Analyzing Spatial Directional Data

Alan E. Gelfand
Duke University
(Joint work with Giovanna Jona Lasinio and Fangpo Wang)
Outline of the Talk

- Review of basics of directional data
- Wrapped normal distributions and wrapped Gaussian processes
- Kriging for directional data
- Examples
- Projected normal distributions and projected Gaussian processes
- Examples

Alternative title could be “How to use Gaussian processes for the analysis of spatial directional data”
Directional data

Directional data, circular data, angular data (in 2 dimensions)

Higher dimensional (on spheres)

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)

Usually convert to \([0, 2\pi)\)

Some of these applications can be spatial
The literature


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
Example: Real angles

**Figure:** Orientations of 76 turtles after laying eggs (Gould's data cited by Stephens, 1969)
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:**
- \( f(\theta) \geq 0 \)
- \( \int_{0}^{2\pi} f(\theta) d\theta = 1 \)
- \( f(\theta + 2k\pi) = f(\theta) \) for any integer \( k \) (\( f \) is periodic)
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 $1^\circ, 0^\circ, 359^\circ$, sample mean is $120^\circ$! Clearly $0^\circ$ more sensible. Sample variance is also silly.
Simulation from a von Mises distribution

Sample mean $\bar{x} = 3.666$

Sample Variance $s^2 = 7.904$

Circular mean $\tilde{\mu} = \arctan \frac{\sum \sin x_i}{\sum \cos x_i} = 6.239$

Circular Concentration $\tilde{\tau} = \sqrt{\left(\sum \cos x_i\right)^2 + \left(\sum \sin x_i\right)^2} = 0.936$
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, \( Z = e^{i\theta} \).

\[
E(Z^r) = E(e^{ir\theta}) \equiv \rho_r e^{i\mu_r}, \text{ i.e, value of the characteristic function at } r.
\]

- 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 \( \mu = \arctan^* \frac{E \sin(\theta)}{E \cos(\theta)} \).

- We have the resultant/concentration \( \rho (\equiv c) = \sqrt{(E \cos(\theta))^2 + (E \sin(\theta))^2} \leq 1 \).
Descriptive, exploratory ideas

Data \{ \theta_i, i = 1, 2, \ldots n \}

Equivalently \{ (\cos \theta_i, \sin \theta_i), i = 1, 2, \ldots n \}

Let \( \bar{C} = \frac{1}{n} \sum_i \cos \theta_i, \bar{S} = \frac{1}{n} \sum_i \sin \theta_i \).

Then, we have method of moments estimators using
\( \bar{C} = \hat{\rho} \cos \hat{\mu}, \bar{S} = \hat{\rho} \sin \hat{\mu} \)

That is, \( \hat{\mu} = \arctan^* \frac{\bar{S}}{\bar{C}} \) and \( \hat{\rho} = \sqrt{\bar{C}^2 + \bar{S}^2} \)
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 $\mu$ is mean direction, $\kappa$ 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 (so not flexible)

- For regression, mean direction is a circular parameter so a link function to the circle, e.g., arctan(·) from $\mathbb{R}^1$ to $(-\pi, \pi)$; difficult with multiple regressors

- Hopeless for high dimensional multivariate data

- For spatial or temporal or space-time data, cond’ly indep von Mises with process models for $\mu, \kappa$
Wrapping

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

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

- Convenient choice is a multivariate normal
Projection Approach

- An embedding approach - unit circle within $\mathbb{R}^2$
- $\mathbf{U} = (U_1, U_2) \sim g(u_1, u_2)$, a density on $\mathbb{R}^2$
- Then $(V_1, V_2) = \left( \frac{U_1}{||\mathbf{U}||}, \frac{U_2}{||\mathbf{U}||} \right)$ where $||\mathbf{U}||$ is the length of $\mathbf{U}$, is a point on the unit circle, associated angle is $\theta = \arctan*\frac{V_2}{V_1} = \arctan*\frac{U_2}{U_1}$
- 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 $\arctan*\frac{E\sin\theta}{E\cos\theta} = \frac{E(V_2)}{E(V_1)} \neq \frac{E(U_2)}{E(U_1)}$
- Concentration is $||E(\mathbf{V})|| \leq 1$
Projected normal distribution. Suppose the random vector $\mathbf{U} \sim N_2(\mu, \Sigma)$, then $\theta \sim PN_2(\mu, \Sigma)$.

