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

Alan E. Gelfand
Duke University

(Joint work with Fangpo Wang, Giovanna Jona Lasinio, Gianluca Mastroantonio, Maria Terres, Shin Shirota)

WOW! Some remarks

- Families of models with skewness and kurtosis

WOW! Some remarks

- Families of models with skewness and kurtosis
- Hypothesis testing and power calculations

WOW! Some remarks

- Families of models with skewness and kurtosis
- Hypothesis testing and power calculations
- Goodness of fit tests

WOW! Some remarks

- Families of models with skewness and kurtosis
- Hypothesis testing and power calculations
- Goodness of fit tests
- Asymptopia

WOW! Some remarks

- Families of models with skewness and kurtosis
- Hypothesis testing and power calculations
- Goodness of fit tests
- Asymptopia
- Need more females

WOW! Some remarks

- Families of models with skewness and kurtosis
- Hypothesis testing and power calculations
- Goodness of fit tests
- Asymptopia
- Need more females
- I live on a different planet

Outline of the Talk

- Some basics of directional data
- Briefly, wrapped normal distributions and wrapped Gaussian processes
- Projected normal distributions and projected Gaussian processes in space and space-time
- Examples
- Spatial analysis of a directional and linear variable
- New work: Angular discrepancy processes; aspect processes; spatial point patterns on a circle

Directional data

- Directional, circular, angular data (here, in 2 dimensions)
- Applications include:
  - meteorology (wind direction)
  - oceanography (wave direction, different from wind direction)
  - ecology (animal movement)
  - periodic data, say daily or weekly, “wrap” it to be circular (time of max ozone level, time and day of a particular type of crime), convert to $[0, 2\pi)$
- Some of these applications can be spatial - wind, wave directions
- Can have a linear variable as well - ozone level, wave height
- Can be dynamic

Here, exclusively a Bayesian view

- Inference on population features, inference on prediction in space and time (kriging)
- Model fitting through MCMC; posterior sampling (analogue of resampling techniques for likelihood-based inference (Pewsey et al., 2013))
- Fitting is easy through introduction of latent variables
- Limited Bayesian literature
  - Damien and Walker (1998) Von Mises distribution
  - Ghosh et al. (2003) One and two sample problems
  - Nuñez-Antonio and Gutierrez-Peña (2005a,b, 2011) Von Mises, wrapped distribution, projected normal
  - Ravindran (2002) Thesis on wrapped distributions
  - Coles (1998) Multivariate circular data; Casson and Coles (1998) von Mises with spatial dependence in location and scale parameters

Intrinsic Approach

- von Mises distribution $M(\mu, \kappa)$, density

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)},$$

  where μ is mean direction, κ is concentration, and I_0 modified Bessel function of the first kind of order 0.
- Most common circular distribution. Circular analogue of normal distribution for linear data
- symmetric, unimodal; mixture models to add flexibility
- Infeasible for multivariate angular data
- For spatial or temporal or space-time data, conditionally independent von Mises with process models at second stage for μ, κ. May be unattractive.

Wrapping

- Wrap a linear variable, i.e., $\theta = Y \text{mod} 2\pi$
- If $g(y)$ is a density on R^1, wrapped density looks like

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2\pi k)$$

- Regression - again, a link function to the circle
- Obviously, can rescale from $[0, L)$ to $[0, 2\pi)$
- Multivariate version (say p-dim) is easy. With multivariate density g on R^p,

$$f(\boldsymbol{\theta}) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_p=-\infty}^{\infty} g(\boldsymbol{\theta} + 2\pi \mathbf{k})$$

- Convenient choice is a multivariate normal

The univariate wrapped normal

- $WN(\mu, \sigma^2)$ density takes the form, for $0 \le \theta < 2\pi$:

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta+2k\pi) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} \text{exp}(-\frac{(\theta+2k\pi-\mu)^2}{2\sigma^2})$$

- $E(Z) = e^{-\sigma^2/2}e^{i\mu}$ so μ is the linear mean with $\mu = \tilde{\mu} + 2\pi K_\mu$ with $\tilde{\mu} \in [0, 2\pi)$ the mean direction and $c = e^{-\sigma^2/2}$ is concentration
- θ is observed; $\theta + 2K\pi$ is the linear variable; K is latent
- Joint density for θ and K is $f(\theta, k) = g(\theta + 2k\pi) =$

$$\frac{1}{\sigma\sqrt{2\pi}}\text{exp}(-\frac{(\theta+2k\pi-\mu)^2}{2\sigma^2}), \quad 0 \le \theta < 2\pi, K \in \{0, \pm 1, \pm 2, \ldots$$

- Removes the doubly infinite sum; suggests adding K as a latent variable

Wrapped Gaussian Processes

- Recall the multivariate wrapped distribution:

$$f(\boldsymbol{\theta}) = \sum_{k_1=-\infty}^{\infty} \dots \sum_{k_p=-\infty}^{\infty} g(\boldsymbol{\theta} + 2\pi\mathbf{k})$$

- Here $\boldsymbol{\theta}$ is observed vector, $\boldsymbol{\theta} + 2\pi\mathbf{K}$ is the linear vector, $\mathbf{K}$ is the latent vector
- Again, the joint density for $(\boldsymbol{\theta}, \mathbf{K})$ is:

$$f(\boldsymbol{\theta}, \mathbf{K}) = g(\boldsymbol{\theta} + 2\pi\mathbf{k})$$

- Since GP's are specifies through their finite dimensional distributions, we can induce a wrapped GP from a linear GP. In particular, if linear GP has covariance function $\sigma^2\rho(s - s'; \phi)$, then

$$\boldsymbol{\theta} = (\theta(s_1), \theta(s_2), \dots, \theta(s_n)) \sim WN(\mu\mathbf{1}, \sigma^2 R(\phi))$$

Projection Approach

- An embedding approach - unit circle within R^2
- $\mathbf{U} = (U_1, U_2) \sim g(u_1, u_2)$, a density on R^2
- Then $(V_1, V_2) = (\frac{U_1}{||\mathbf{U}||}, \frac{U_2}{||\mathbf{U}||})$ where $||\mathbf{U}||$ is the length of $\mathbf{U}$, is a point on the unit circle, associated angle is $\theta =$ atan2$\frac{V_2}{V_1} =$ atan2$\frac{U_2}{U_1}$
- In fact, $U_1 = R$cosθ and $U_2 = R$sinθ, R latent
- $R = ||\mathbf{U}||$, $V_1 =$ cosθ, $V_2 =$ sinθ
- Again, angular mean direction is atan2$\frac{E\text{sin}\theta}{E\text{cos}\theta} = \frac{E(V_2)}{E(V_1)} \neq \frac{E(U_2)}{E(U_1)}$
- Concentration is $||E(\mathbf{V}|| \leq 1$

Projection cont.

- Projected normal distribution. Suppose the random vector $\mathbf{U} \sim N_2(\boldsymbol{\mu}, \Sigma)$, then $\theta \sim PN_2(\boldsymbol{\mu}, \Sigma)$.
- More flexible - can be asymmetric, bimodal
- Easy for regression - linear model in covariates for μ - but may be hard to interpret, a regression for each component
- A nice characterization: The collection of mixtures of projected normals is dense in the class of all circular distributions
- Difficult to work with for dim >2.

Projected normal

- Back to projected normal
- Recall, if $\mathbf{U} = (U_1, U_2) \sim g(u_1, u_2)$, a density on R^2
- Then $(\frac{U_1}{||\mathbf{U}||}, \frac{U_2}{||\mathbf{U}||})$ where $||\mathbf{U}||$ is the length of $\mathbf{U}$, is a point on the unit circle, associated angle is $\theta = \text{atan2}\frac{U_2}{U_1}$
- Projected normal distribution. Suppose the random vector $\mathbf{U} \sim N_2(\boldsymbol{\mu}, \Sigma)$, then $\theta \sim PN_2(\boldsymbol{\mu}, \Sigma)$.
- The density can be obtained explicitly but is very messy.
- Instead, we would use polar coordinates working with the joint density of (θ, R) derived as a transformation from (U_1, U_2), treating R as a latent variable
- $f(r, \theta|\boldsymbol{\mu}, \Sigma) = (2\pi)^{-1}|\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{(r\mathbf{u}-\boldsymbol{\mu})'\Sigma^{-1}(r\mathbf{u}-\boldsymbol{\mu})}{2}\right) r$

PN density $p = 2$

By transforming the bivariate random vector $\mathbf{x} = (x_1, x_2)'$ into polar co-coordinates (r, θ) and integrating over r for a given θ, the density function for θ , $f(\theta; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$ is obtained as,

$$f(\theta) = \frac{\phi(\mu_1, \mu_2; \mathbf{0}, \mathbf{\Sigma}) + aD(\theta)\Phi\{D(\theta)\}\phi\left[a\{C(\theta)\}^{-\frac{1}{2}}(\mu_1 \sin\theta - \mu_2 \cos\theta)\right]}{C(\theta)}$$

where

$$\begin{aligned} a &= \{\sigma_1\sigma_2\sqrt{1-\rho^2}\}^{-1} \\ C(\theta) &= a^2(\sigma_2^2\cos^2\theta - \rho\sigma_1\sigma_2\sin 2\theta + \sigma_1^2\sin^2\theta) \\ D(\theta) &= a^2\{C(\theta)\}^{-\frac{1}{2}}\{\mu_1\sigma_2(\sigma_2\cos\theta - \rho\sigma_1\sin\theta) + \mu_2\sigma_1(\sigma_1\sin\theta - \rho\sigma_2\cos\theta)\} \end{aligned}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

cont.

- What do projected normal densities look like?
- The form with general Σ has only been considered theoretically; data analysis and inference has only been considered so far for the case $\Sigma = I$.
- In this latter case, the PN densities are symmetric, unimodal (and the uniform arises when $\mu_1 = \mu_2 = 0$).
- When $\Sigma = I$, the mean direction $\mu =$ atan2$\frac{\mu_1}{\mu_2}$, closed form for ρ (Kendall, 1974).
- In this case, the PN can be compared with the von Mises. Both have two parameters and can line up their directions and resultants.

cont.

