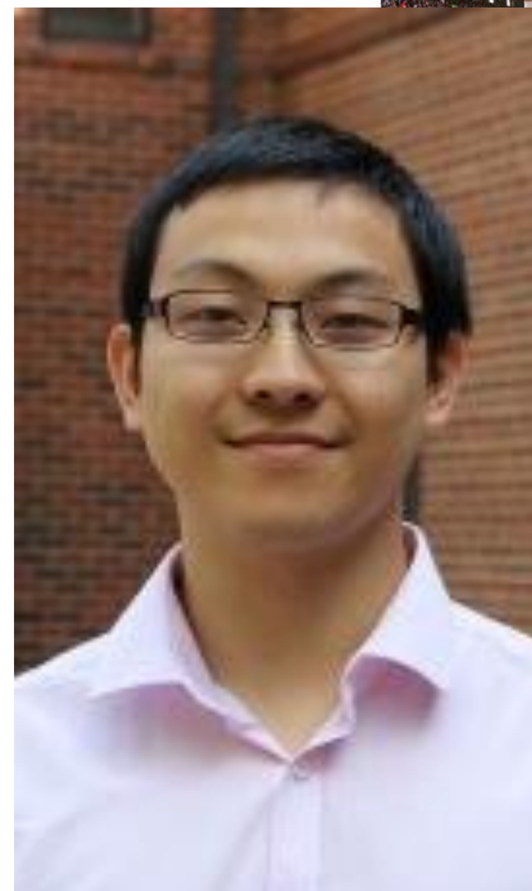
Modeling Space and Space-Time Directional Data Using Projected Gaussian Processes

- Presenter: Yingjun Guan
- Course: STAT545

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

About Me

- Yingjun Guan
- First-year doctoral student from School of Information Sciences (iSchool)
- Advisor: Vetle Torvik
- Research Interest: data mining, text mining, machine learning, and data visualization, specifically on how DM and TM can help analyze sociotechnical data such as medical data, publication data, or geographical data, etc; how to apply different ML algorithms to different projects and invent new ones; and how to better visualize and story tell the data.