More flexible - can be asymmetric, bimodal

Easy for regression - linear model in covariates for $\mu$ - 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 dim $>2$. 
The univariate wrapped normal

\textit{WN}(\mu, \sigma^2) \text{ density takes the form, for } 0 \leq \theta < 2\pi:

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

\[ E(Z) = e^{-\sigma^2/2}e^{i\mu} \text{ so } \mu \text{ is the linear mean with } \mu = \tilde{\mu} + 2\pi K \mu \text{ with } \tilde{\mu} \in [0, 2\pi) \text{ the mean direction and } c = e^{-\sigma^2/2} \text{ is concentration} \]

\[ \theta \text{ is observed; } \theta + 2K\pi \text{ is the linear variable; } K \text{ is latent} \]

Joint density for \( \theta \) and \( K \) is

\[ f(\theta, k) = g(\theta + 2k\pi) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(\theta + 2k\pi - \mu)^2}{2\sigma^2}\right), \quad 0 \leq \theta < 2\pi \]

Standard dist theory for conditionals and marginals
Issues

- As $\sigma \to \infty$ WN tends to circular uniform. Mean direction
does not exist for uniform ($E\sin\theta = E\cos\theta = 0$) and $c = 0$

- Can truncate doubly infinite sum according to $\sigma$, e.g.,
  $\sigma < 2\pi/3$ implies $P(K \in \{-1, 0, 1\}) > .997$;
  $2\pi/3 < \sigma < 4\pi/3$ implies $P(K \in \{-2, -1, 0, 1, 2\}) > .997$

- So $K$ can be large iff $\sigma$ large, suggesting an
  identifiability issue

- If $\sigma$ large, hard to distinguish from a circular uniform

- “close to” uniform implies inference problems for $\mu$ and $c$

- Perhaps a preliminary uniformity test (Rayleigh;
  Kuiper-Watson; Rao - Circ Stats R library)
Bayesian model fitting

- Data $\theta_i, i = 1, 2, ..., n$
- Introduce latent $K_i$ to work with joint distribution of $(\theta_i, K_i)$
- Prior for $\mu$ and $\sigma^2$. We assume $\mu \sim N(\mu_0, \sigma_0^2)$. Induces a WN prior on $\tilde{\mu}$ with a latent $K_\mu$. Can not learn about $\mu$; can learn about $\tilde{\mu}$.
- For $\sigma^2$, from previous slide, a right truncated inverse Gamma
- Use moments estimators to help with priors and also to initiate MCMC chains
- Then, adaptively update each $K_i$ given $\mu$, $\sigma^2$ (using foregoing truncation), and $\theta_i$; update $\mu$ and $\sigma^2$ through conjugacy, given the $k_i$’s and $\theta_i$’s
Wrapped Gaussian Processes

- Recall the multivariate wrapped distribution:

\[
f(\theta) = \sum_{k_1=-\infty}^{\infty} \cdots \sum_{k_p=-\infty}^{\infty} g(\theta + 2\pi k)
\]

- Here \(\theta\) is observed vector, \(\theta + 2\pi K\) is the linear vector, \(K\) is the latent vector

- Again, the joint density for \((\theta, K)\) is:

\[
f(\theta, K) = g(\theta + 2\pi k)
\]

- 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

\[
\theta = (\theta(s_1), \theta(s_2), \ldots, \theta(s_n)) \sim WN(\mu 1, \sigma^2 R(\phi))
\]
Remarks

- We use a common mean, $\mu$, for all locations, hence a common $\tilde{\mu}$. A regression form would be for $\tilde{\mu}(s)$ and requires a suitable link function (as above for the von Mises distribution).

- Can directly extend to wrapped t-process using usual mixing of GP to t-process.

- We differ from Coles (1998) who uses a general fixed covariance matrix and requires replications. We have a structured "$\Sigma$" for any $n$ and any set of locations and do not require replications.

- We define the wrapped GP through the dependence structure of the linear GP. No unique notion of 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 $\theta_1, \theta_2$ indep

Jammalamadaka and Sarma (1988) provide the following measure which satisfies the above properties

$$\rho_c(\theta_1, \theta_2) = \frac{E(\sin(\theta_1 - \mu_1)\sin(\theta_2 - \mu_2))}{\sqrt{\text{Var}(\sin(\theta_1 - \mu_1))\text{Var}(\sin(\theta_2 - \mu_2))}}$$
For the Wrapped Normal

- For the WN, this measure takes a simple form

\[
\begin{pmatrix}
\theta_1 \\
\theta_2
\end{pmatrix} \sim WN\left(\begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}, \begin{pmatrix}
\sigma^2 & \rho \sigma^2 \\
\rho \sigma^2 & \sigma^2
\end{pmatrix}\right),
\]

\[
\rho_c(\theta_1, \theta_2) = \frac{\sinh(\rho \sigma^2)}{\sinh(\sigma^2)}
\]