- We work with the more general Σ case
- Can draw pictures of the density in terms of five parameters in $\boldsymbol{\mu}$ and Σ. We can achieve asymmetry and bimodality
- With regard to inference, an identifiability issue: Note that if we scale $\mathbf{U}$ by a, the distribution of θ doesn't change
- We simply set $\Sigma = \begin{pmatrix} \tau^2 & \rho\tau \\ \rho\tau & 1 \end{pmatrix}$
- We have a four parameter model
- No simple form for μ or c now; ugly functions of the four parameters but we can compute them numerically
- So, no role for the usual EDA ideas here

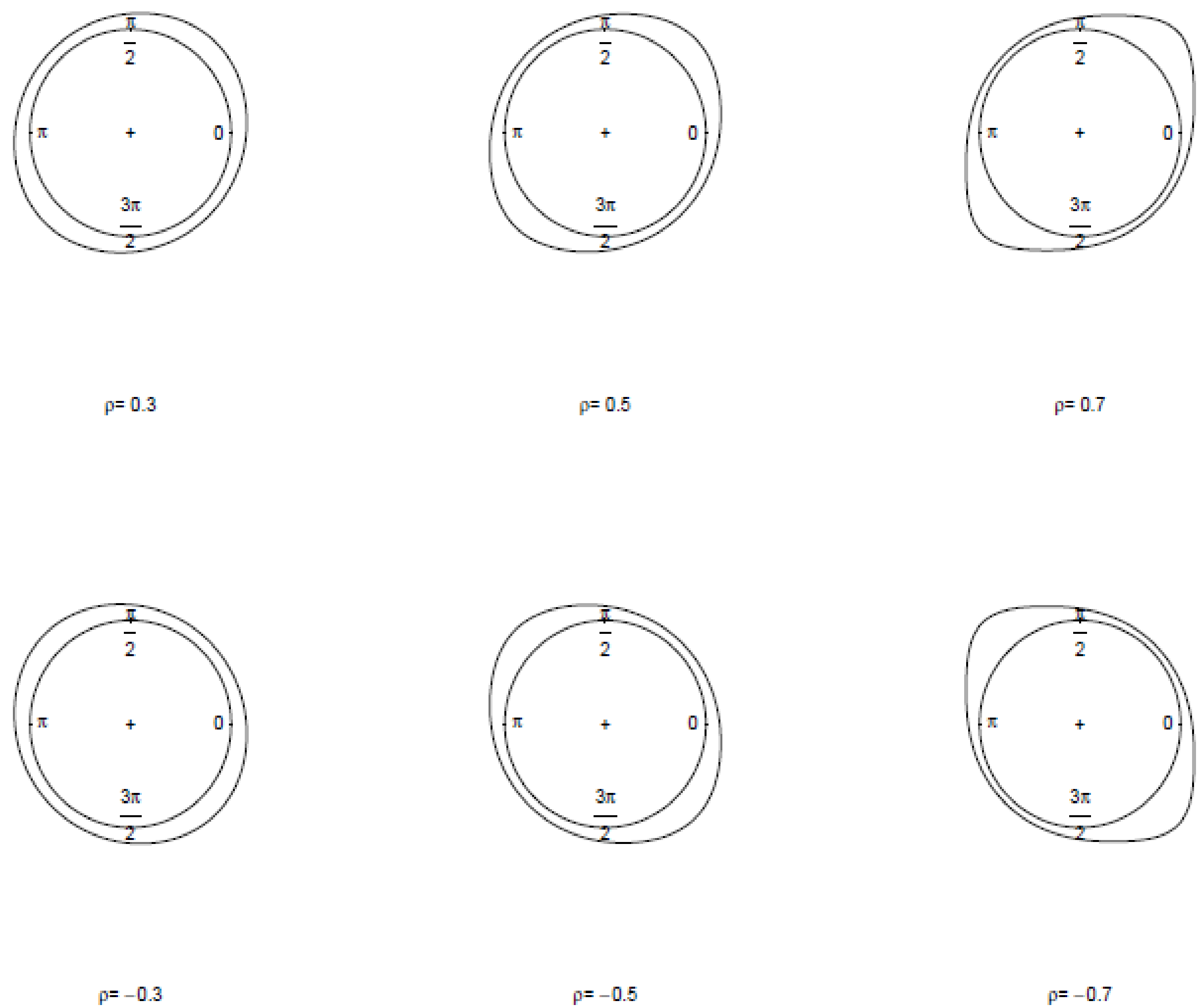
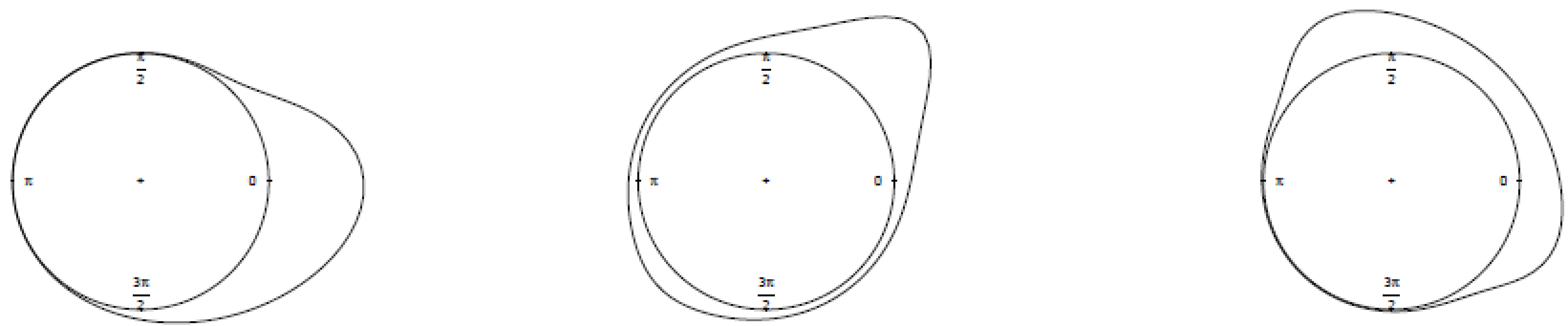


Figure 2. Density of θ for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = \sigma$ and different values of ρ



(a) Asymmetry (b) Antipodality (c) Bimodality

Figure 3. Shape of the general projected normal distribution

Model fitting and inference

- Bayesian model fitting is straightforward. With observed θ_i's and latent R_i's, we convert to U_{1i}'s and U_{2i}'s. Update β's and τ^2 and ρ under a standard bivariate Gaussian setup.
- The R_i's have an explicit closed form full conditional (M-H step with Gamma proposal)
- When $\Sigma \neq I$, we achieve better out-of-sample prediction using the correct model rather than the incorrect model
- Comparing an observed hold-out θ with an estimate of its mean direction is not sensible with bimodal densities
- With holdout, we use a predictive log likelihood loss (PLSL) and the cumulative rank probability score (CRPS; Grimit et al., 2006)