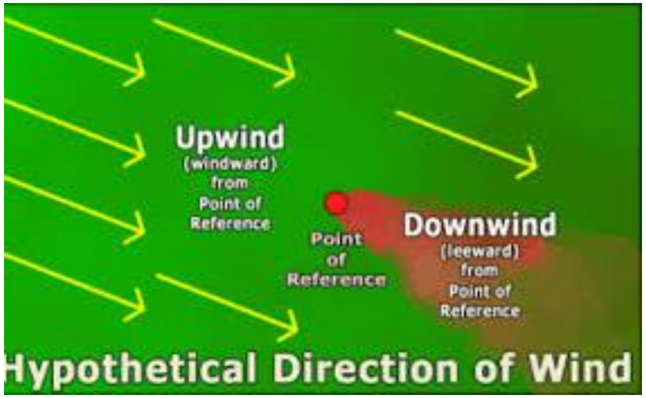


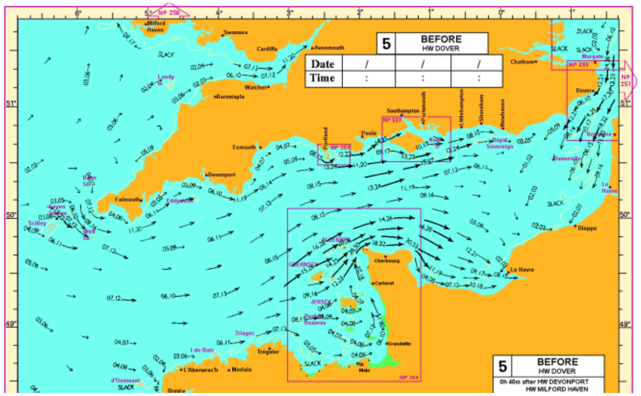
Content

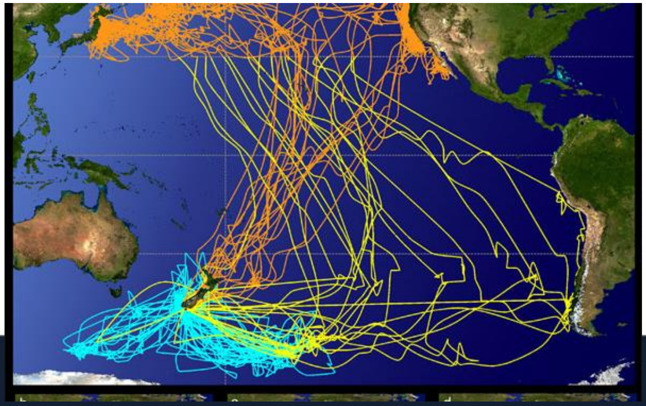
- 1. Introduction
- 2. Projected Gaussian Spatial Processes
- 3. Model fitting and inference
- 4. Examples
- 5. Space-Time models
- 6. Future work

1. Introduction

- Examples of maps on directional variables (meteorology, oceanography, biology, etc.)







1.1 Motivating example

- Figure 1 displays such data in a region of Adriatic sea at a particular time point during a calm period. The observations are wave angles at a fixed set of spatial locations over time.
- Furthermore, such data are obtained every 2 hr and we anticipate temporal dependence.
- The project also considers a two-day window during each of two sea motion states, calm and storm.

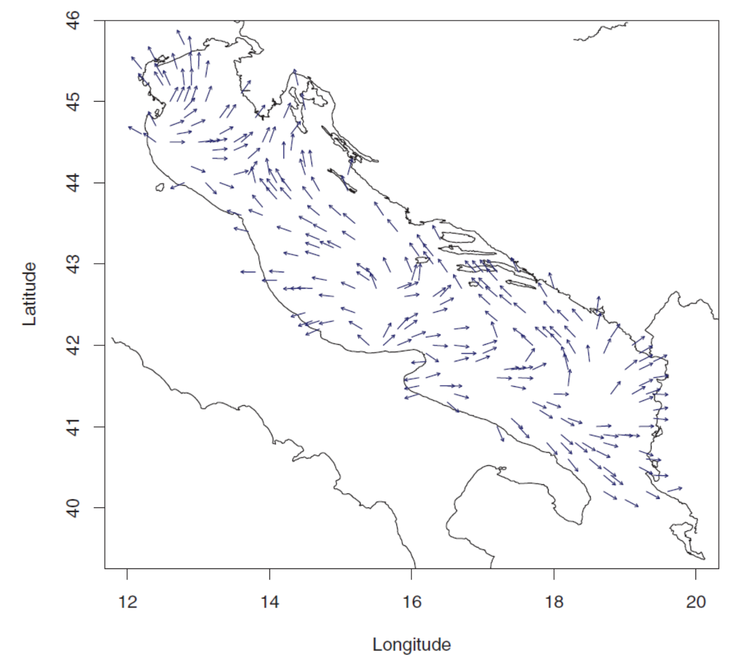


Figure 1. Outgoing wave directions at 06:00 on April 2, 2010, at Adriatic sea.

1.2 Measurement and data source

- Monitored ***buoy data***, supplying wave direction measurements, would seem to be attractive for analyzing wave directions.
- **However,** at present, buoy networks are too sparse to be used as a data source for spatial analysis.
- Therefore, we employ an alternative data source, outputs from **deterministic models (ISPRA model)**, usually climatic forecasts computed at several *spatial* and *temporal* resolutions.
- Deterministic models for the prediction of wave heights and directions are increasingly accurate and these models may eventually be calibrated when there is more buoy data.

