Wrapped Gaussian processes: a short review and some new results

Giovanna Jona Lasinio[1], Gianluca Mastrantonio[2] and Alan Gelfand[3]

1-Università Sapienza di Roma 2- Università RomaTRE 3- Duke University

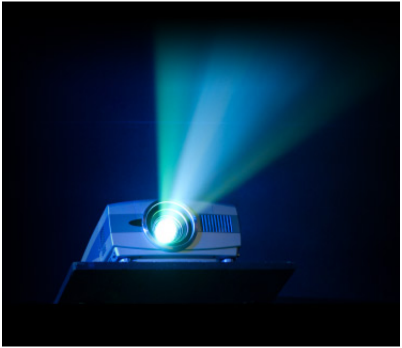
Advances in Directional Statistics
Bruxelles 20-22 May 2014

Topics

1. Introduction
   - Outline
   - Wrapped distributions
2. Spatial and spatio-temporal processes
   - Wrapped Gaussian
   - Projected Normal model
3. Prediction
4. Computational issues
   - Simulation examples
   - Real Data
5. Concluding Remarks
6. Essential References

Aim

- We review modeling strategies based on wrapped Gaussian processes defined to model directional spatio-temporal data.
- We compare the Wrapped Normal approach to Projected Normal models in terms of computational efficiency/convenience.
- We present a simulation study and some real data examples (marine data)



We start in the univariate setting [4, 2]

- Let Y be a real valued random variable on $\mathbb{R}$,(*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 $X = Y \bmod 2\pi$ and $0 \leq X < 2\pi$.
- 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:
$$f(x) = \sum_{k=-\infty}^{\infty} g(x + 2k\pi), \;\; 0 \leq x < 2\pi \tag{1}$$
- $g(\cdot)$ is the distribution of $Y = X + 2K\pi$, Y determines X and K through the modulus operation, and X is a wrapped version of Y

- The distribution for K is easily obtained:
  $P(K = k) = \int_0^{2\pi} g(x + 2k\pi)dx$.
- And $K|X = x$ is such that
  $P(K = k|X = x) = g(x + 2k\pi)/\sum_{j=-\infty}^{\infty} g(x + j2\pi)$
- And 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 we obtain a multivariate wrapped distribution for $\mathbf{X} = (X_1, X_2, ..., X_p)$ starting with a multivariate linear distribution for $\mathbf{Y} = (Y_1, Y_2, ..., Y_p) \sim g(\cdot)$ where $g(\cdot)$ is a $p-$variate distribution on $\mathbb{R}^p$, g is a family of distributions indexed by $\boldsymbol{\theta}$;
- Let $g(\cdot)$ be a $p-$variate normal distribution. $\mathbf{K} = (K_1, K_2, ..., K_p)$ is such that $\mathbf{Y} = \mathbf{X} + 2\pi\mathbf{K}$. Then, the joint distribution of $\mathbf{X}$ and $\mathbf{K}$ is $g(\mathbf{x} + 2\pi\mathbf{k})$ for $0 \leq x_j < 2\pi, j = 1, 2, ..., p$ and $k_j \in \mathbb{Z}, j = 1, 2, ..., p$. The marginal distribution of $\mathbf{X}$ is, directly

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

Again we introduce latent K_j's to facilitate the model fitting.

$\mathbf{X}$ has a p-variate wrapped normal distribution (WN) when $g(\cdot;\theta)$ is a multivariate normal and $\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

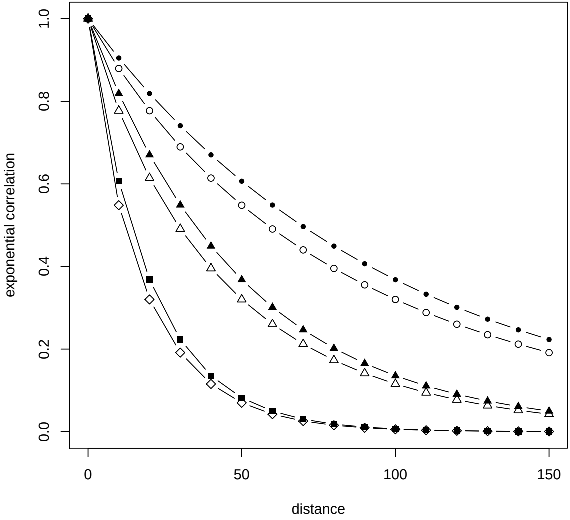
Using standard results, the conditional distribution of Y_j given $\{Y_l, l \neq j\}$ and $\boldsymbol{\theta}$ and the distribution of X_j, K_j given $\{X_l, K_l, l \neq j\}$ and $\boldsymbol{\theta}$ are immediate. In [3] it is shown how to truncate the series when $g(\cdot)$ is Gaussian based on distribution variability.