- 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
Exponential spatial correlation $\exp\{-\phi D\}$ (dashed) and its circular equivalent (solid)
Fitting a wrapped GP model

- Approach follows that above. Introduce latent $K_i$'s. With a WN prior for $\mu$ and a right-censored inverse Gamma prior for $\sigma^2$
- What about $\phi$?
- With Matérn correlation function, we have difficulty identifying both $\sigma^2$ and $\phi$.
- However, we have been able to work with uniform priors for $\phi$, jointly updating $\phi$ and $\sigma^2$
- 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 $K$; only interested in those for $\mu, \sigma^2$
Kriging with Wrapped GP’s

- From linear GP, we have joint distribution of $\theta + 2\pi K, \theta(s_0) + 2\pi K(s_0)$, hence the the conditional distribution of $\theta(s_0) + 2\pi K(s_0)$ given $\theta + 2\pi K$, hence the WN model for $\theta(s_0)$ given $\theta + 2\pi K$

- Thus we have $E(e^{i\theta(s_0)}|\theta, K, \mu, \sigma^2, \phi) = e^{-\sigma_0^2} e^{i\mu_0}$ with $\mu_0, \sigma_0^2$ conditional mean and variance. But, we want $E(e^{i\theta(s_0)}|\theta)$

- Marginalizing over $K$ is hopeless. Even if each $K_i$ takes only three values, we have a sum over $3^n$ terms

- So, instead a Monte Carlo integration using samples of $K$’s and the parameters from fitting the model

- Yields a posterior mean kriged direction and a posterior kriged concentration (but no posterior samples from predictive distribution for $\theta(s_0)|\theta$)
Measurement error

- Particularly with directional data, can envision measurement error - wind monitors, buoy monitors
- Suppose we introduce conditionally independent measurement error as $\theta_{i,\text{obs}} \sim \text{WN}(\theta_{i,\text{true}}, \tau^2)$ with previous WN model for the $\theta_{i,\text{true}}$
- In the univariate case, cannot separate $\tau^2$ from $\sigma^2$.
- In the spatial case $\tau^2$ becomes a nugget and can be separated. Just another hierarchical layer to the modeling
- Kriging can be implemented as on previous slide.
Locations for the simulation examples; fitted (circles, 50 pts), validation (squares, 50 pts)
Simulation example; n=50 fitting, n=50 holdout; spatial (left, exponential cov fcn), nonspatial (right) 
\( \mu = \pi, \ \phi = 0.02 \) (range=140km, max dist=290km)

<table>
<thead>
<tr>
<th></th>
<th>n = 30</th>
<th>n = 70</th>
<th>n = 30</th>
<th>n = 70</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu} ) (95% CI)</td>
<td>3.124 (2.716, 3.517)</td>
<td>3.135 (2.500, 3.505)</td>
<td>3.102 (2.981, 3.223)</td>
<td>3.110 (3.034, 3.188)</td>
</tr>
<tr>
<td>( \hat{c} ) (95% CI)</td>
<td>0.926 (0.823, 0.965)</td>
<td>0.929 (0.548, 0.962)</td>
<td>0.947 (0.914, 0.968)</td>
<td>0.948 (0.928, 0.962)</td>
</tr>
<tr>
<td>( \hat{\phi} ) (95% CI)</td>
<td>0.013 (0.005, 0.066)</td>
<td>0.017 (0.007, 0.037)</td>
<td>0.023 (0.007, 0.032)</td>
<td>0.023 (0.007, 0.032)</td>
</tr>
<tr>
<td>Average Prediction Error</td>
<td>0.034</td>
<td>0.023</td>
<td>0.048</td>
<td>0.042</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>n = 30</th>
<th>n = 70</th>
<th>n = 30</th>
<th>n = 70</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu} ) (95% CI)</td>
<td>3.164 (2.399, 3.777)</td>
<td>3.229 (2.518, 3.898)</td>
<td>2.781 (2.535, 3.030)</td>
<td>2.925 (2.710, 3.140)</td>
</tr>
<tr>
<td>( \hat{c} ) (95% CI)</td>
<td>0.708 (0.547, 0.811)</td>
<td>0.748 (0.596, 0.841)</td>
<td>0.794 (0.680, 0.871)</td>
<td>0.749 (0.649, 0.823)</td>
</tr>
<tr>
<td>( \hat{\phi} ) (95% CI)</td>
<td>0.018 (0.009, 0.039)</td>
<td>0.015 (0.008, 0.032)</td>
<td>0.240 (0.064, 0.823)</td>
<td>0.170 (0.064, 0.823)</td>
</tr>
<tr>
<td>Average Prediction Error</td>
<td>0.058</td>
<td>0.085</td>
<td>0.240</td>
<td>0.170</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>n = 30</th>
<th>n = 70</th>
<th>n = 30</th>
<th>n = 70</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu} ) (95% CI)</td>
<td>2.916 (2.416, 3.514)</td>
<td>2.928 (2.261, 3.673)</td>
<td>2.869 (2.514, 3.217)</td>
<td>2.785 (2.577, 3.001)</td>
</tr>
<tr>
<td>( \hat{c} ) (95% CI)</td>
<td>0.594 (0.470, 0.693)</td>
<td>0.608 (0.480, 0.706)</td>
<td>0.640 (0.473, 0.765)</td>
<td>0.677 (0.578, 0.755)</td>
</tr>
<tr>
<td>( \hat{\phi} ) (95% CI)</td>
<td>0.049 (0.022, 0.190)</td>
<td>0.025 (0.015, 0.048)</td>
<td>0.335 (0.473, 0.765)</td>
<td>0.382 (0.578, 0.755)</td>
</tr>
<tr>
<td>Average Prediction Error</td>
<td>0.188</td>
<td>0.099</td>
<td>0.335</td>
<td>0.382</td>
</tr>
</tbody>
</table>
Italian wave direction data