Projected Gaussian Spatial Processes

- ## Univariate Projected Normal Distribution
- For univariate circular variables the von-Mises distribution (also known as the circular normal distribution) is most commonly used; it is **unimodal** and **symmetric**. The density takes the form,

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

- where μ is the mean direction, κ is a concentration parameter

Projected Gaussian Spatial Processes

- Projected Normal distribution
- Suppose a random vector $\mathbf{Y} = (Y1, \ldots, Yp)^T$ follows a p dimensional multivariate normal distribution, with mean $\boldsymbol{\mu}$ and covariance matrix ($p \geq 2$). The corresponding random unit vector $\mathbf{U} = \mathbf{Y}/\|\mathbf{Y}\|$ is said to follow a *projected normal* distribution (Small 1996; Mardia and Jupp 2000) with the same parameters.
- Mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$
- Covariance matrix

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

Projected Gaussian Spatial Processes

- Bivariate Spatial Process
- By projecting a bivariate spatial process on R2 we can create a spatial stochastic process of random variables taking values on a circle.
- Let $\mathbf{Y}(\mathbf{s}) = (Y1(\mathbf{s}), Y2(\mathbf{s}))^T$ denote a 2×1 vector of random variables at location $\mathbf{s}$, D be the domain of interest and $\{\mathbf{Y}(\mathbf{s}) : \mathbf{s} \in D\}$ be a bivariate stochastic process.
- Letting $(\cos(\mathbf{s}), \sin(\mathbf{s}))^T = (Y1(\mathbf{s}), Y2(\mathbf{s}))^T / ||\mathbf{Y}(\mathbf{s})||$, one obtains a circular process $(\mathbf{s})$. This projected process inherits properties of the inline bivariate process, such as stationarity.

Projected Gaussian Spatial Processes

- *Theorem.*

- If the bivariate process $\mathbf{Y}(\mathbf{s})$ is strictly stationary, the induced circular projected process $(\mathbf{s})$ is strictly stationary.

Model

- Suppose we have a projected Gaussian spatial process model, $\Theta(\mathbf{s})$, which is induced from a linear bivariate process $\mathbf{Y}(\mathbf{s})$ with mean $\boldsymbol{\mu}(\mathbf{s})$ and the separable cross-covariance $C(\mathbf{s}, \mathbf{s}') = \varrho(\mathbf{s} - \mathbf{s}'; \phi) \cdot T$. For simplicity, we only provide details of model fitting with constant mean $\boldsymbol{\mu}(\mathbf{s}) \equiv \boldsymbol{\mu} = (\mu_1, \mu_2)^{\mathrm{T}}$ and an exponential correlation function in the cross-covariance denoted by $\varrho(\mathbf{s} - \mathbf{s}'; \phi) = e^{-\phi\|\mathbf{s}-\mathbf{s}'\|}$. As previously defined,
- $$T = \begin{pmatrix} \tau^2 & \rho\tau \\ \rho\tau & 1 \end{pmatrix}.$$
- Therefore, we have five parameters: $\mu_1, \mu_2, \tau, \rho, \varphi$.

Projected Gaussian Spatial Processes

- The reason we import projected normal distribution is to solve asymmetric and multi-modality problems.