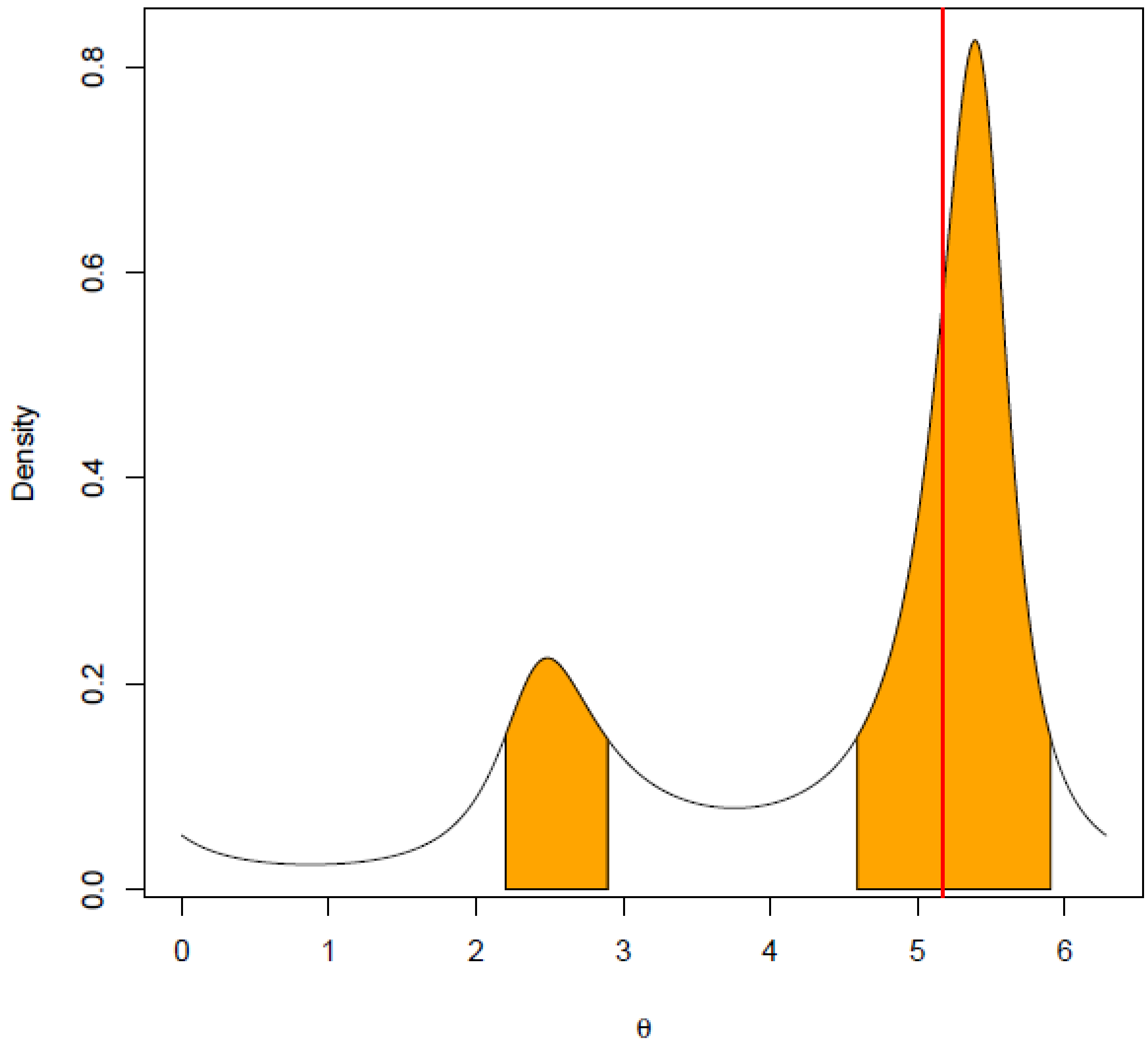


Figure 4. disjoint HPD sets for bimodal predictive distribution

Density estimation, n=50

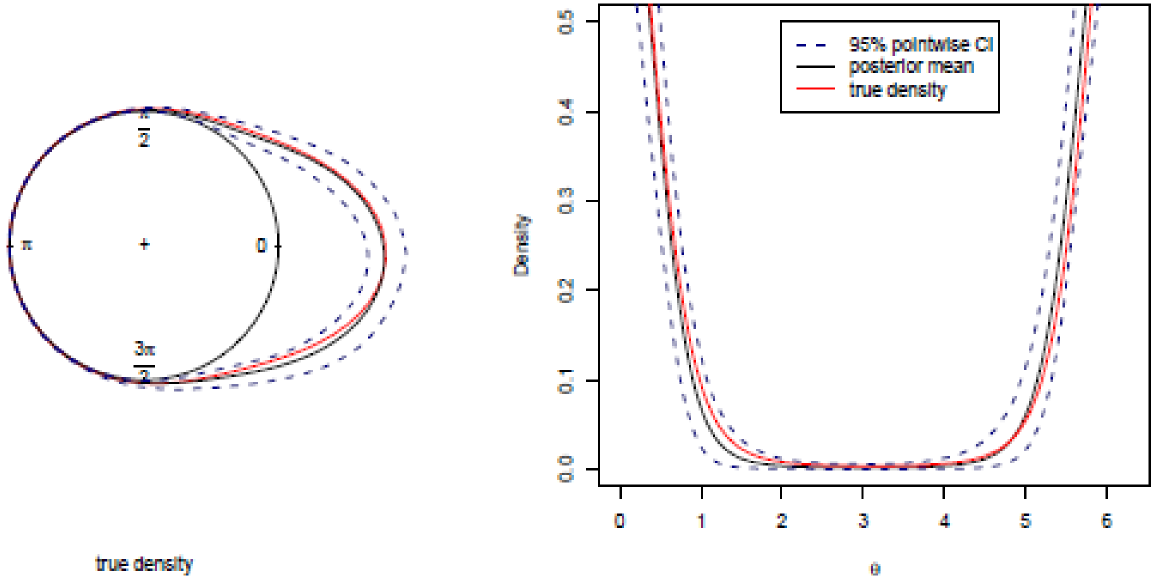


Figure: Location: $\mu_1 = 2$, $\mu_2 = -0.1$, $\tau^2 = 1$ and $\rho = 0$

Density estimation, n=50

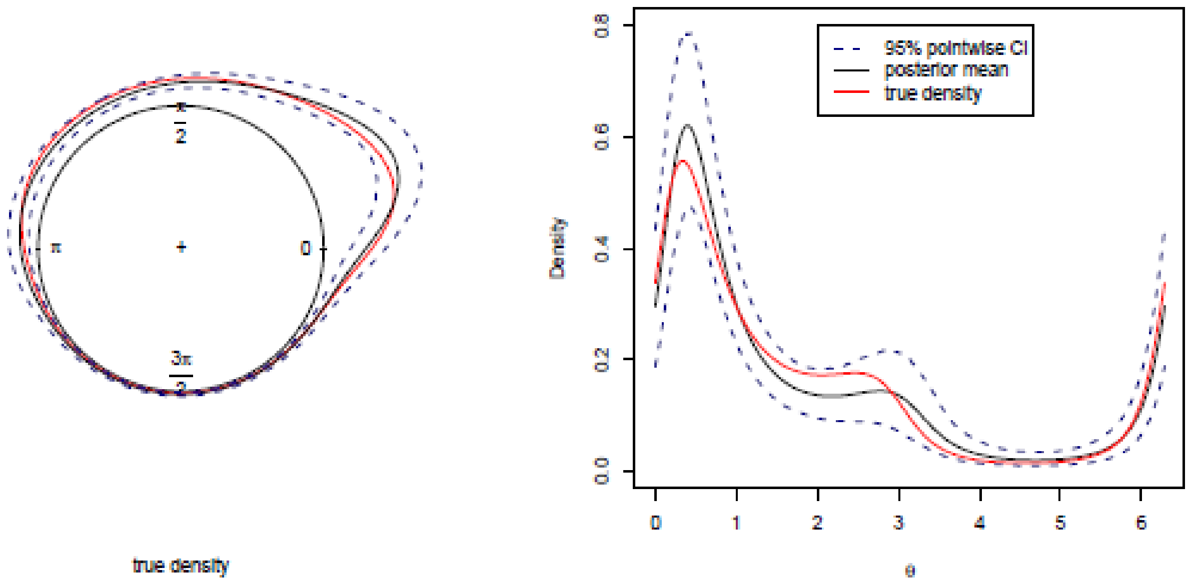
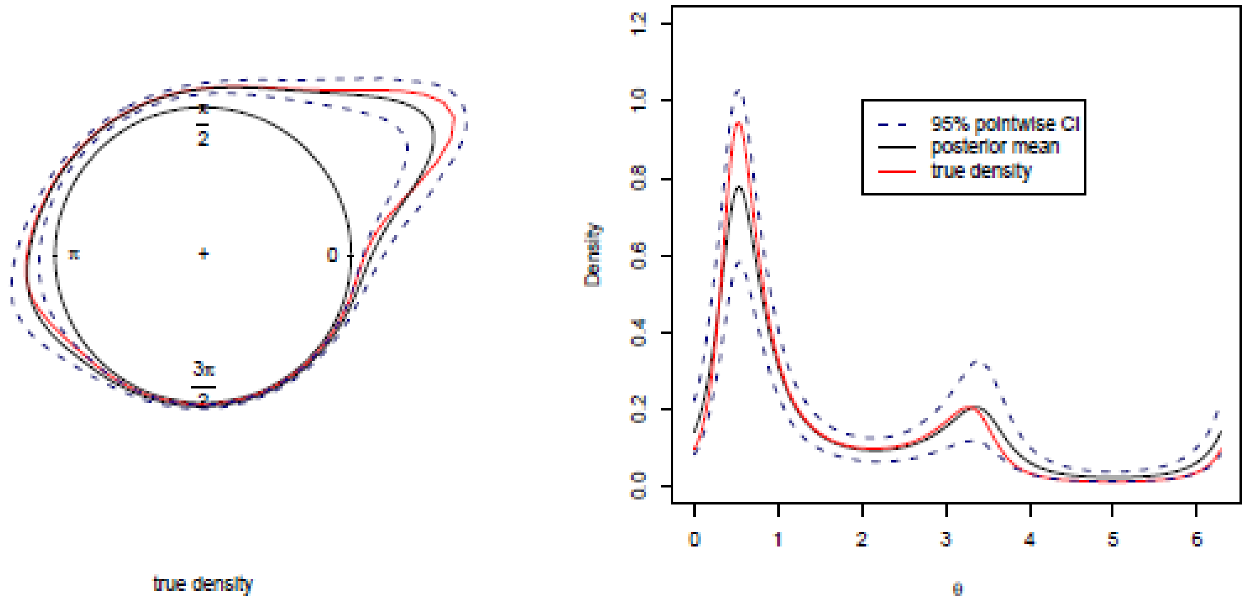


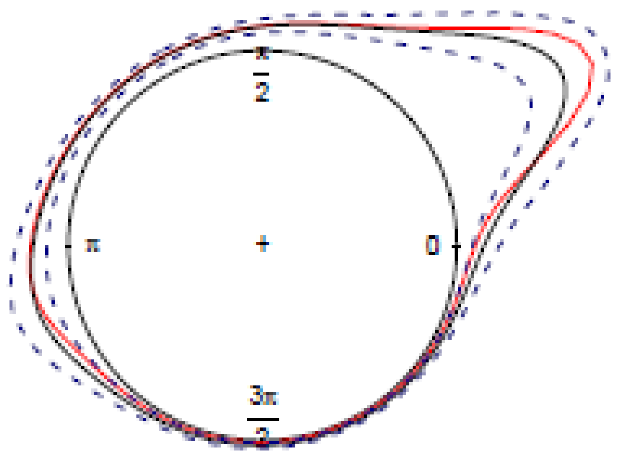
Figure: $\mu_1 = 1$, $\mu_2 = 1$, $\tau^2 = 4$ and $\rho = 0$