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**)
WAM grid; squares denote buoy locations
The analysis

- Use 45 grid points
- $\hat{\mu} = 0.5540$, $\hat{\rho} = 0.8447$
- 30K MCMC iterations, 20K for burn-in
- “Downscaling” to a grid of 222 cells with 10km resolution
- Krige directions and concentrations
- A frequent definition of “circular distance” is $d(\theta_1, \theta_2) = 1 - \cos(\theta_1 - \theta_2)$. We use this to define prediction error as $1 - \cos(\hat{\theta} - \theta_{obs})$
- Average prediction error is 0.0243
- Average prediction error for nonspatial model is 0.1533
Figure 3. Prior (dashed) and posterior (solid) distributions for the WAM data
Figure 4. (a) Kriged estimate vs. observed using leave-one-out validation, (b) locations of the 5 outliers present in the data (grey squares)
Kriging results: posterior wave directions and concentrations
Projected normal

Back to projected normal

Recall, if $\mathbf{U} = (U_1, U_2) \sim g(u_1, u_2)$, a density on $\mathbb{R}^2$

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 $\theta = \arctan^* \frac{U_2}{U_1}$

Projected normal distribution. Suppose the random vector $\mathbf{U} \sim N_2(\mu, \Sigma)$, then $\theta \sim PN_2(\mu, \Sigma)$.

The density can be obtained explicitly but is very messy.

Instead, we would use polar coordinates working with the joint density of $(\theta, R)$ derived as a transformation from $(U_1, U_2)$, treating $R$ as a latent variable

$$f(r, \theta|\mu, \Sigma) = (2\pi)^{-1}|\Sigma|^{-\frac{1}{2}} \exp \left(-\frac{(r\mu-\mu)'\Sigma^{-1}(r\mu-\mu)}{2}\right) r$$
By transforming the bivariate random vector \( \mathbf{x} = (x_1, x_2)' \) into polar co-coordinates \((r, \theta)\) and integrating over \(r\) for a given \(\theta\), the density function for \(\theta\), \(f(\theta; \mu_1, \mu_2, \sigma_1, \sigma_2, \rho)\) is obtained as,

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

where

\[
a = \left\{\sigma_1 \sigma_2 \sqrt{1 - \rho^2}\right\}^{-1}
\]

\[
C(\theta) = a^2(\sigma_2^2 \cos^2 \theta - \rho \sigma_1 \sigma_2 \sin 2\theta + \sigma_1^2 \sin^2 \theta)
\]

\[
D(\theta) = a^2\{C(\theta)\}^{-\frac{1}{2}}\left\{\mu_1 \sigma_2(\sigma_2 \cos \theta - \rho \sigma_1 \sin \theta) + \mu_2 \sigma_1(\sigma_1 \sin \theta - \rho \sigma_2 \cos \theta)\right\}
\]

\[
\Sigma = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}.
\]
What do projected normal densities look like? The form with general $\Sigma$ 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 $\mu = \arctan^* \frac{\mu_1}{\mu_2}$, closed form for $\rho$ (Kendall, 1974).

In this case, the PN can be compared with the von Mises. Both have two parameters and we can line up their directions and resultants.

