CBMS Lecture 10

Alan E. Gelfand
Duke University

Outline

- 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

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

Challenges

- Support restriction is not just $[0, 2\pi)$ but circularity, i.e., sensitivity to the starting point.
- An angle or, equivalently, a real number on $[0, 2\pi)$ given a fixed orientation
- However, inference should not depend upon "choice of origin", "sense of rotation"
- Circle is very different topologically from the line; beginning coincides with the end
- "Direction has no magnitude." No ordering or ranking.
- Is it 2-dim, e.g., angle associated with resultant of a N-S direction and an E-W direction
- Sample mean and variance don't mean anything, e.g., for the sample $1^o, 0^o, 359^o$, sample mean is 120^o! Clearly 0^o more sensible. Sample variance is also silly.

Circular distributions

- A probability distribution whose entire mass is on the circumference of a unit circle
- We work with the absolutely continuous case (w.r.t. Lebesgue measure on the circle) with density $f(\theta)$.
- Properties:
  - cdf: $F(\theta + 2\pi) - F(\theta) = 1$, $\theta \in R^1$
  - with density: $f(\theta) \geq 0$
  - $\int_0^{2\pi} f(\theta) d\theta = 1$
  - $f(\theta + 2k\pi) = f(\theta)$ for any integer k (f is periodic)

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. Computationally unattractive.

Moment properties

- Expectations under circular densities are hard to compute, e.g., for the Von Mises, wrapped distributions and projected normals. Work with associated complex variable on the unit circle in the complex plane, $Z = e^{i\theta}$.
- $E(Z^r) = E(e^{ir\theta}) \equiv \rho_r e^{i\mu_r}$, i.e, value of the characteristic function at r.
- Thus, we have $\rho_r \cos(\mu_r) = E\cos(r\theta)$ and $\rho_r \sin(\mu_r) = E\sin(r\theta)$.
- When $r = 1$, we have $\rho\cos\mu = E(\cos\theta)$ and $\rho\sin\mu = E(\sin\theta)$.
- Solving for the **mean direction** $\mu = \text{atan2}\frac{E\sin(\theta)}{E\cos(\theta)}$
- We have the **resultant/concentration** $\rho(\equiv c) = \sqrt{(E\cos(\theta))^2 + (E\sin(\theta))^2} \leq 1$

Wrapping

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

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

- Obviously, can rescale from $[0, L)$ to $[0, 2\pi)$
- Multivariate version (say p-dim) is easy to specify. 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 $0 \le \theta < 2\pi$:

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

- $E(Z) = e^{-\sigma^2/2} e^{i\mu}$ 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 $f(\theta, k) = g(\theta + 2k\pi) =$

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

- 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 specified 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

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

where $R(\phi)_{ij} = \rho(s_i - s_j; \phi)$

Remarks

- A common mean, μ, for all locations, hence a common $\tilde{\mu}$. A regression form would be for $\tilde{\mu}(s)$ and requires a suitable link function
- Can directly extend to wrapped t-process using usual mixing of GP to t-process
- We define the wrapped GP through the dependence structure of the linear GP.
- Correlation measure for a pair of directions? Can we obtain a sensible induced covariance function?

Association structure

- What should association between $\theta(s)$ and $\theta(s')$ mean?
- What good properties should we require for a correlation between dependent directions?
- In particular, how can we connect the covariance function of the linear GP to that of the wrapped GP?
- Properties of a circular correlation coefficient:
  - $\rho_c(\theta_1, \theta_2)$ should not depend upon the "zero" direction
  - $\rho_c(\theta_1, \theta_2) = \rho_c(\theta_2, \theta_1)$
  - $|\rho_c(\theta_1, \theta_2)| \leq 1$
  - $\rho_c(\theta_1, \theta_2) = 0$ if θ_1, θ_2 indep
- Jammalamadaka and Sarma (1988) provide the following measure which satisifies the above properties

$$\rho_c(\theta_1, \theta_2) = \frac{E(\sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2))}{\sqrt{\mathrm{Var}(\sin(\theta_1 - \mu_1))\mathrm{Var}(\sin(\theta_2 - \mu_2))}}$$

For the Wrapped Normal

- For the WN, this measure takes a simple form
- For $\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} \sim \mathrm{WN}(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix})$, we have $\rho_c(\theta_1, \theta_2) = \frac{\sinh(\rho\sigma^2)}{\sinh(\sigma^2)}$
- So, if linear GP has cov function $\sigma^2\rho(s, s')$, induced covariance function is $\sinh(\sigma^2\rho(s, s'))$
- Easy to see that this is a valid covariance function