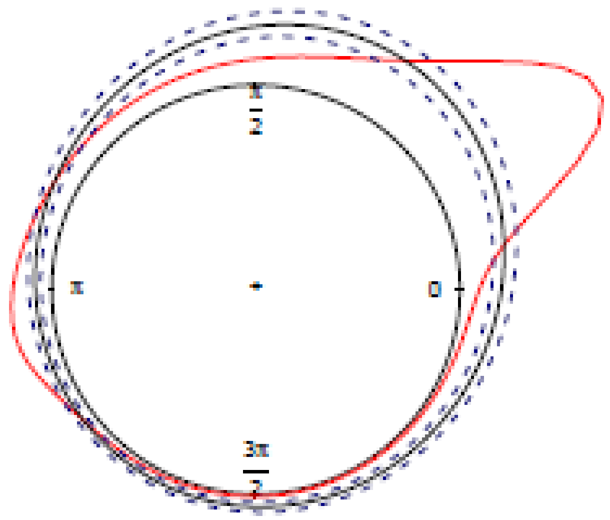
Density estimation, n=50



Correlation parameter: $\mu_1 = 1$, $\mu_2 = 1$, $\tau^2 = 4$ and $\rho = 0.7$

Comparing Σ vs. I





Example: Real angles

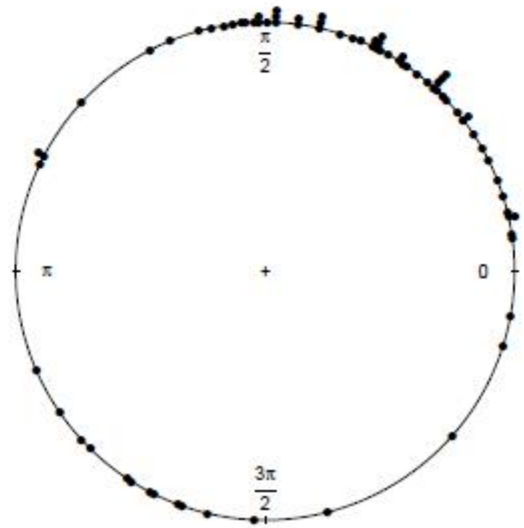
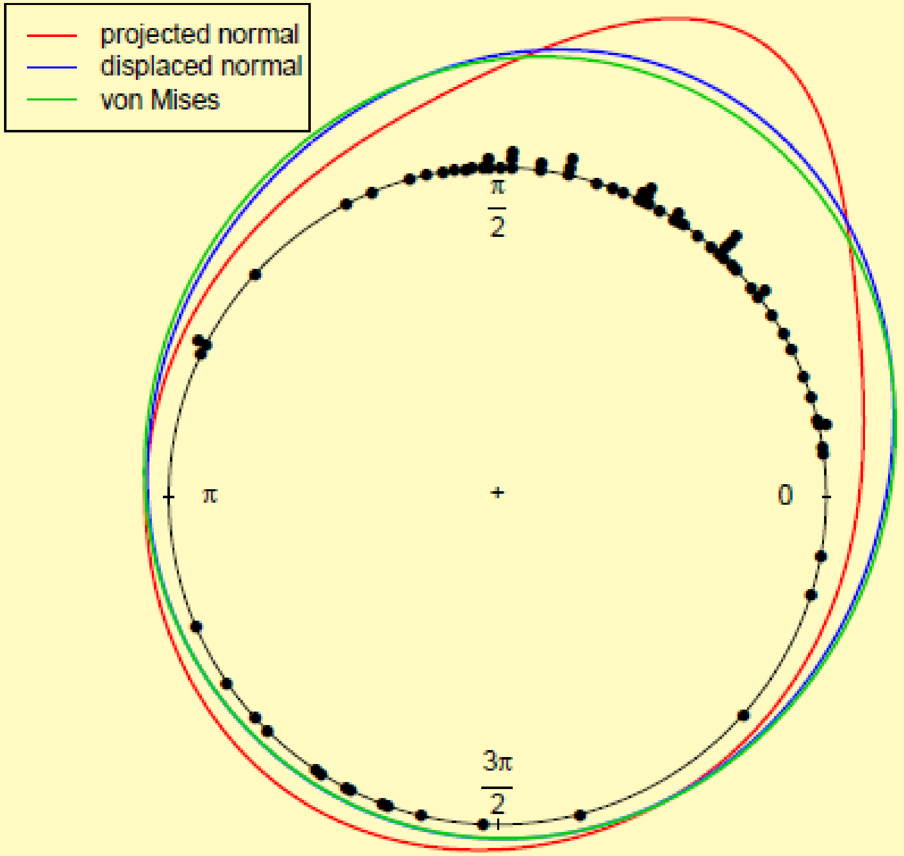


Figure: Orientations of 76 turtles after laying eggs (Gould's data cited by Stephens, 1969)

Real Data Example

The raw data are the directions in which 76 female turtles moved after laying their eggs on a beach. (Gould's data)



Spatial PN models

- Finally, we return to the case of $\{\theta(s_i), i = 1, 2, ..., n\}$
- In the independence case, we had latent independent U_i's modeled as bivariate normal
- Now, we assume latent $\mathrm{U}(s_i)$ from a bivariate Gaussian process
- This induces a spatial process for the $\theta(s_i)$ which we call the Projected Normal GP
- Many ways to specify the bivariate GP; separable cross-covariance function
- Kriging is, again, straightforward. We can krige posterior predictive samples of say $\mathrm{U}(s_0)$ which, in turn induce posterior predictive samples of $\theta(s_i)$
- We can easily insert spatial regressors, $\mathbf{X}(s)$ in the $\boldsymbol{\mu}(s)$, analogous to the independence case.

Model fitting

- From the joint distribution of $\{\mathbf{U}(s_i)\}$ can write the joint distribution of $\{\theta(s_i), r(s_i)\}$.
- So, now need to update $r(s_i)|$"everything else". But same idea as before; now the conditional distribution of $\mathbf{U}(s_i)|$"everything else" is a conditional normal so again, an explicit form for the full conditional for $r(s_i)$.
- Start with separable cross covariance functions for $\mathbf{U}(s)$
- From the separable cross covariance function, we can explore the induced covariance function for $\theta(s)$
- ρ is stationary in $\mathbf{U}(s)$ process, joint dist for $(\theta(s), \theta(s'))$ will depend on $s - s'$ but no implied form for correlation.
- General form proposed in Jammalamadaka and Sarma. Not likely to be a *valid* correlation function

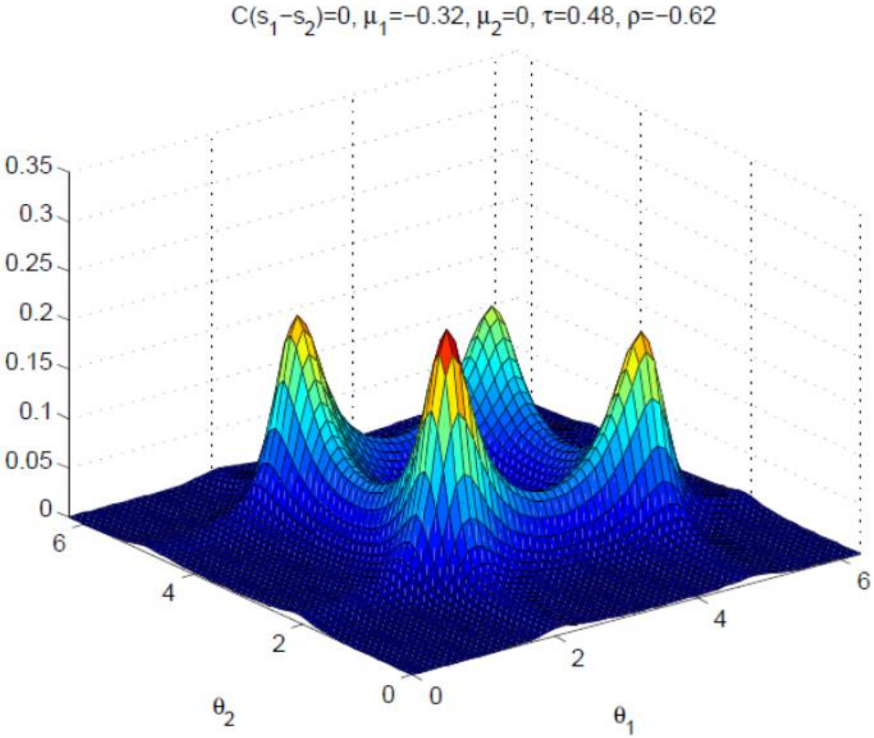
Circular correlation

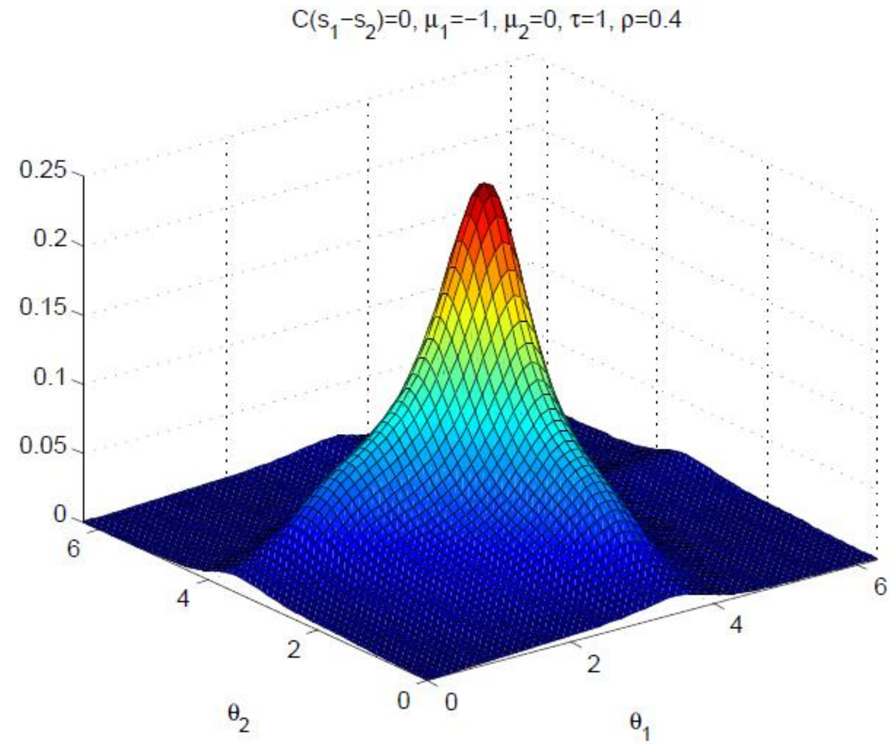
- Jammalamadaka and Sarma (1988) proposed a measure of circular correlation as

$$\rho_c(\alpha, \beta) = \frac{E\{\sin(\alpha - \mu)\sin(\beta - \nu)\}}{\sqrt{\text{Var}(\sin(\alpha - \mu))\text{Var}(\sin(\beta - \nu))}}. \tag{1}$$