Introduce dependence in space and/or time it is easily achieved by equipping the linear variable with a structured covariance function.

- In general we can say that given a linear spatio-temporal Gaussian process with: mean $\boldsymbol{\mu}_Y$ and covariance $\boldsymbol{\Sigma}_Y = \sigma^2\mathbf{R}$ ($\mathbf{R}(\phi)$ space-time correlation function parametrized by ϕ),
- it induces a **Wrapped spatio-temporal Gaussian process**

$$\mathbf{X} \sim WN(\boldsymbol{\mu}_X, \sigma^2\mathbf{R}(\phi))$$

Spatial linear exponential correlation (solid) and its circular counterpart $\rho_c(s, s') = \sinh(\sigma^2 \rho(s, s'))/\sinh(\sigma^2)$ (empty)



Which covariance/correlation

In what follows we consider a very general and flexible covariance function for the linear variable:

$$\mathrm{Cor}(\mathbf{h}, u) = \frac{1}{(a|u|^{2\alpha}+1)^{\tau}} \exp\left(-\frac{c\|\mathbf{h}\|^{2\gamma}}{(a|u|^{2\alpha}+1)^{\beta\gamma}}\right), \ (\mathbf{h}; u) \in \mathbb{R}^{d} \times \mathbb{R} \quad (3)$$

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 [1]

Model

We write the linear GP $Y(\mathbf{s},t) = X(\mathbf{s},t) + 2\pi K(\mathbf{s},t)$ as:

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

- $\mu_Y(\mathbf{s},t)$: mean function,
- $\omega_Y(\mathbf{s},t)$: space-time GP with zero mean and covariance function $\sigma^2(\mathbf{s},t)\mathrm{Cor}(\mathbf{h},u)$, where $\mathrm{Cor}(\mathbf{h},u)$ is defined in (3)
- $\varepsilon_Y(\mathbf{s},t) \sim N(0,\phi_Y^2)$ is an independent random error (nugget effect or measurement error).

Model

- We consider several situations with different complexity degree in terms of mean and variance structure.
- The simplest one has constant mean and variance
- We adopt an ANOVA-type model for the mean and/or the variance when an auxiliary information allows us to imagine that there exist n_1 possible mean or variance levels
- We adopt a regression structure for the mean when auxiliary information is available (wave heights and directions). Hence here we have to consider an appropriate link function connecting the linear and the circular mean $\Longrightarrow$ 2atan[1]

[1] We use the atan definition of [2, page 13]

5 models

- *WN1*: $\mu_Y(\mathbf{s}, t) = \mu_Y$ and $\sigma^2_Y(\mathbf{s}, t) = \sigma^2_Y$;
- *WN2*: ANOVA parametrization for the mean and $\sigma^2_Y(\mathbf{s}, t) = \sigma^2_Y$;
- *WN3*: $\mu_Y(\mathbf{s}, t) = \mu_Y$ and ANOVA parametrization for the variance;
- *WN4*: atan link for the mean and $\sigma^2_Y(\mathbf{s}, t) = \sigma^2_Y$;
- *WN5*: atan link for the mean and ANOVA parametrization for the variance.

Priors

We suggest the following choices:

- covariance: $a \sim Gamma(a_a, b_a)$, $c \sim Gamma(a_c, b_c)$, $\alpha \sim Beta(\nu_{1,\alpha}, \nu_{2,\alpha})$, $\beta \sim Beta(\nu_{1\beta}, \nu_{2,\beta})$, $\gamma \sim Beta(\nu_{1\gamma}, \nu_{2,\gamma})$,
- process: $\sigma^2_Y \sim InGamma(a_\sigma, b_\sigma)$, $\sigma^2_{Y,i} \sim InGamma(a_{\sigma_i}, b_{\sigma_i})$, $\phi^2_Y \sim InGamma(a_\varepsilon, b_\varepsilon)$,
- the mean: Wrapped Gaussian

Our benchmark: the spatio-temporal Projected Gaussian process, defined by [7].