The advantage to the PN is convenient regression through, say $\mu_{1i} = X_i^T \beta_1$ and $\mu_{2i} = X_i^T \beta_2$. This induces a regression for the mean direction, i.e., $\mu(X_i)$. 
Figure: PN density vs von Mises
cont.

- We want to work with the more general $\Sigma$ case.
- We can draw pictures of the density in terms of five parameters in $\mu$ and $\Sigma$. We can achieve asymmetry and bimodality.
- However, with regard to inference, an identifiability issue: Note that if we scale $U$ by $a$, the distribution of $\theta$ 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 $\mu$ 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 $\theta$ for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = \sigma$ and different values of $\rho$
Figure 3. Shape of the general projected normal distribution
Model fitting and inference

- Bayesian model fitting is straightforward. With observed $\theta_i$'s and latent $R_i$'s, we convert to $U_{1i}$'s and $U_{2i}$'s. Update $\beta$'s and $\tau^2$ and $\rho$ under a standard bivariate Gaussian setup.

- Also, the $R_i$'s have an explicit closed form full conditional (M-H step with Gamma proposal)

- We simulate regression examples under the identified $\Sigma$ and can retrieve the parameters of the original model

- We show that, 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 $\theta$ 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 $\Sigma$ vs. $I$
Model comparison

<table>
<thead>
<tr>
<th>((n, \tau, \rho))</th>
<th>((60, 1, 0.4))</th>
<th>((200, 1, 0.4))</th>
<th>((60, 1.6, 0.4))</th>
<th>((200, 1.6, 0.4))</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLSL</td>
<td>39.16</td>
<td>54.60</td>
<td>37.64</td>
<td>42.30</td>
</tr>
<tr>
<td>average CRPS</td>
<td>0.2278</td>
<td>0.2452</td>
<td>0.2149</td>
<td>0.2251</td>
</tr>
<tr>
<td>average length of predictive CI</td>
<td>1.74</td>
<td>3.11</td>
<td>1.83</td>
<td>2.35</td>
</tr>
<tr>
<td>percentage (%)</td>
<td>100</td>
<td>100</td>
<td>97.5</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2. Model comparison: under each combination of \((n, \tau^2, \rho)\), general PN \(\Sigma \neq I\) (left column) and displaced normal \(\Sigma = I\) (right column)
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 \( U(s_i) \) 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.

Kriging is, again, straightforward. We can krig posterior predictive samples of say \( U(s_0) \) which, in turn induce posterior predictive samples of \( \theta(s_i) \).

We can easily insert spatial regressors, \( X(s) \) in the \( \mu(s) \), analogous to the independence case.
Model fitting

- From the joint distribution of \(\{U(s_i)\}\) we can write the joint distribution of \(\{\theta(s_i), r(s_i)\}\).

- So, only change from independence case is that we now need to update \(r(s_i)\)\ everything else. But same idea as before; now the conditional distribution of \(U(s_i)\)\ 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 \(U(s)\)

- From the separable cross covariance function, we can explore the induced covariance function for \(\theta(s)\)

- \(\rho\) is stationary in \(U(s)\) 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
Some joint distributions

C(s₁ - s₂) = 0, μ₁ = -0.32, μ₂ = 0, τ = 0.48, ρ = -0.62

C(s₁ - s₂) = 0, μ₁ = -1, μ₂ = 0, τ = 1, ρ = 0.4
Two joint distributions from separable correlation function with same cross 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 $\rho_c$ under projected normal process model. However, we can approximate this quantity using Monte Carlo integration.

Fisher's definition of circular correlation coefficient doesn't apply in the spatial setting.
∑≠1, \( \mu_1=\mu_2=1 \) (left), \( \mu_1=-1, \mu_2=1 \) (right)
Simulation: strong spatial dependence

$\mu_1 = -0.21, \mu_2 = 0.24, \tau = 0.92, \rho = -0.15, \phi = 1$
Simulation: weak spatial dependence

\[ \mu_1 = -0.21, \mu_2 = 0.24, \tau = 0.92, \rho = -0.15, \phi = 5 \]
Back to the wave direction data
Real Data Example

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

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

For the wrapped setting, since we have the associated wave heights, we would like a joint model. Hopefully, a joint model will provide “co-kriging” that will improve upon current spatial prediction. Challenge: a joint spatial model for a linear and an angular variable.

For the projected normal setting, we are comfortable with the spatial setting and we are adding dynamics in order to follow storms. We are also exploring the mixture setting. (There has been some work with mixtures of von Mises distributions.)