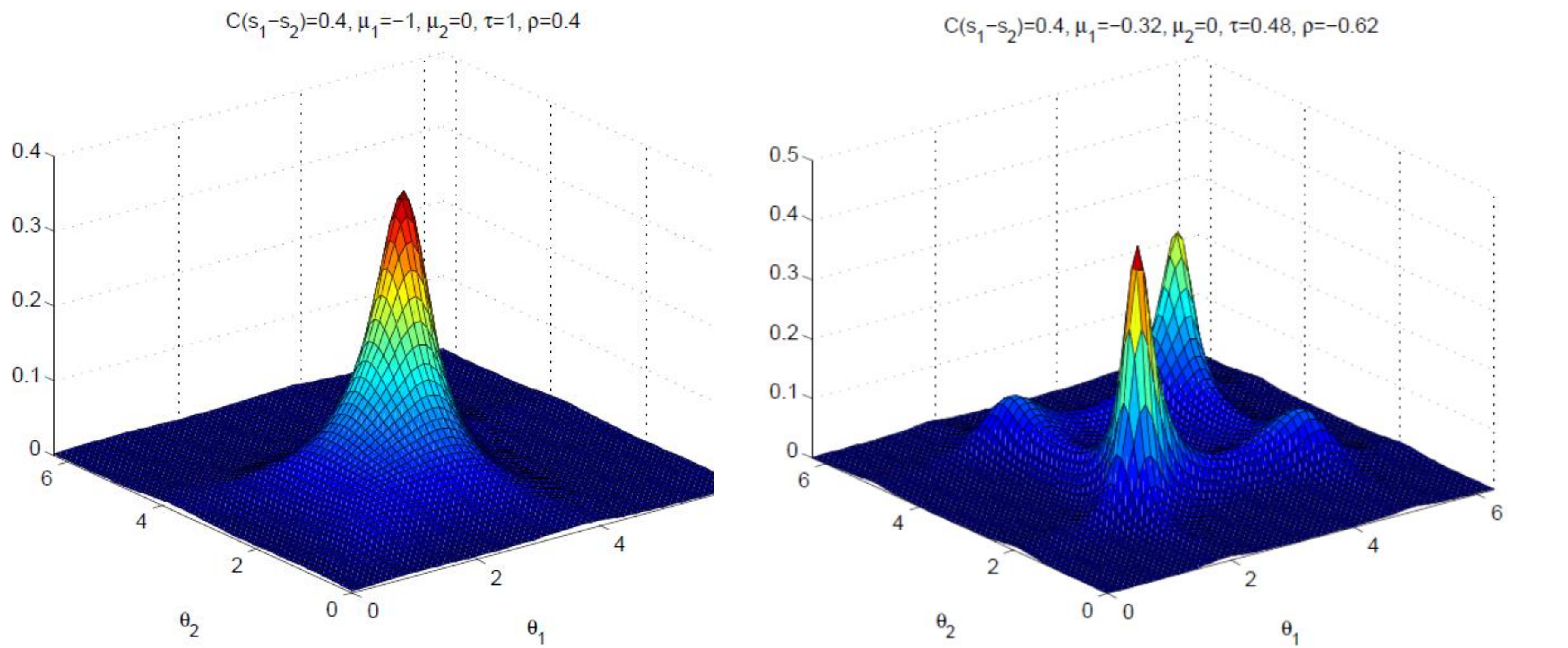
- It is difficult to obtain an explicit form for ρ_c under projected normal process model. However, we can approximate this quantity using Monte Carlo intergration.
- Fisher's definition of circular correlation coefficient doesn't apply in the spatial setting.

Some joint distributions





Two joint distributions from separable correlation function with same cross correlation



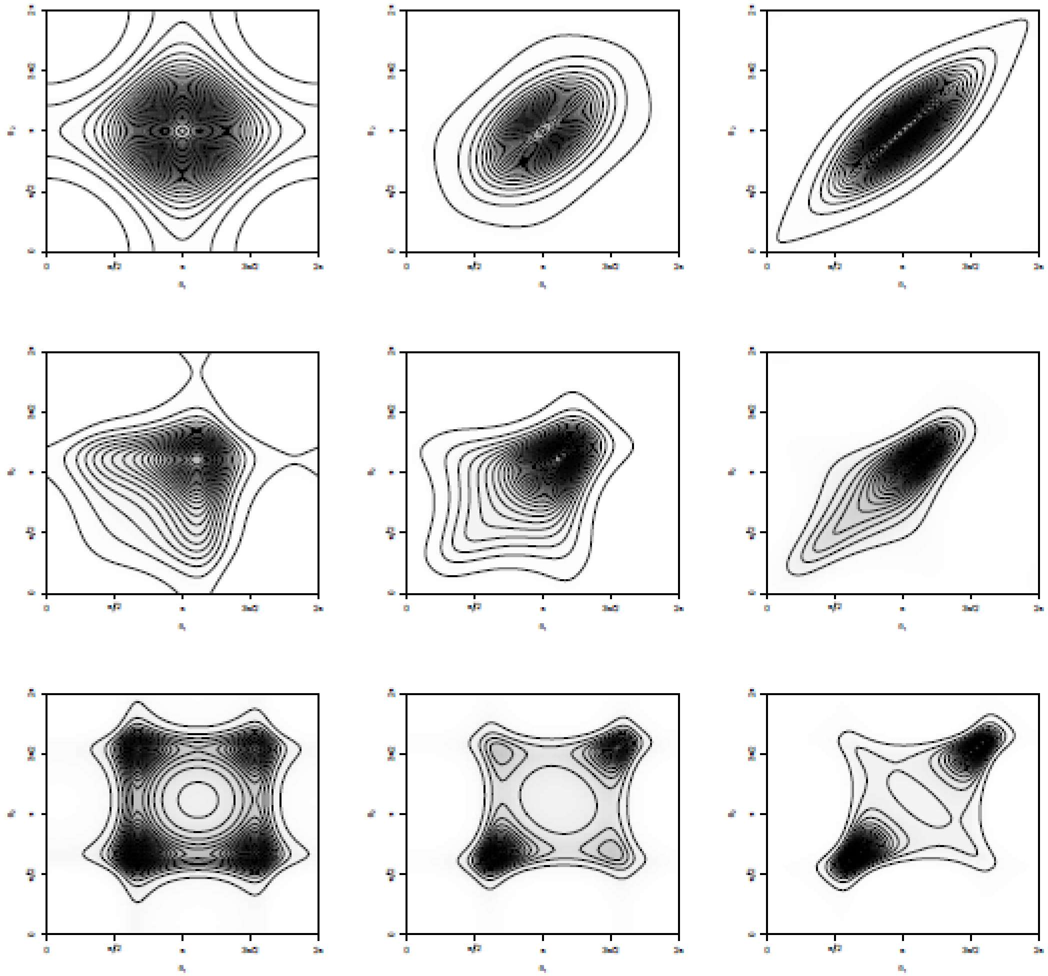
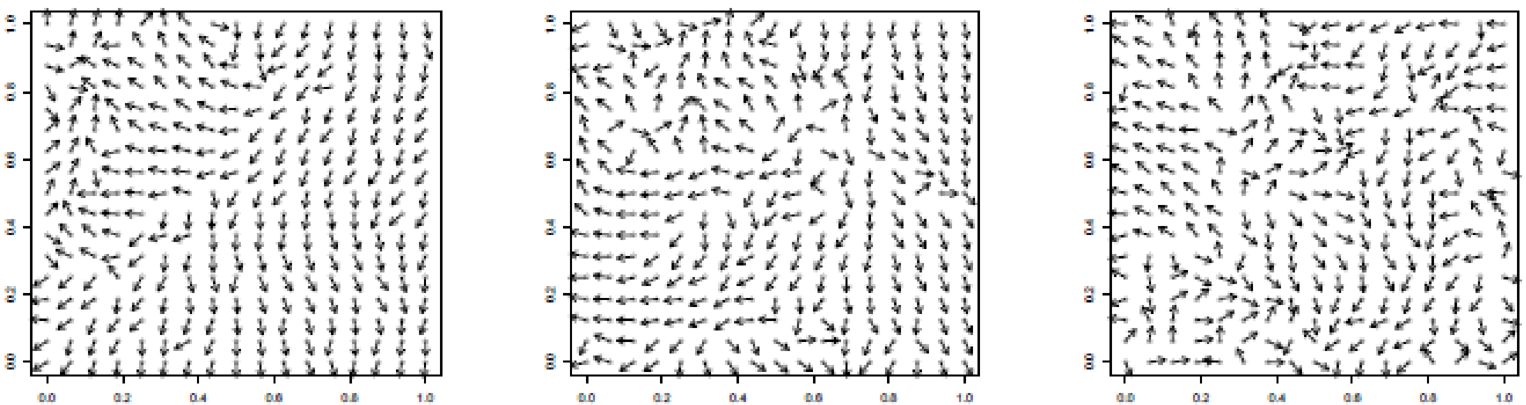


Figure 3. Bivariate joint distribution of $\Theta(\mathbf{s})$ and $\Theta(\mathbf{s}')$ of three different marginals (rows) and three different levels of spatial dependence (columns). Explicit values of the parameters are provided in the text



$(a)\phi = 1$ $(b)\phi = 2$ $(c)\phi = 5$

Figure 6. Simulated draws from projected Gaussian process with different values of ϕ.

Italian wave direction data

Fourteen buoys of RON

Available Data Sources:

- WAM (WAve Model) data in **deep water** on a grid (25×25 km cells)
- RON (Rete Ondamerica Nazionale - National Wave-measure Network) data

Data of Interest:

- Wave Heights (H)
- Wave Directions (D) -circular data-

Aim: assimilation of values produced by WAM with data recorded by RON in order to improve (calibrate) WAM estimates. The final target is to use these data on a higher resolution grid (10×10 km) to perform coastal prediction (**shallow waters**)

Real Data Example

Wave direction at 224 locations, we hold out observations at 50 locations.