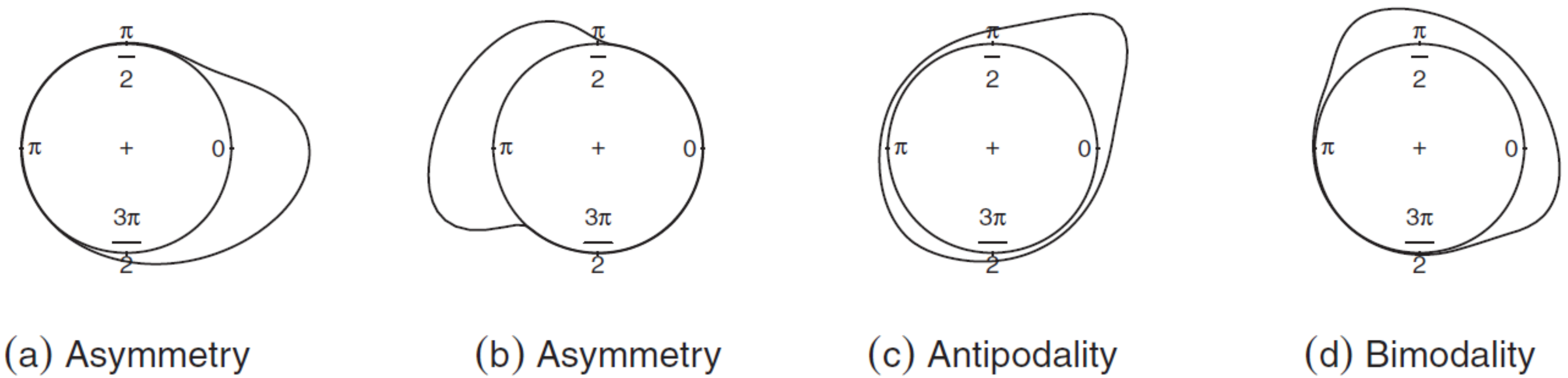
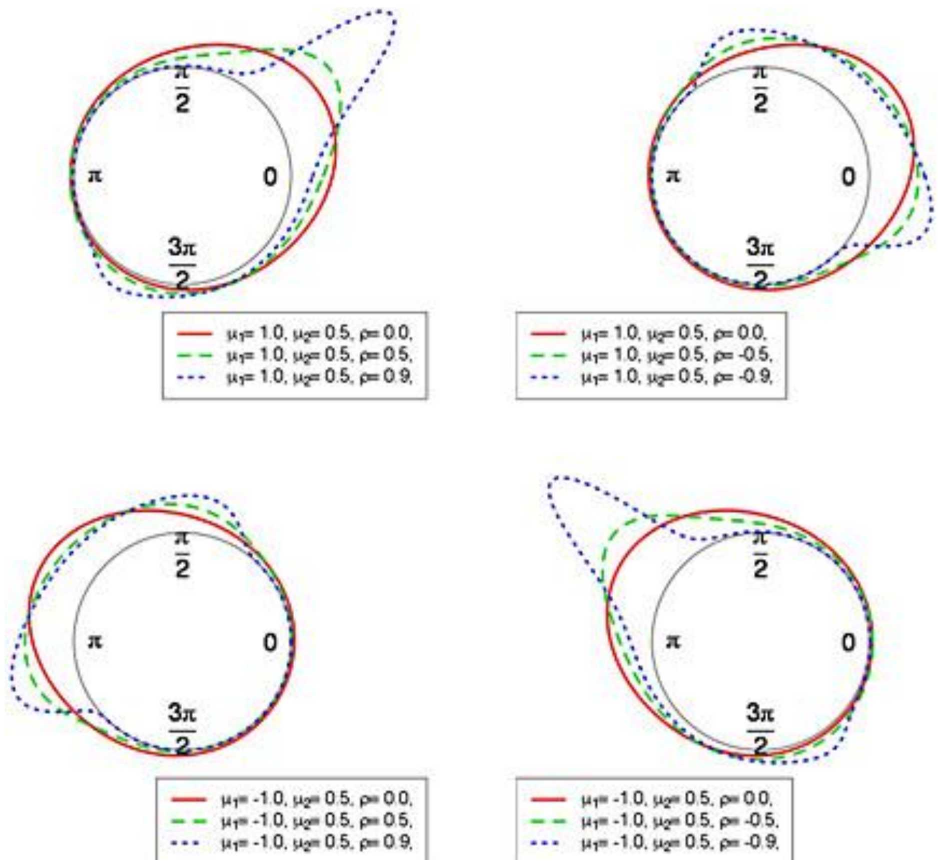


Figure 2. Density shapes for the general projected normal distribution.

