Note that the full conditional of σ^2 is Inverse Gamma and then this parameter could be updated with a Gibbs sampler step, but it is well known, see for example [25], that there are identifiability issues involving both the correlation parameters and the variance σ^2 for Gaussian process. A possible solution is to block-sample the set of parameters in a Metropolis step [6]). The full conditional of the vector of winding numbers has independent components and each component are simulated using the adaptive truncation strategy as proposed by [15]. The computational time is dominated by the inversion of matrices, the covariance matrix of $\mathbf{w}$'s full conditional, that requires $O(n^3)$ operations, where n is the total number of observations, and the evaluation and inversion of the covariance matrix in (4) again of complexity $O(n^3)$ giving an overall complexity of $O(2n^3)$. Note that we could marginalize over $\mathbf{w} = \{w(s, t)\}$ saving $O(n^3)$ operations, but then the full conditional of k 's would have dependent components and its simulation, following [15], would require n inversions of $(n-1) \times (n-1)$ covariance matrices ($O(n(n-1)^3)$). As an alternative to the adaptive truncation we could sample a vector of dependent real valued random variables z associated with the winding numbers via the transformation $k = \lfloor z + 0.5 \rfloor^2$ in a Metropolis step with an adaptive proposal, as for the correlation parameters, that requires the inversion of the covariance matrix of the proposal ($O(n^3)$); in that case the computational complexity is again $O(2n^3)$. Then given the same condition in terms of computational complexity, we prefer to sample $\mathbf{w}$ with a Gibbs sampler without marginalizing it.

3.2 Simulations

To asses model behaviour we generate $N = 600$ (50 locations and 12 time points) data from the linear model in (5) with mean $\mu_Y = \pi$ and all combinations of $(a, c) = \{(0.2, 1), (1, 0.2)\}$, $\alpha = \{0.5, 0.8\}$, $\beta = \{0, 0.5, 0.9\}$, $\gamma = \{0.5, 0.8\}$ and $(\sigma^2, \sigma_\epsilon^2) = \{(0.1, 0.01), (0.5, 0.1)\}$ for the covariance parameters, given a total of 48 different datasets. We obtain the circular values as in (1), spatial coordinates are uniformly sampled from $[0, 10]$. We use 425 points between the 1st and 10th time for estimation and the remaining 225 points for validation purposes. Following [15], to asses model performance we compute an average prediction error (APE), defined as the average

² $\lfloor \cdot \rfloor$ is the floor operation rounding a real number to the smallest closest integer

circular distance³ between observed and estimated values, over the validation set of observed values. The rationale behind these examples is as follows: we distinguish between “small variance” and “medium variance”, i.e. a signal to noise ratio of 0.1 and 0.2 respectively. Then we consider separable ($\beta = 0$) and non separable ($\beta = 0.5, 0.9$) datasets and several levels of smoothness of the covariance function (α, γ). Eventually we consider weak and strong dependence in time and/or space ruled by parameters a and c .

3.3 Real Data

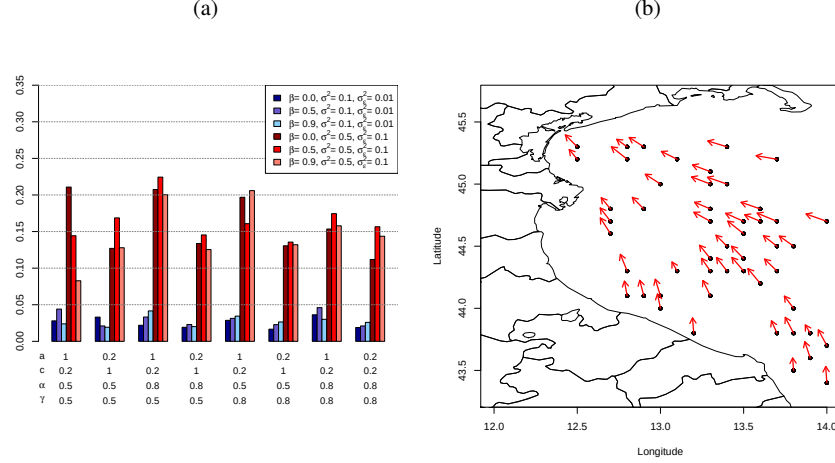
As an example we use data outputs from a deterministic model implemented by ISPRA⁴. The model starts from a wind forecast model predicting the surface wind over the entire Mediterranean. The hourly evolution of sea wave spectra is obtained solving energy transport equations using as input the wind forecast. 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 output is affected by considerable uncertainty that is not directly accounted for. Further more we are interested in changes in the spatial and temporal scale of these data, i.e. we want to estimate values on unobserved points in space and/or time. Among all the observations available, we choose to restrict our attention on the north-west area of the Adriatic sea, we use the wave direction expressed in radiant on 50 locations, Figure 1, from the 00.00 of the 1st to the 18:00 of the 3rd of April 2010, getting values every 6 hours. 425 observations between April 1, 00:00 and April 3, 06:00 are used for model estimation, the remaining, 175 points, are used for model validation. On this same dataset we estimate a model based on the projected normal distribution (PN) [23] as the natural competitor of our proposal that is expected to perform better or equivalently to the WN model. Without going into details, our implementation of the projected is similar to the one proposed by [23] with some modifications: Wang uses a separable covariance structure while here we adopt a correlation structure derived from (4) (divided by σ^2). The parameters priors are set so that their modes are: π for μ_Y , 0.5 for a and c , 0.4 for α and γ , 0.5 for β , 0.5 for σ^2 and 0.1 for σ_e^2 . The latter choices anticipate the presence of a strong spatio-temporal correlation structure, with correlation in space always greater than 0.2 and in time becoming less than 0.05 after about 8 time lags. In the projected normal model the priors structure is the same as in the WN setting for covariance parameters, while the two means of the PN are Gaussian with zero mean and large variance. The correlation between the PN components is Uniform in $(-1, 1)$ and the variance of one component⁵ is inverse gamma with parameters $(2, 2)$.