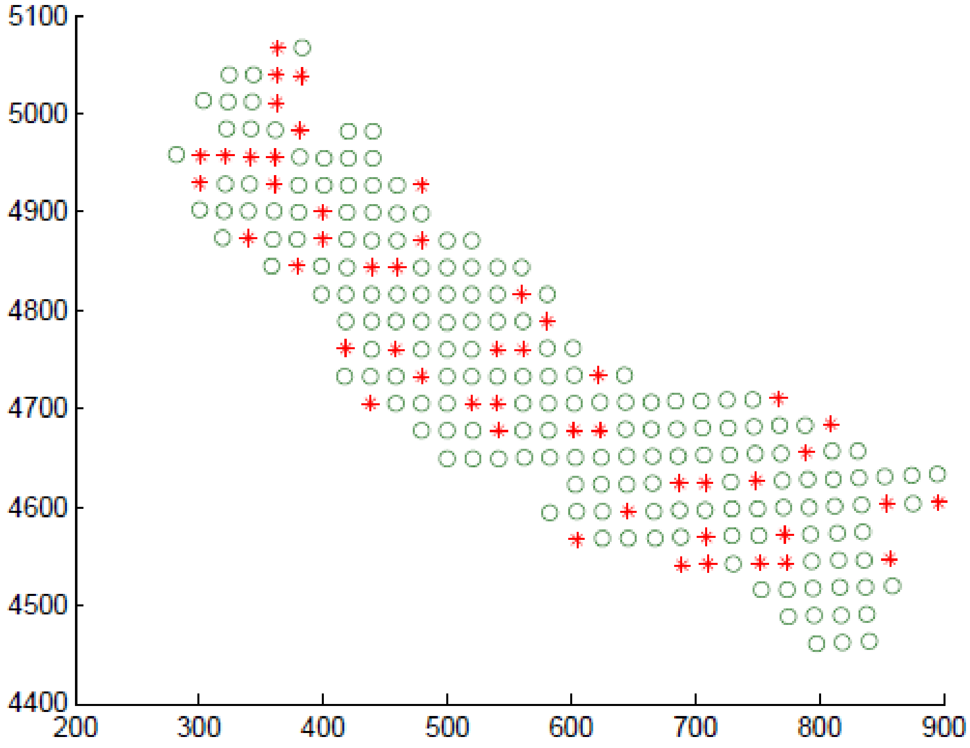
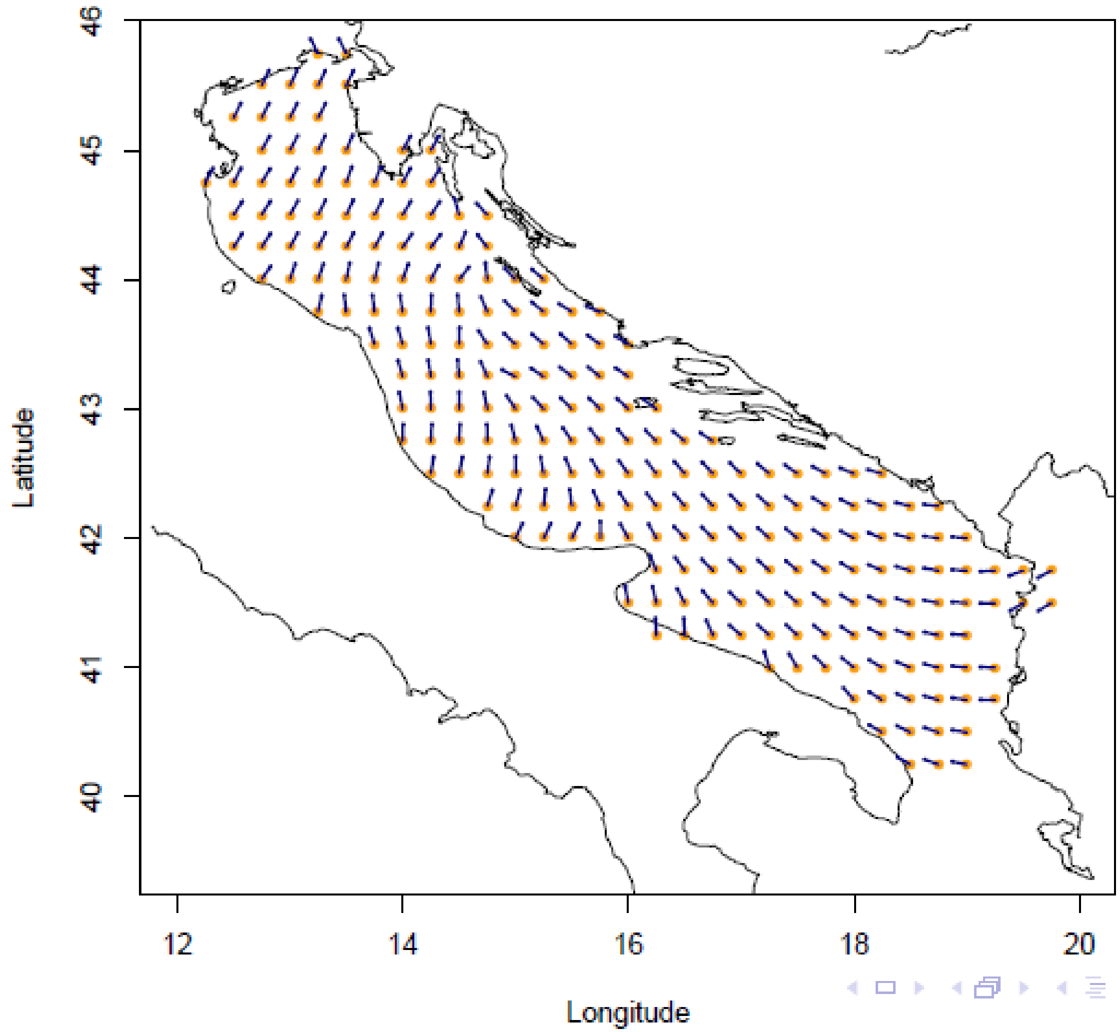


Figure: Training set (circles) and holdout set (stars)

Adriatic Wave Data

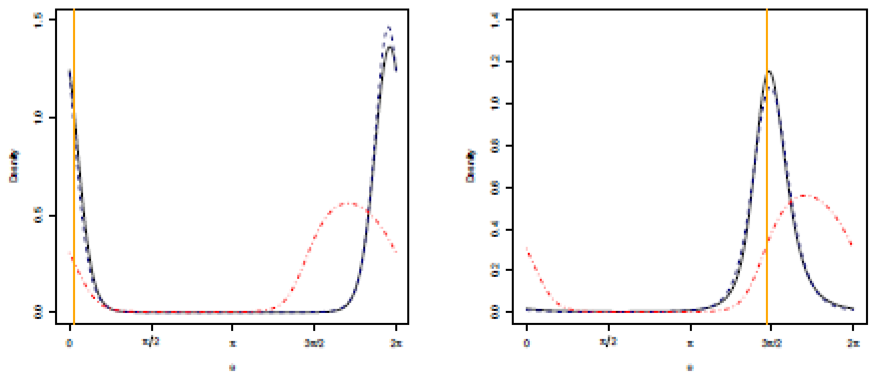
- Data outputs from a deterministic wave model implemented by ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale) for the Adriatic Sea
- On a grid with $\approx 12.5 \times 12.5$ km cells.
- A random set of 250 irregular locations, 50 for validation
- Static spatial analysis - a single time slice (hour) separately during a calm period and a stormy period
- Very strong spatial dependence for the directions yields very slow decay implying a range beyond the largest pairwise distance in our dataset.

Dec 3, 2002, 0:00

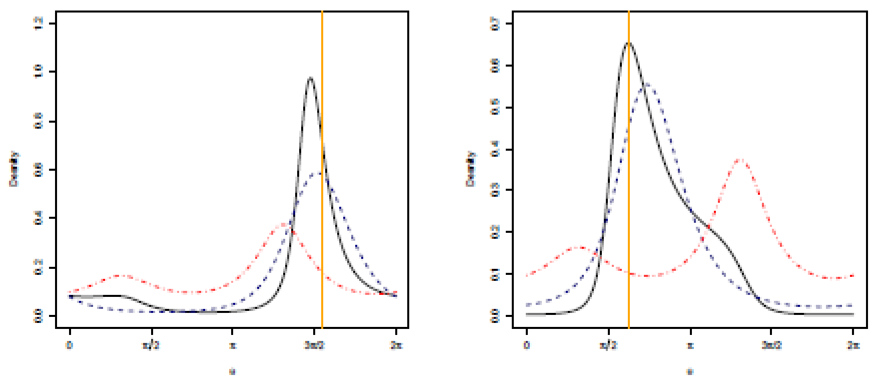


Inference

- For kriging, CRPS values in calm period are 0.0889 (genPGP), 0.0970 (T=I PGP) and 0.6860 nonspatian genPGP)
- The PLSL values are -62.81, -39.49 and 172.13.
- For storm, CRPS values are 0.0726, 0.0682 and 0.5432, PLSL values are -110.14, -99.17 and 140.46
- General PGP and wrapped GP comparison - kriging for hold out locations and corresponding circular distances
- For calm time slice, 0.0222 for general PGP, 0.3743 for the wrapped GP; for storm time slice, 0.0217 and 0.1516
- More variation in wave directions during a calm period. General PGP outperforms T=I PGP.
- In storm not much difference. Concentration in a common direction; data does not require genPGP



(a) asymmetric marginal and long range spatial dependence



(b) bimodal marginal and long range spatial dependence

Figure 9. Predictive density at two hold out locations: the solid line is the predictive density for the projected Gaussian model with general T, the dashed line is for the projected Gaussian with $T = I$ and the dash-dot line is for the non-spatial model. The vertical lines denote the held out angle.

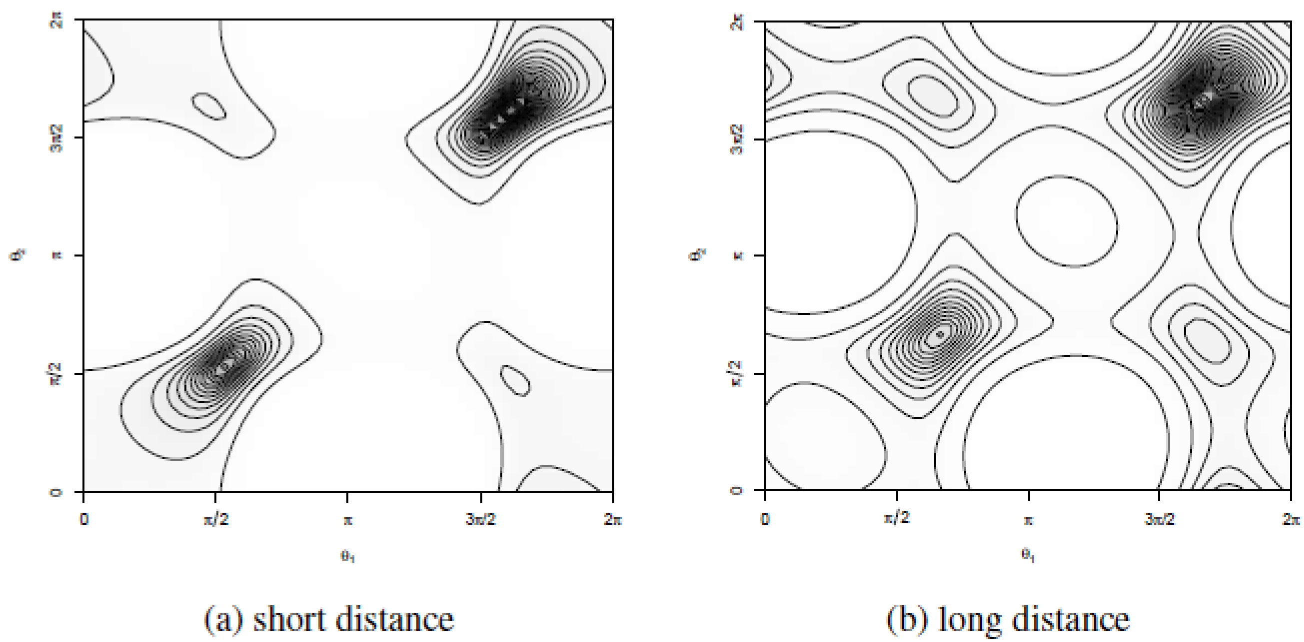


Figure 11. Posterior bivariate density plot based on observations during a calm period.