- Let $\mathbf{Z} = (Z_1, Z_2)$ be bivariate Gaussian with mean $\boldsymbol{\mu}_Z = (\mu_{Z,1}, \mu_{Z,2})$ and covariance matrix
$$\begin{pmatrix} \sigma^2_{Z,1} & \sigma_{Z,1}\sigma_{Z,2}\rho_z \\ \sigma_{Z,1}\sigma_{Z,2}\rho_z & \sigma^2_{Z,2} \end{pmatrix}$$
- We transform $\mathbf{Z}$ into an angular variable, Θ, with the transformation $\Theta = \text{atan}\frac{Z_2}{Z_1}$
- Θ is distributed as a Projected Normal variable [4, pag. 52] with parameters $\mu_{Z,1}, \mu_{Z,2}, \sigma^2_{Z,1}, \sigma^2_{Z,2}, \rho_z$.
- [6] note that the distribution of Θ is invariant if we multiply $\mathbf{Z}$ by a positive constant
- an identification constraint is required and the authors suggestion is: $\sigma^2_{Z,2} = 1$
$$\mathbf{V} = \begin{pmatrix} \sigma^2_{Z,1} & \sigma_{Z,1}\rho_z \\ \sigma_{Z,2}\rho_z & 1 \end{pmatrix}$$

- Working with the Projected Normal distribution is analytically not "easy"
- It is convenient to introduce a latent variable $R = ||\mathbf{Z}||$ and work with the joint density of (Θ, R)

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

- We can move back and forth between the linear variables and (Θ, R) using $Z_1 = R\cos\Theta$, $Z_2 = R\sin\Theta$ and the atan transformation.

- Let $\mathbf{Z}(\mathbf{s}, t) = (Z_1(\mathbf{s}, t), Z_2(\mathbf{s}, t))$ be a 2-dimensional space-time process with constant mean $\boldsymbol{\mu}_Z$ and cross covariance function $C_\theta\left((\mathbf{s}, t), (\mathbf{s}', t')\right)$
- 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 Σ_Z (see [5] for details).
- As before the latent variable $R(\mathbf{s}, t)$ is introduced to facilitate model fitting.

Model details

- We define, for each s, t the 2-dimensional linear process:

$$
\begin{aligned}
Z_1(\mathbf{s}, t) &= \mu_{Z,1} + \omega_{Z,1}(\mathbf{s}, t) + \tilde{\varepsilon}_{Z,1}(\mathbf{s}, t) \\
Z_2(\mathbf{s}, t) &= \mu_{Z,2} + \omega_{Z,2}(\mathbf{s}, t) + \tilde{\varepsilon}_{Z,2}(\mathbf{s}, t)
\end{aligned}
$$

- where $\boldsymbol{\mu}_Z = (\mu_{Z,1}, \mu_{Z,2})$ is the mean, $\boldsymbol{\omega}_Z(\mathbf{s}, t) = (\omega_{Z,1}(\mathbf{s}, t), \omega_{Z,2}(\mathbf{s}, t))$ is a bivariate Gaussian process with zero mean and covariance function $\mathbf{V} \otimes Cor(\mathbf{h}, v)$
- $Cor(\mathbf{h}, v)$ is the Gneiting correlation introduced before
- $\tilde{\varepsilon}_Z(\mathbf{s}, t)$ is a bivariate error with zero mean and covariance matrix $\phi_Z^2 \mathbf{I}$.

Model details

- We marginalized over $\boldsymbol{\omega}_Z(\mathbf{s}, t)$ above:

$$Z_1(\mathbf{s}, t) = \mu_{Z,1} + \varepsilon_{Z,1}(\mathbf{s}, t)$$
$$Z_2(\mathbf{s}, t) = \mu_{Z,2} + \varepsilon_{Z,2}(\mathbf{s}, t)$$

- Now $\varepsilon_Z(\mathbf{s}, t)$ is a bivariate Gaussian process with zero mean and covariance function $\mathbf{V} \otimes \mathrm{Cor}(\mathbf{h}, v) + \mathbf{I}_4 \phi_Z^2$
- Then $\Theta = \mathrm{atan}\frac{Z_1(s,t)}{Z_2(s,t)}$ is a circular process with constant mean $\boldsymbol{\mu}_Z$, a nugget (measurement error) and, as in the WN setting, correlation between the circular variable induced by the Gneiting spatio-temporal correlation function. (PN1)
- A model without nugget is readily obtained by removing $\tilde{\varepsilon_Z}$ in the previous expression (PN2)
- **Priors**: Gaussian with large variance for $\mu_{Z,i}, i = 1, 2$, $\rho_Z \sim N(\mu_\rho, \sigma_\rho) I(-1, 1)$, $\phi_Z^2 \sim \mathit{InvGamma}$, for the correlation parameters the same priors as in the WN

Wrapped Normal

- To perform prediction with a space-time wrapped GP we need 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 asses 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.
- Then we can compute Monte Carlo approximations of the desired mean.