Fitting a wrapped GP model

- Introduce latent K_i's. With a WN prior for μ and a right-censored inverse Gamma prior for σ^2
- What about ϕ?
- With Matérn correlation function, we have difficulty identifying both σ^2 and ϕ.
- However, can work with uniform priors for ϕ, jointly updating ϕ and σ^2
- The full conditionals for the K_i are immediate from the conditional normal distribution of $[\theta_i + 2\pi k_i | \theta_j + 2\pi k_j, j \neq i, \mu, \sigma^2, \phi]$
- We use adaptive truncation for the K_i's as above. Discard posterior samples of $\mathbf{K}$; only interested in those for μ, σ^2
- Kriging is straightforward

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 $\theta = \text{atan2}\frac{V_2}{V_1} = \text{atan2}\frac{U_2}{U_1}$
- In fact, $U_1 = R\cos\theta$ and $U_2 = R\sin\theta$, R latent
- $R = ||\mathbf{U}||$, $V_1 = \cos\theta$, $V_2 = \sin\theta$
- Again, angular mean direction is $\text{atan2}\frac{E\sin\theta}{E\cos\theta} = \frac{E(V_2)}{E(V_1)} \neq \frac{E(U_2)}{E(U_1)}$
- 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 $\boldsymbol{\mu}$ - 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 dimension > 2.

Projected normal

- 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
- $f(r, \theta | \boldsymbol{\mu}, \Sigma) = (2\pi)^{-1} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{(r\mathbf{u} - \boldsymbol{\mu})' \Sigma^{-1} (r\mathbf{u} - \boldsymbol{\mu})}{2}\right) r$

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 $\mu = \text{atan2}\frac{\mu_1}{\mu_2}$, 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.

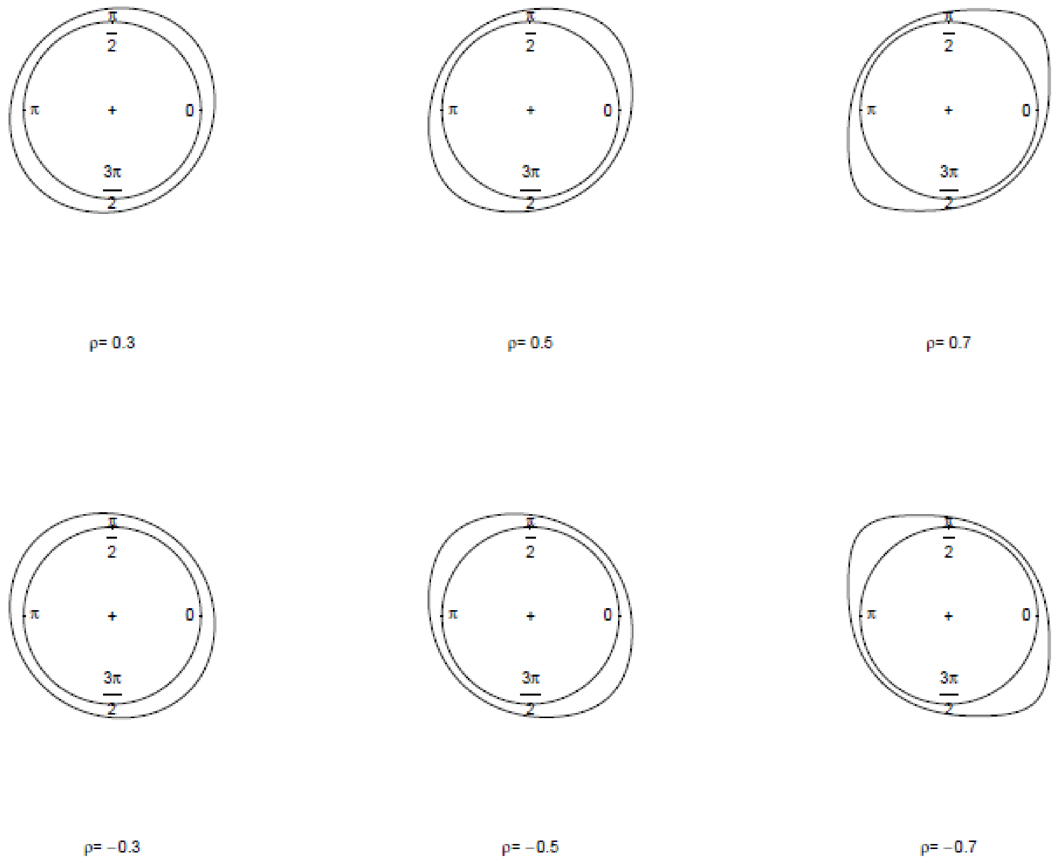
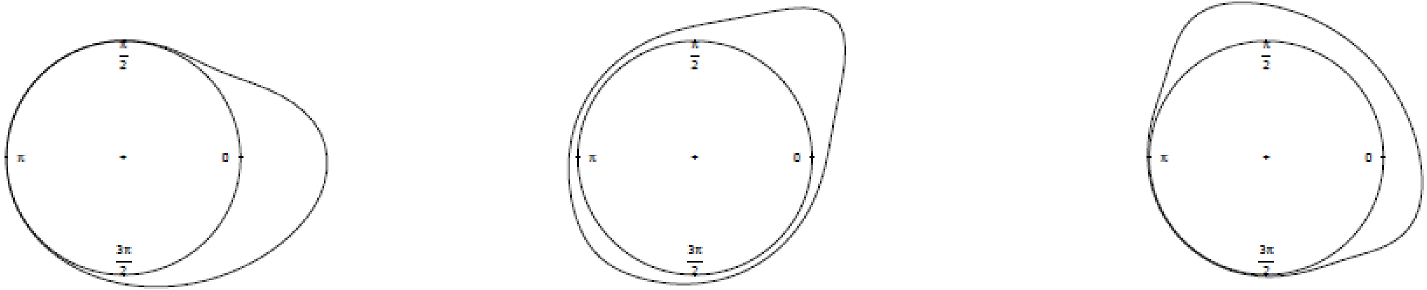