Space-time Models

- Wave directions are available over 24 time points, two hours apart
- Linear space-time bivariate GP induces a space-time genPGP
- Circular variables $\Theta(\mathrm{s}, t)$ over $\mathrm{s} \in D$ and $t \in (0, \Gamma)$.
- Bivariate space-time GP, constant means, separable covariance structure, T as earlier

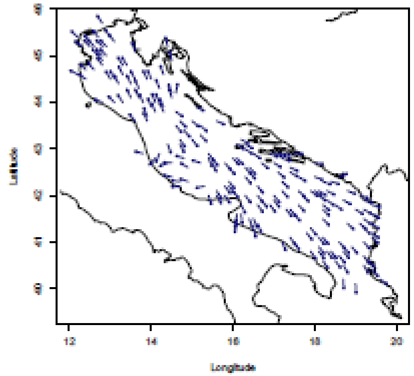
$$\mathrm{cov}(\theta(\mathrm{s}, t), \theta(\mathrm{s}', t') = \rho_{\mathrm{s}}(\mathrm{s} - \mathrm{s}'; \phi_{\mathrm{s}})\rho_t(t - t'; \phi_t) \cdot T,$$

- Again, only one space-time covariance function. Assume separability in space and time
- One-step ahead prediction; with $t = t_1, t_2, ..., t_k$, forecast to t_{k+1} for the n observed locations

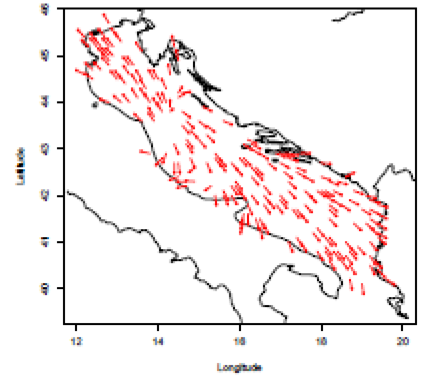
Space-time wave direction example

- Again, data during a calm period and during a stormy period, each over 24 time points two hours apart.
- For the calm period, outputs from April 2nd, 2010 and April 3rd, 2010, for the storm period, from April 5th, 2010 and April 6th, 2010.
- Observations are provided at regular discrete time points, that is, 00:00, 02:00 $, \ldots ,$22:00.
- Hold out data and predict for $t = 24$
- Model adequacy: empirical vs 90% nominal coverage for the 200 predictions. For the calm period, 88.5%; for the stormy period, 95.5%
- Compare the observed vs. predicted at $t = 24$ at each of the 200 locations

Calm period, 22:00 on April 3rd, 2010

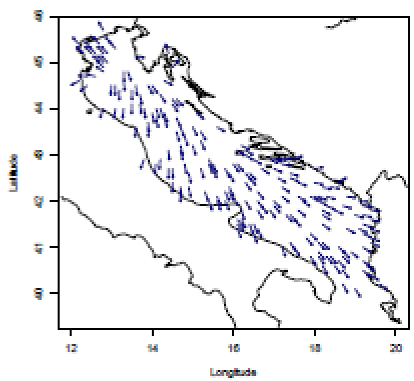


(a) Data

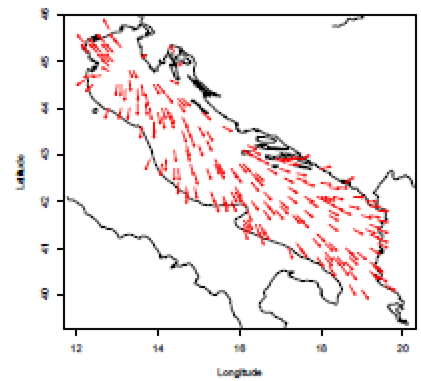


(b) Kriged result

Stormy period, 22:00 on April 6th, 2010



(a) Data



(b) Kriged result

Figure 15. Comparison of hold out data and model fitting results for space-time model.

Challenge of a dynamic model

- Circular variables, $\theta_t(\mathrm{s}_i)$ at each s_i and each $t = 1, \ldots, N_t$, $i = 1, \ldots, n$.
- Want a hierarchical model to capture both spatial and temporal dependence
- At the data level, would assume $\theta_t(\mathrm{s}_i)$ follow a conditionally independent genPGP with time-varying parameters Φ_t and say a common separable cross-covariance.
- At the second level, need to specify dynamics for Φ_t.
- Strong interplay between the parameters in Φ_t with regard to distributional shape $\Rightarrow$ a four dimensional dynamic second stage specification with dependence between all of the parameters

Wave direction-wave height models

- Realizations of a circular variable Θ and a linear variable X as pairs $(\theta_1, x_1), \ldots, (\theta_n, x_n)$.
- Usually, *conditional* modeling.
- If $\Theta|X$, *Linear-Circular regression* model.
- Usually, von Mises distribution with a suitable link function, $\mathbb{R}^k$ to $(-\pi, \pi)$, e.g., the arctan function
- If $X|\Theta$, *Circular-Linear regression* model.
- Again, a link function is employed; A flexible approach is to employ trigonometric polynomials

Measuring association between a circular and a linear variable

- The pair (Θ, X) has the support on a cylinder.
- If their relationship can be written as $X = a + b\cos(\Theta) + c\sin(\Theta)$, then ordinary multiple correlation in the regression setting between X and $(\cos(\Theta), \sin(\Theta))$ (Johnson-Wehrly-Mardia).
- Our modeling in this spirit, conditioning X on Θ.
- A joint model through a conditional $\times$ marginal specification; marginal specification is a space or space-time directional data process and, conditionally, a space or space-time linear process.
- For wave height given wave direction we find a *natural* link function to be as above.

The model details

- A single linear variable and a single directional variable, e.g., wave height and wave direction.
- A joint parametric model for wave height H and wave direction Θ by introducing a latent R in the form

$$f(H, \Theta | \mathbf{\Psi}_h, \mathbf{\Psi}_\theta) = \int f(H | \Theta, R, \mathbf{\Psi}_h) f(\Theta, R | \mathbf{\Psi}_\theta) dR$$

- $f(H | \Theta, R, \mathbf{\Psi}_h)$ is a customary *geostatistical* model, $H(\mathrm{s}) = g(\Theta(\mathrm{s}), R(\mathrm{s})) + w(\mathrm{s}) + \epsilon(\mathrm{s})$
- Residual: spatial effect $w(\mathrm{s})$, a mean 0, stationary GP, and uncorrelated errors term $\epsilon(\mathrm{s})$
- Link function $g(\cdot)$: the linear regression of $H(\mathrm{s})$ on $Y_1(\mathrm{s})$ and $Y_2(\mathrm{s})$ from the “unobserved” linear GP

cont.

- Explicitly, we have

$$H(\mathrm{s}) = \beta_0 + \beta_1 R(\mathrm{s})\cos\Theta(\mathrm{s}) + \beta_2 R(\mathrm{s})\sin\Theta(\mathrm{s}) + w(\mathrm{s}) + \epsilon(\mathrm{s})$$

$$= \beta_0 + \beta_1 Y_1(\mathrm{s}) + \beta_2 Y_2(\mathrm{s}) + w(\mathrm{s}) + \epsilon(\mathrm{s})$$

- Altogether, $(H(\mathrm{s}), Y_1(\mathrm{s}), Y_2(\mathrm{s}))^T$ specifies a trivariate GP
- Mean structure and cross-covariance structure easily calculated.

cont.

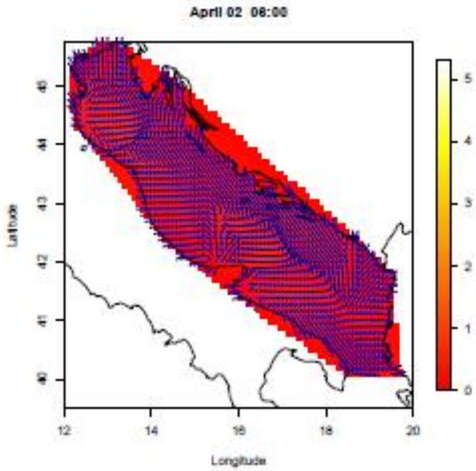
- β_1 and β_2 inform about association between Θ and H.
- When β_1 and β_2 are both 0, the joint model fits the wave heights and the wave directions separately
- Again, the square of the multiple correlation coefficient, $R^2_{H|\mathbf{Y}}$, measures strength of conditional dependence.
- A regression where the covariates are not observed.
- However, if $\Delta = \beta_1^2\tau_\theta^2 + \beta_2^2 + 2\beta_1\beta_2\tau_\theta\rho$, then

$$R^2_{H|\mathbf{Y}} = \frac{\Delta}{\Delta + \sigma_h^2 + \tau_h^2}.$$

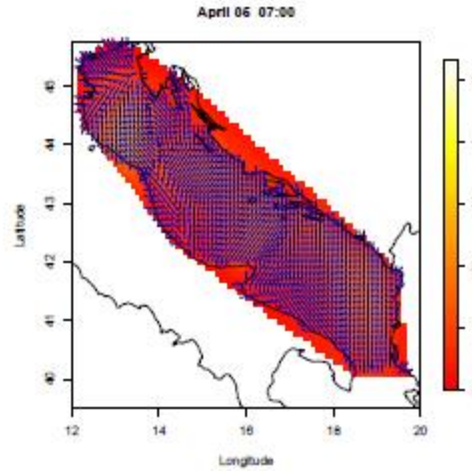
- The posterior distribution of $R^2_{H|\mathbf{Y}}$ directly from posterior samples of the parameters