Projected Normal

- Predictive distribution available
- Given $(\mathbf{s}_0, t_0)$ we can infer about the circular mean and the concentration at the unobserved spatio-temporal location using posterior samples of the parameters and the latent $\mathbf{R}$.
- At each iteration of the McMC we draw a sample $\mathbf{z}(\mathbf{s}_0, t_0)^b$ from the distribution of $\mathbf{Z}(\mathbf{s}_0, t_0)|\boldsymbol{\Theta}, \mathbf{R}^b, \boldsymbol{\Psi}_Z^b$ and then we convert it to the associated circular process.
- Recall that the circular mean is $atan\left(\frac{\mathrm{E}\cos\Theta}{\mathrm{E}\sin\Theta}\right)$ and the concentration is $\sqrt{\mathrm{E}\cos^2\Theta + \mathrm{E}\sin^2\Theta}$ then we can compute them using their McMC approximation

Implementation

- We need several computational “tricks” to speed up convergence and ensure identifiably
- block sample nugget and variance
- Adaptive metropolis
- Large matrices to be inverted $\rightarrow O(n^3)$ operations

Simulation scheme

Parameters in common to all dataset:

- 20 locations and 12 time points.
- coordinates uniformly generated in $[0, 10] \times [0, 10]$.
- 170 points between the 1^{th} and 10^{th} time are used for model estimation, the remaining 70 for model validation.
- Correlation parameters:
  - $(a, c) = \{(0.2, 1); (1, 0.2)\}$.
  - $\beta = \{0; 0.5; 0.9\}$.
  - $\alpha = \{0.5; 0.8\}$.
  - $\gamma = \{0.5; 0.8\}$.

Simulation scheme

Wrapped Models parameters:

- $\phi_Y^2 = (0.01, 0.1)$ nugget.
- constant σ_Y^2: $(\sigma_Y^2, \phi_Y^2) = (= 0.1, 0.01)$ and $(\sigma_Y^2, \phi_Y^2) = (0.5, 0.1)$ (WN1,WN2,WN4).
- constant mean: $\mu_Y = \pi$ (WN1,WN3).
- ANOVA - type mean: 3 possible values for each (s, t) with probability $1/3$, $(1, \pi, 5)$ (WN2).
- ANOVA-type $\sigma_Y^2(\mathbf{s}, t)$: 3 values for the variance: $(0.1, 0.5, 1)$ again with probability $1/3$(WN3,WN5).
- Regression function for the mean: $\pi + 0.5 * U(s, t)$ with $U(s, t) \sim Unif(-10, 10)$

Simulation scheme

Projected models parameters:

- We simulate with two sets of parameters for the PN1:
  - $(\mu_{Z,1}, \mu_{Z,2}, \sigma^2_{Z,1}, \rho_Z, \phi^2_Z) = (2, 2.5, 1, 0.2, 0.01)$.
  - $(\mu_{Z,1}, \mu_{Z,2}, \sigma^2_{Z,1}, \rho_Z, \phi^2_Z) = (1.2, 1.2, 1, 0.2, 0.1)$.
- In the model PN2 we used the same parameters of the model PN1 excluding ϕ^2_Z

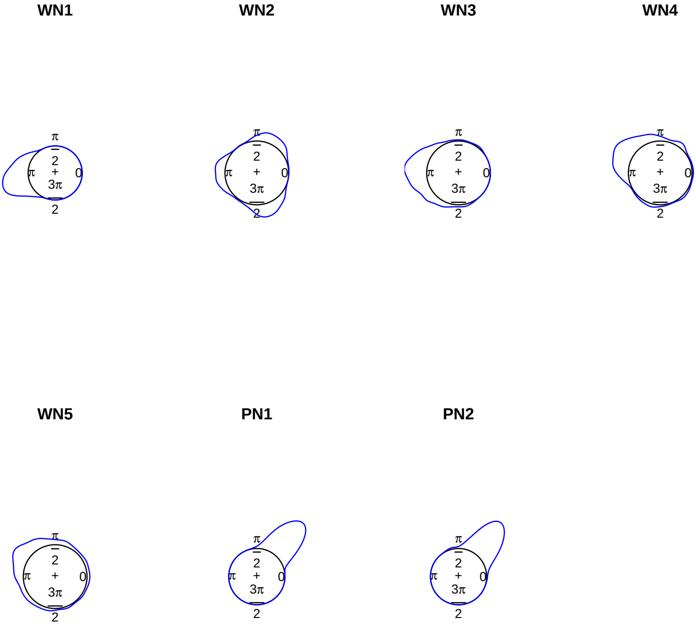
Note that:

- We simulate unimodal and slightly asymmetric projected Normal distribution
- We consider parameters combinations that induce circular variances of the same order of magnitude as in the WN simulations, i.e. "small" and "large" variance

- To asses model performance we compute an **average prediction error** (APE), defined as the average circular distance[2] between a validation dataset and model predicted values.
- All runs are implemented on a large computers cluster the Bari INFN high performance Grid computing infrastructure Bc2S [3], about 250 computing knots (4000 CPU cores) and it allows data management and storage on a 1650 TB shared hard disk.

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

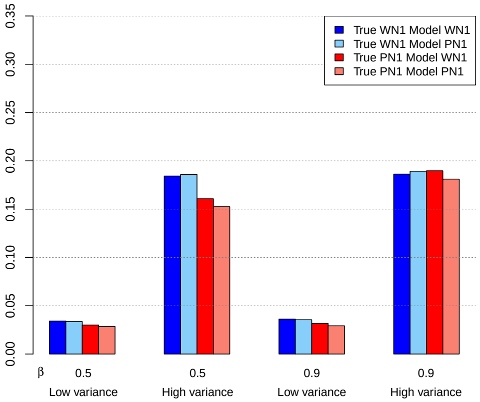
[3] Bari Computer Center for Science


WN1
WN2
WN3
WN4
WN5
PN1
PN2
$\frac{\pi}{2}$
π
0
$+$
$\frac{3\pi}{2}$

Computational efficiency

Model	Average Iterations	Average Time
PN 1	700000	28h
PN 2	700000	10h
WN 1:3	150000	2h
WN 4:5	300000	4h

WN vs PN

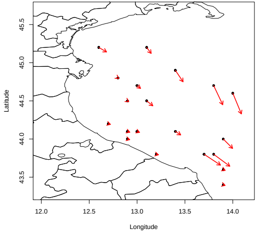


0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
True WN1 Model WN1
True WN1 Model PN1
True PN1 Model WN1
True PN1 Model PN1
β
0.5
Low variance
0.5
High variance
0.9
Low variance
0.9
High variance

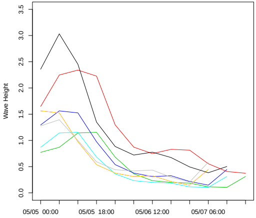
- Outputs from a deterministic model implemented by ISPRA[4]. The model starts from a wind forecast model predicting the surface wind over the entire Mediterranean. The hourly evolution of sea wave spectra is obtained solving energy transport equations using as input the wind forecast. Wave spectra are 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.
- It produces several waves parameters, here we consider significant wave height and direction.
- It is affected by a large uncertainty, spatial resolution is 0.1 degree longitude about $12.5 \times 12.5km$ cells, time resolution is 1hour.

[4] Istituto Superiore per la Protezione e la Ricerca Ambientale

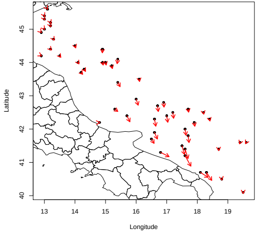
We choose two datasets:



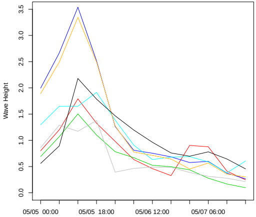
(Dataset1.D)



(Dataset1.H selected locations)



(Dataset2.D)



(Dataset2.H selected locations)

- Both datasets cover the period between the 00:00 of May 5, 2010 and the 18:00 of May 7, 2010.
- We select values every 6 hours.
- The association between wave height and direction is used as auxiliary information for both the regression and the ANOVA-type models. In the latter we define 3 groups wave height $\leq 1m$, $(1, 2]m$, $> 2m$.

Dataset1:

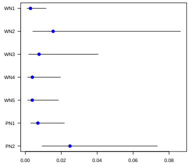
- North-West area of the Adriatic sea.
- 20 spatial points.
- 170 points between the 5^{th} 00:00 and the 7^{th} 06:0 are used for model estimates, the remaining 70 for model validation.