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

⁴ Istituto Superiore per la Protezione e la Ricerca Ambientale

⁵ the other is fixed equal to one to ensure identifiability

Fig. 1: (a) Simulated examples: Averaged Prediction error (APE), (b) Wave directions in the north east Adriatic sea - The direction of the arrows represent the mean direction and the length the circular variance of the observed direction in the time window under study.



4 Results and Conclusions

The simulated examples show that with non separable or separable data and small variance (see Figure 1 darker bars) prediction is very accurate with $\text{APE} \leq 0.05$ in all considered situations. When the total variance increases ($\sigma^2 + \sigma_\varepsilon^2$, red bars in Figure 1) the accuracy of predictions decreases, with a worst case leading to a maximum value of the APE equal to 0.22. Posterior parameters estimates are always correct including the “true” value in the 95% posterior credibility interval.

Prediction of the real data example is very accurate under the WN model leading to an overall APE of 2.3×10^{-3} . The left hand side of the credibility interval of $\hat{\beta} = 0.70$ (CI= (0.25, 0.93)) is far away from zero suggesting a non separable spatio-temporal covariance structure. The PN model performs in a similar way with an overall APE equal to 1.6×10^{-3} and a $\hat{\beta} = 0.65$ (CI= (0.15, 0.95)) estimate confirming the non separable covariance choice.

As a concluding remark we can say that the PN model is slower than the WN one in reaching convergence while the predictive capability of the two models is equivalent when the data are unimodal. While when data shows multimodal distributions, the PN is expected to perform way better than the WN (examples are now running).

Immediate future developments of the WN model will include a more complex modeling of the process mean. It is reasonable to imagine a mean evolving in time, possibly following the evolution of other variables. For example considering wave directions and waves significant height we can model the mean direction as a re-

gression with significant wave heights or use an analysis of variance approach. In fact according to meteorologists, the three sea status: calm, transition between calm and storm and storm, roughly correspond to levels of wave heights less than 1 meter, between 1 and 2 meters, greater than 2 meters respectively, and it is well known that in each sea state a different level of waves direction is observed. Further investigations will involve the development of non Gaussian wrapped processes.

References

1. Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive mcmc. *Statistical computing*, **18**, 343–373.
2. Breckling, J. (1989). *The analysis of directional time series: applications to wind speed and direction*. Lecture notes in statistics. Springer-Verlag.
3. Bulla, J., Lagona, F., Maruotti, A., and Picone, M. (2012). A multivariate hidden markov model for the identification of sea regimes from incomplete skewed and circular time series. *Journal of Agricultural, Biological, and Environmental Statistics*, **17**(4), 544–567.
4. Coles, S. (1998). Inference for circular distributions and processes. *Statistics and Computing*, **8**(2), 105–113.
5. Damien, P. and Walker, S. (1999). A Full Bayesian Analysis of Circular Data Using the von Mises Distribution. *The Canadian Journal of Statistics*, **27**, 291–298.
6. Finley, A. O., Banerjee, S., and Gelfand, A. E. (2013). spbayes for large univariate and multivariate point-referenced spatio-temporal data models. *Journal of Statistical software* (forthcoming).
7. Fisher, N. I. (1996). *Statistical Analysis of Circular Data*. Cambridge University Press.
8. Fisher, N. I. and Lee, A. J. (1992). Regression Models for an Angular Response. *Biometrics*, **48**(3), 665–677.
9. Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, **97**(458), 590–600.
10. Guttorp, P. and Lockhart, R. A. (1988). Finding the location of a signal: A bayesian analysis. *Journal of the American Statistical Association*, **83**(402), 322–330.
11. Harrison, D. and Kanji, G. K. (1988). The Development of Analysis of Variance for Circular Data. *Journal of Applied Statistics*, **15**, 197–224.
12. Holtzman, H., Munk, A., Suster, M., and Zucchini, W. (2006). Hidden markov models for circular and linear-circular time series. *Environmental and Ecological Statistics*, **13**, 325–347.
13. Hughes, G. (2007). *Multivariate and Time Series Models for Circular Data with Applications to Protein Conformational Angles*. Ph.D. thesis, University of Leeds, Leeds, UK.
14. Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific Pub Co Inc, har/dskt edition.
15. Jona-Lasinio, G., Gelfand, A., and Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped gaussian processes. *Annals of Applied Statistics*, **6**(4).
16. Kato, S. (2010). A markov process for circular data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **72**(5), 655–672.
17. Kato, S. and Shimizu, K. (2008). Dependent models for observations which include angular ones. *Journal of Statistical Planning and Inference*, pages 3538 – 3549.
18. Lagona, F. and Picone, M. (2011). A latent-class model for clustering incomplete linear and circular data in marine studies. *Journal of Data Science*, **9**, 585–605.
19. Lee, A. (2010). Circular data. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**(4), 477–486.
20. Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*. Wiley, 1 edition.

21. Ravindran, P. (2002). *Bayesian Analysis of Circular Data Using Wrapped Distributions*. Ph.D. thesis, North Carolina State University.
22. Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, **22**(4), 1701–1728.
23. Wang, F. (2013). *Space and Space-Time Modeling of Directional Data*. Ph.D. thesis, Duke University.
24. Wang, F. and Gelfand, A. E. (2012). Modeling space and space-time directional data using projected gaussian processes. Technical report, Duke university.
25. Zhang, H. (2004). Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics. *Journal of the American Statistical Association*, **99**(465), 250–261.