Figure 2. Density of θ for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = \sigma$ and different values of ρ



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

Figure 3. Shape of the general projected normal distribution

cont.

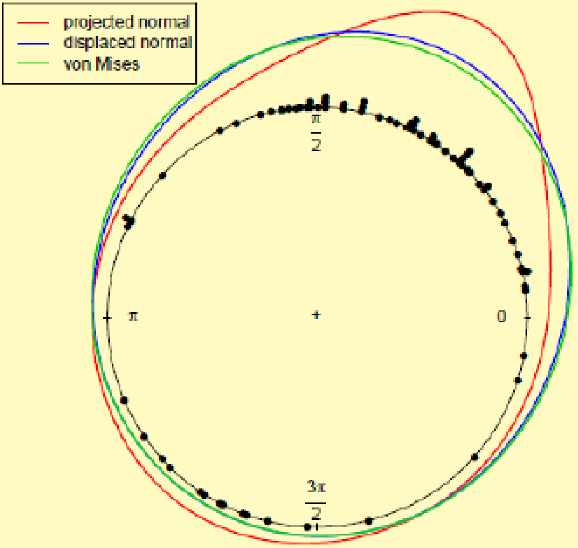
- We work with the more general Σ case
- Can draw pictures of the density in terms of five parameters in $\boldsymbol{\mu}$ and Σ. We can achieve asymmetry and bimodality
- With regard to inference, an identifiability issue: Note that if we scale $\mathbf{U}$ by a, the distribution of θ doesn't change
- To make identifiable, set $\Sigma = \begin{pmatrix} \tau^2 & \rho\tau \\ \rho\tau & 1 \end{pmatrix}$
- We have a four parameter model

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)

Real Data Example

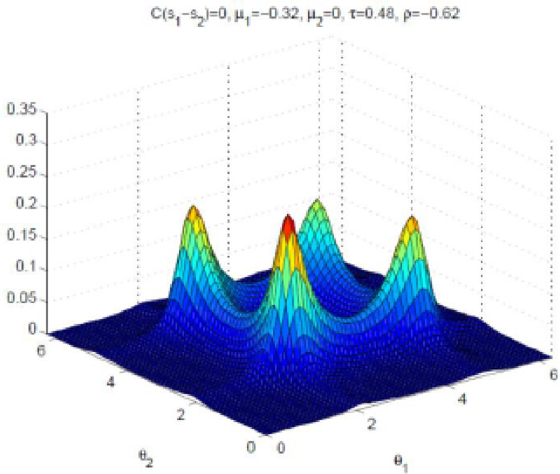
The raw data are the directions in which 76 female turtles moved after laying their eggs on a beach. (Gould's data)