Projected Gaussian Spatial Processes

- Five parameters ($\mu 1$, $\mu 2$, τ , ρ, φ) affect the shape of the density shape of each specific general projected normal distribution.
- (Hernandez-Stumpfhauser, 2017),
- (Wang, 2013)



Dependence under projected Gaussian Process

- Relationship between correlation and distance (and decay parameter).
- Exponential spatial correlation function is commonly used.
- This paper prefers PN correlation function.

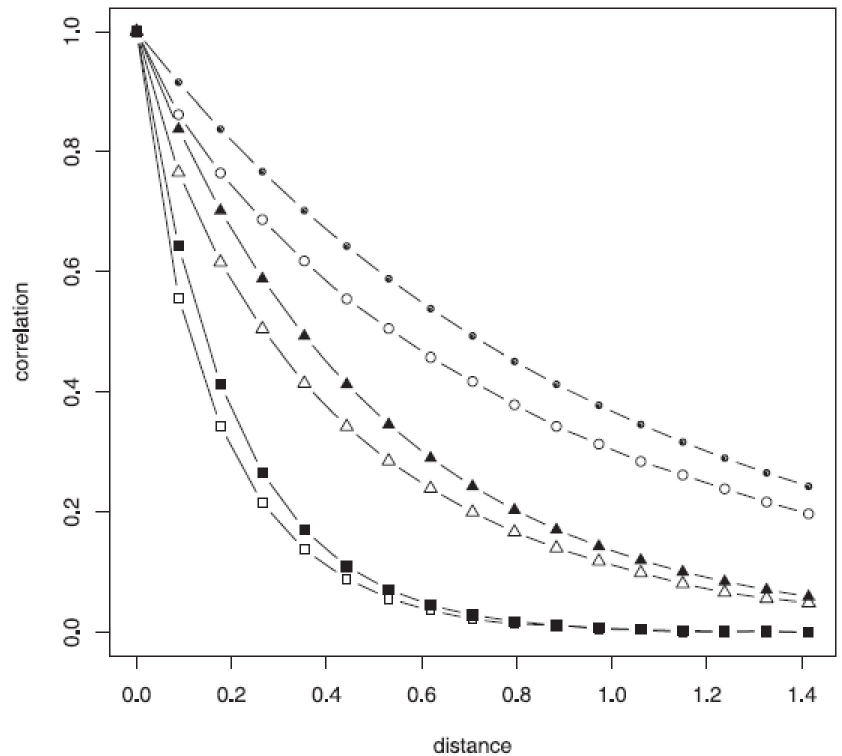


Figure 4. Exponential spatial correlation function (solid circle $\phi = 1$, solid triangle $\phi = 2$, solid square $\phi = 5$) and the corresponding PN correlation (empty circle $\phi = 1$, empty triangle $\phi = 2$, empty square $\phi = 5$) for 3 values of the decay parameter ϕ.

4. Real examples

- We generate samples of size 250 from a projected Gaussian process with constant mean $\boldsymbol{\mu}(\mathbf{s}) \equiv (\mu 1, \mu 2)^{\mathrm{T}}$ and separable cross-covariance $C(\mathbf{s}, \mathbf{s}) = (\mathbf{s} - \mathbf{s}; \varphi) \cdot T$, with $(\mathbf{s} - \mathbf{s}; \varphi)$ the exponential correlation function. Locations are generated uniformly on the unit square.
- We use different combinations of $(\mu 1, \mu 2, \tau^2, \rho, \varphi)$ to generate six simulated datasets; arrow plots of the simulated data for the different settings are shown in Figure 7, with the holdout observations marked in bold. (50 hold-out sites.)

- Figure 7. Display of simulated data of two different marginals (asymmetric a-c and bimodal d-f) and three different ranges of spatial dependence, long (a and d), medium (b and e), and short (c and f).
- Decay parameters and dependence distances are of inverse relation.
- Hold-out sites are in black.