Dataset2:

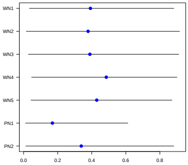
- The entire Adriatic sea.
- 50 spatial points.
- 425 points between the 00:00 May 5, 2010 and the 06:00 of May 7,2010 are used for model estimation, the remaining 175 points for model validation.

Parameters estimates

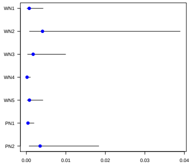
Nuggets estimates are the same up to the 3rd figure in WN1 and PN1



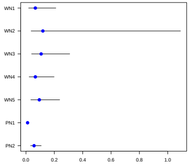
(Dataset1: $\hat{a}$)



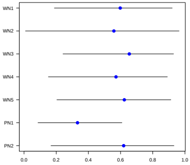
(Dataset1: $\hat{\beta}$)



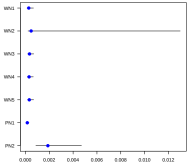
(Dataset1: $\hat{c}$)



(Dataset2: $\hat{a}$)

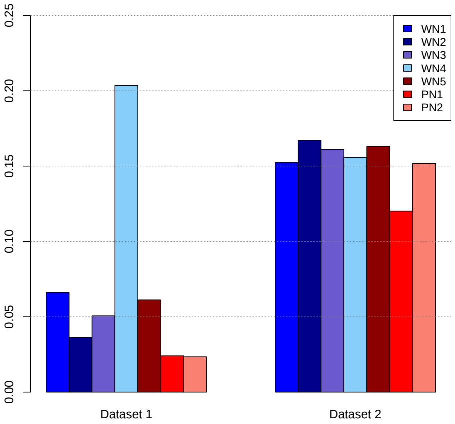


(Dataset2: $\hat{\beta}$)

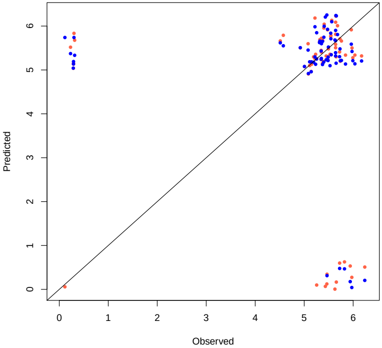


(Dataset2: $\hat{c}$)

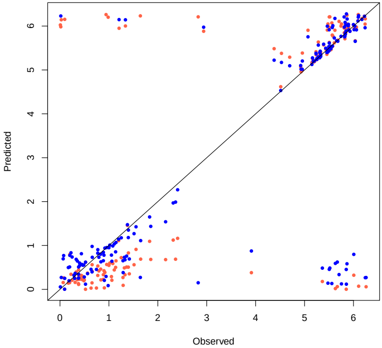
APE


0.00
0.05
0.10
0.15
0.20
0.25
WN1
WN2
WN3
WN4
WN5
PN1
PN2
Dataset 1
Dataset 2

Given the parameters estimates and the APE we'd choose the PN1 for both datasets model but both datasets can be reasonably handled using WN1



Dataset 1



Dataset 2

red WN , blue PN

Pros and Cons

- Efficiency, both computational and statistical, depends on the process variance for both models
- The Wrapped Normal is a computationally very convenient model when reasonably symmetric and unimodal data are available
- The WN can reasonably approximate the PN
- In the WN model no full Bayesian inference for the space-time process.
- The Projected normal model can handle multimodal situations and $d > 2$ and full Bayeian inference is possible.
- The PN model parameters are not easily interpreted.

Coming soon:

- More simulations to compare WN and PN when data are multimodal
- HMM using projected normals
- An R package implementing the proposed models
- Wrapped point processes (?)

[1] Gneiting, T. (2002). Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, **97**(458), 590–600.

[2] Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific Pub Co Inc, har/dskt edition.

[3] 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).

[4] Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Pr.

[5] Wang, F. (2013). *Space and Space-Time Modeling of Directional Data*. Ph.D. thesis, Duke University.

[6] Wang, F. and Gelfand, A. E. (2012). Modeling space and space-time directional data using projected gaussian processes. Technical report, Duke university.

[7] Wang, F., Gelfand, A. E., and Jona-Lasinio, G. (2014). Joint spatio-temporal analysis of a linear and a directional variable: space-time modeling of wave heights and wave directions in the adriatic sea. *Statistica Sinica*.

Thanks for your attention