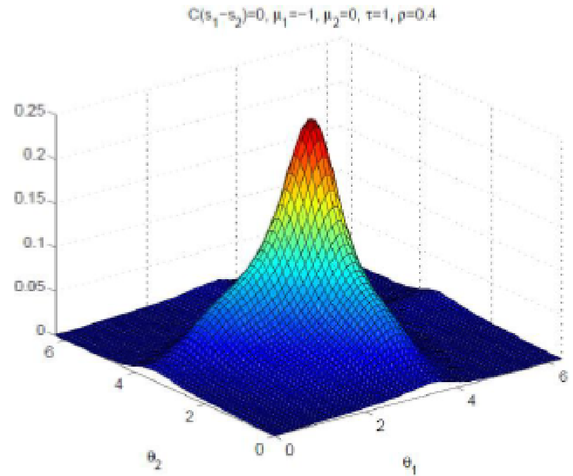


Spatial PN models

- Finally, we return to the case of $\{\theta(s_i), i = 1, 2, ..., n\}$
- In the independence case, we had latent independent $\mathbf{U}_i$'s modeled as bivariate normal
- Now, we assume latent $\mathbf{U}(s_i)$ 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 $\mathbf{U}(s_0)$ 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.

Some joint distributions





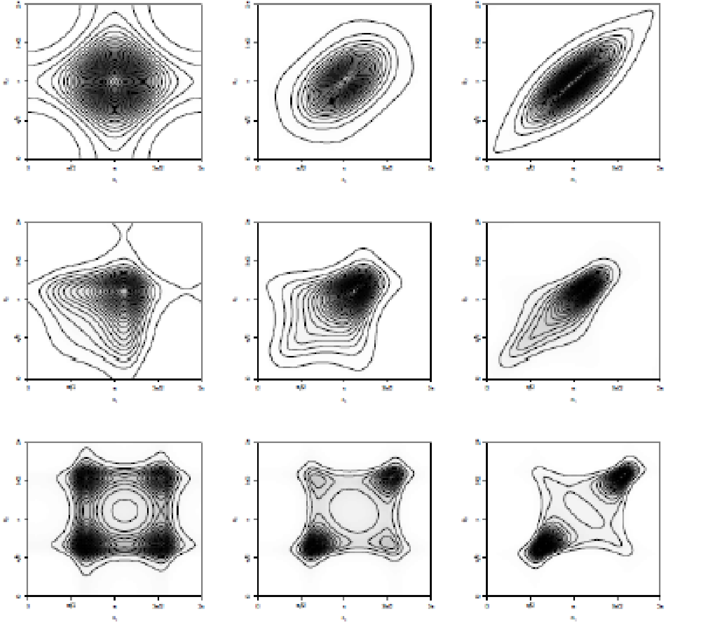
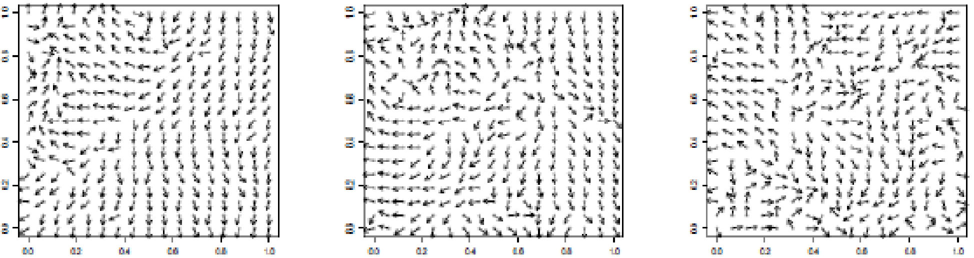


Figure 3. Bivariate joint distribution of $\Theta(\mathbf{s})$ and $\Theta(\mathbf{s}')$ of three different marginals (rows) and three different levels of spatial dependence (columns). Explicit values of the parameters are provided in the text



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

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

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

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.