Back to the Adriatic wave data

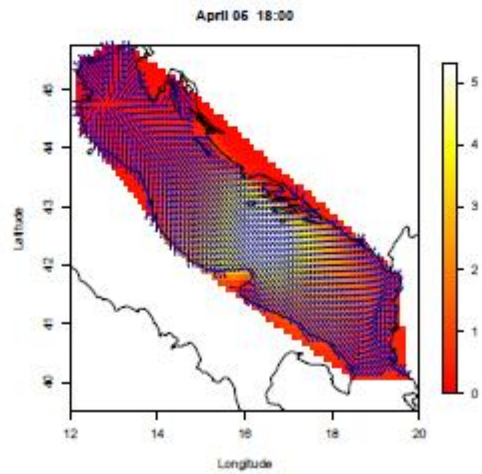
- Data are significant wave heights, essentially the average height of the highest waves in a wave group.
- Outgoing wave directions preferred to incoming, in degrees relative to a fixed orientation.
- Over the Adriatic Sea area, we have 1494 points.
- Definition: “calm” is wave height lower than 1 meter.
- Between 1 and 2 meters, a pre-storm warning
- Above 2 meters a storm warning.
- An illustrative hour under each sea state
- Stationarity assumption is unrealistic due to anticipated orographic influences; so only an initial attempt.
- Compare joint with independent modeling of the wave heights and wave directions



(a) calm



(b) transition



(c) storm

Figure 2: Plot of wave heights (meters) and wave directions under three sea states: (a) calm, (b) transition and (c) storm

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

Feature	(a) calm		(b) transition		(c) storm	
	(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

Table 3: The posterior summaries of the parameter $\boldsymbol{\beta}$

	(calm)			(storm)		
parameter	mean	2.5%	97.5%	mean	2.5%	97.5%
β_0	0.1612	0.0343	0.2816	0.5483	-0.0242	1.1745
β_1	-0.0117	-0.0405	0.0153	-0.1439	-0.2770	-0.0062
β_2	0.0329	0.0091	0.0549	0.2061	0.0670	0.3564

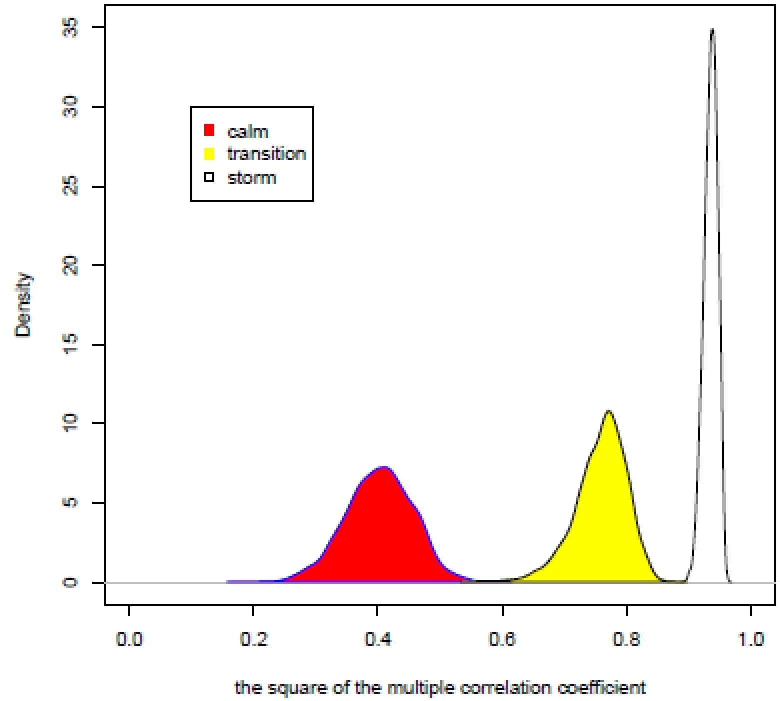


Figure 4: The posterior densities of the square of the multiple correlation coefficient $R^2_{H\,|\,\mathbf{Y}}$

Space-time directions and heights

- Envision an underlying process for heights and directions in continuous space and time
- Hourly resolution using the ISPRA output
- 10 illustrative locations, time series of hourly heights, first ten days in April, 2010
- A range of behavior, including a transition from calm to storm and back to calm.
- At the conditional level,

$$H(\mathrm{s},t) = \beta_0 + \beta_1 Y_1(\mathrm{s},t) + \beta_2 Y_2(\mathrm{s},t) + w(\mathrm{s},t) + \epsilon(\mathrm{s},t).$$

- Simplify $w(\mathrm{s},t)$ to $w(\mathrm{s})$, a zero-centered GP. Otherwise, identifiability challenges
- Note that β, changes with sea state so we only fit the model in either a calm window or a storm window.

An Example

- Randomly select 200 locations; goal is prediction at t_{k+1}
- Calm period, April 2nd to April 3rd, 2010, 24 time points.
- Storm period, April 5th to April 6th, 2010, 24 time points.
- During the calm period, β_1 and β_2 are nearly zero, approximate independence
- During a storm period, β_1 and β_2 clearly depart from zero, strong dependence
- σ_h^2 dominates the variation in heights, a twenty-fold increase during the storm period.
- $\phi_{\theta,\mathrm{s}}$ is roughly double (range roughly half) in calm compared with storm; $\phi_{\theta,t}$ is roughly half (range roughly double) in calm compared with storm
- Prediction at t_{24}. Predict direction better in a storm and height dramatically better during a calm

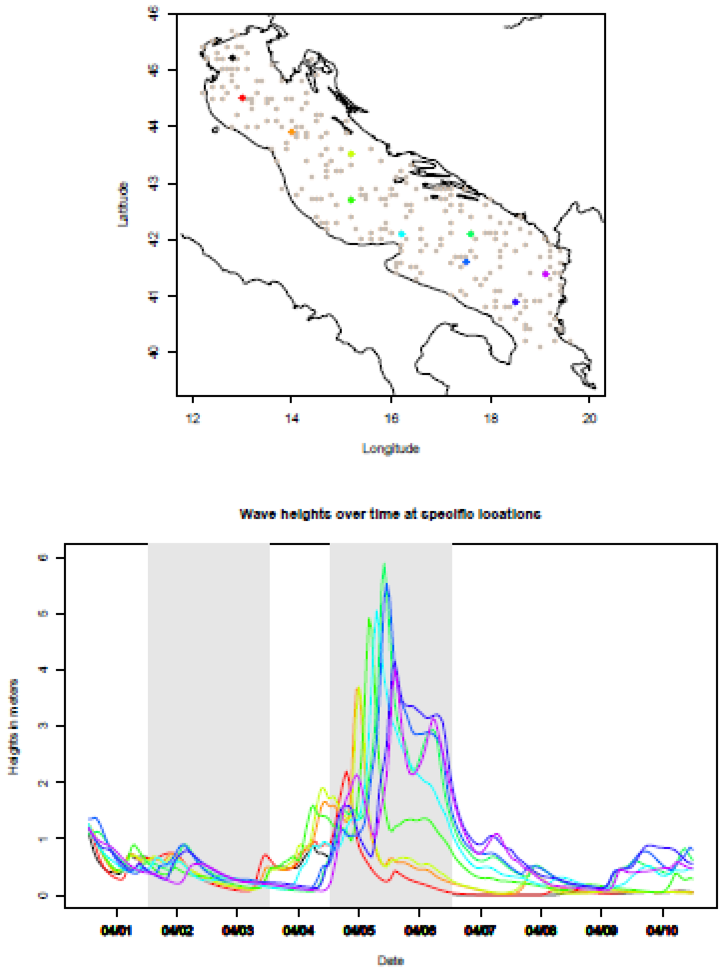


Figure 5: Time series of wave heights (lower panel) for 10 illustrative locations (upper panel) from $n = 200$ in the Adriatic Sea region in April 2010. The two time windows (calm and stormy) used in Section 3.3 are highlighted.

Table 4: Summaries for one-step ahead prediction

	calm	storm
Predictive Mean Square Error (height)	0.0522	0.4497
mean CRPS for wave direction	0.0973	0.0647

Some New Work

Aspect process

- To estimate surface gradients and slopes from a (possibly) irregularly spaced dataset
- Usual approaches are descriptive - projection of the elevation data to a regular grid, slope calculated at each point of the grid by comparison of the elevation to a set of neighboring grid points.
- Usually the 8 compass neighbors (N, NE, E, SE, S, SW, W and NW) are used
- *Aspect* is direction of max slope
- Though slope and aspect are obtained as angles, no awareness of the literature on angular variables

cont.

- If elevation surface is viewed as a realization of a Gaussian process, then,
- max slope at $\mathbf{s}$ is $R(\mathbf{s}) = ||\nabla_Y(\mathbf{s})||$. So we have a max slope process and
- direction of max slope is $\mathbf{u} = \frac{\nabla_Y(\mathbf{s})}{||\nabla_Y(\mathbf{s})||}$ and, formally, the aspect at $\mathbf{s}$ is the angle associated with this maximum slope direction, i.e., $\theta_{asp}(\mathbf{s}) = \texttt{atan2}\frac{D_{(0,1)}Y(\mathbf{s})}{D_{(1,0)}Y(\mathbf{s})}$, an aspect process
- Starting with a bivariate GP, we can model the aspect process as a projected GP
- Fully resolved with full inference for slope and aspect
- Computational challenge

Angular discrepancy process

- Suppose we have two dependent surfaces we seek to compare, e.g., a surface of species abundance $Y(\mathbf{s})$ and a surface of elevation, $X(\mathbf{s})$.
- Suppose the surfaces are described as a realization of a bivariate Gaussian process
- Associated with each surface we can obtain directional gradients at each location, $D_{\mathbf{u}}Y(\mathbf{s})$ and $D_{bu}X(\mathbf{s})$
- We can explicitly obtain the direction of max gradient at each location for each surface: $\frac{\nabla_Y(\mathbf{s})}{||\nabla_Y(\mathbf{s})||}$ and $\frac{\nabla_X(\mathbf{s})}{||\nabla_X(\mathbf{s})||}$.
- Associated angles are: $\theta_Y(\mathbf{s})$ and $\theta_X(\mathbf{s})$.
- Using the circular distance, $1 - \cos(\theta_Y(\mathbf{s}) - \theta_X(\mathbf{s}))$, yields an *angular discrepancy process*
- A useful sensitivity measure for comparing a response and a covariate

Point patterns on a circle

- Suppose we look at crime times during a period, say a week, for a city (or a district within a city)
- We wrap the times onto a circle; the number of times (crimes) is random, so a point pattern on a circle
- To model the point pattern, an intensity function
- If the intensity function is viewed as a realization of a stochastic process over a circular domain, we need a *circular covariance function*
- An example is $K(\theta) = e^{-\frac{2\sin(\theta/2)}{\theta}}$ (Soubeyrand et al.)
- Suppose we add the spatial location of the crime. Now, a space-time point pattern with circular time
- Now, a space-time intensity over $D \times (0, 2\pi]$ with space-time interaction
- Computational challenges