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**)

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

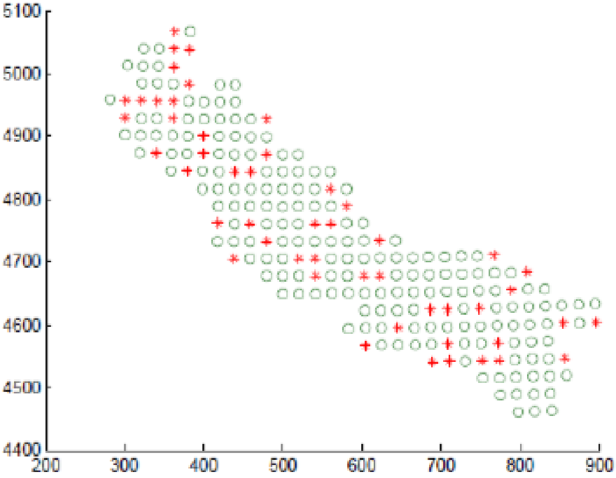
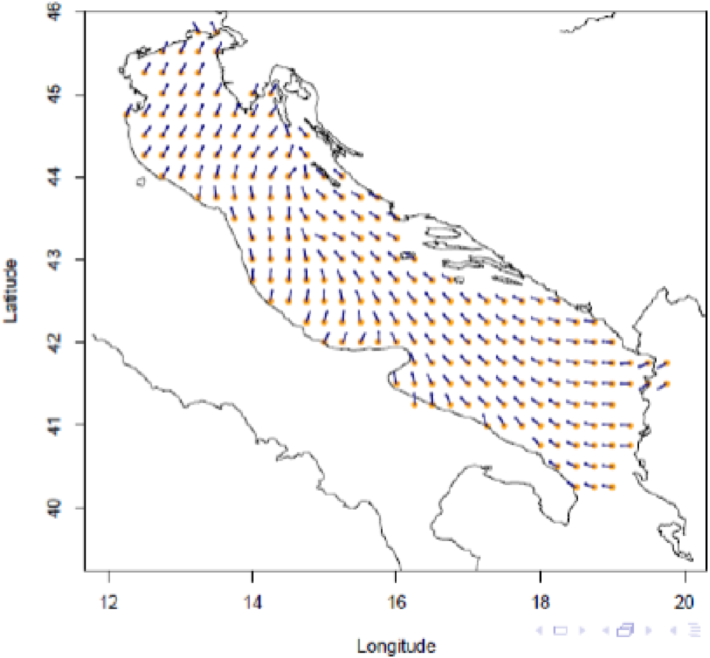


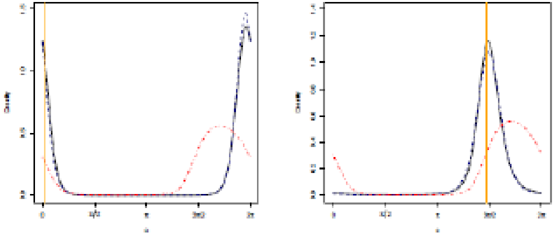
Figure: Training set (circles) and holdout set (stars)

Dec 3, 2002, 0:00

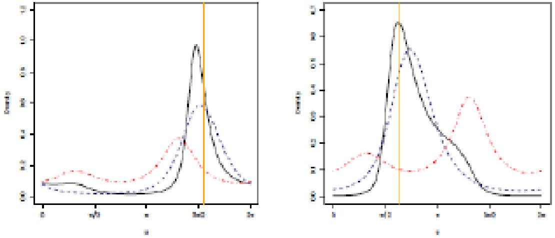


Inference

- For kriging, CRPS values in calm period are 0.0889 (genPGP), 0.0970 (T=I PGP) and 0.6860 nonspatial 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

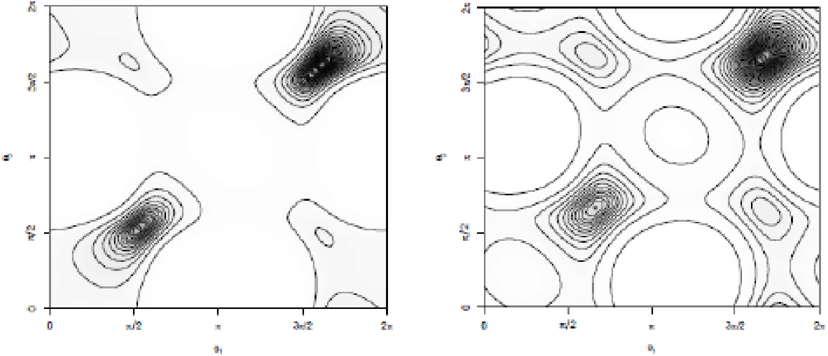


(a) asymmetric marginal and long range spatial dependence



(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.



(a) short distance

(b) long distance

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