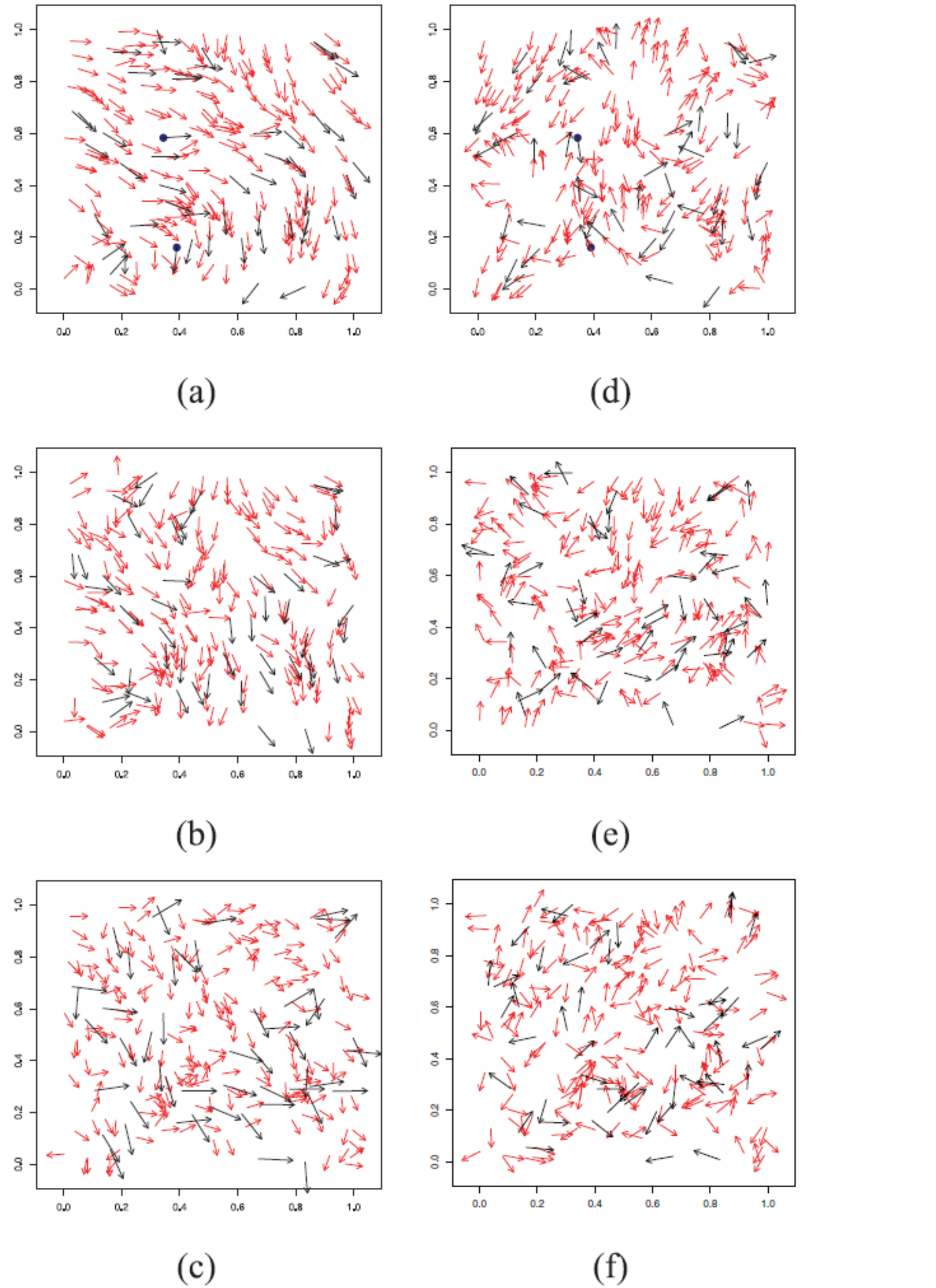


Figure 7. Display of simulated data of two different marginals (asymmetric (a)–(c) and bimodal (d)–(f)) and three different ranges of spatial dependence, long (a) and (d), medium (b) and (e), and short (c) and (f). Hold-out sites are in black.

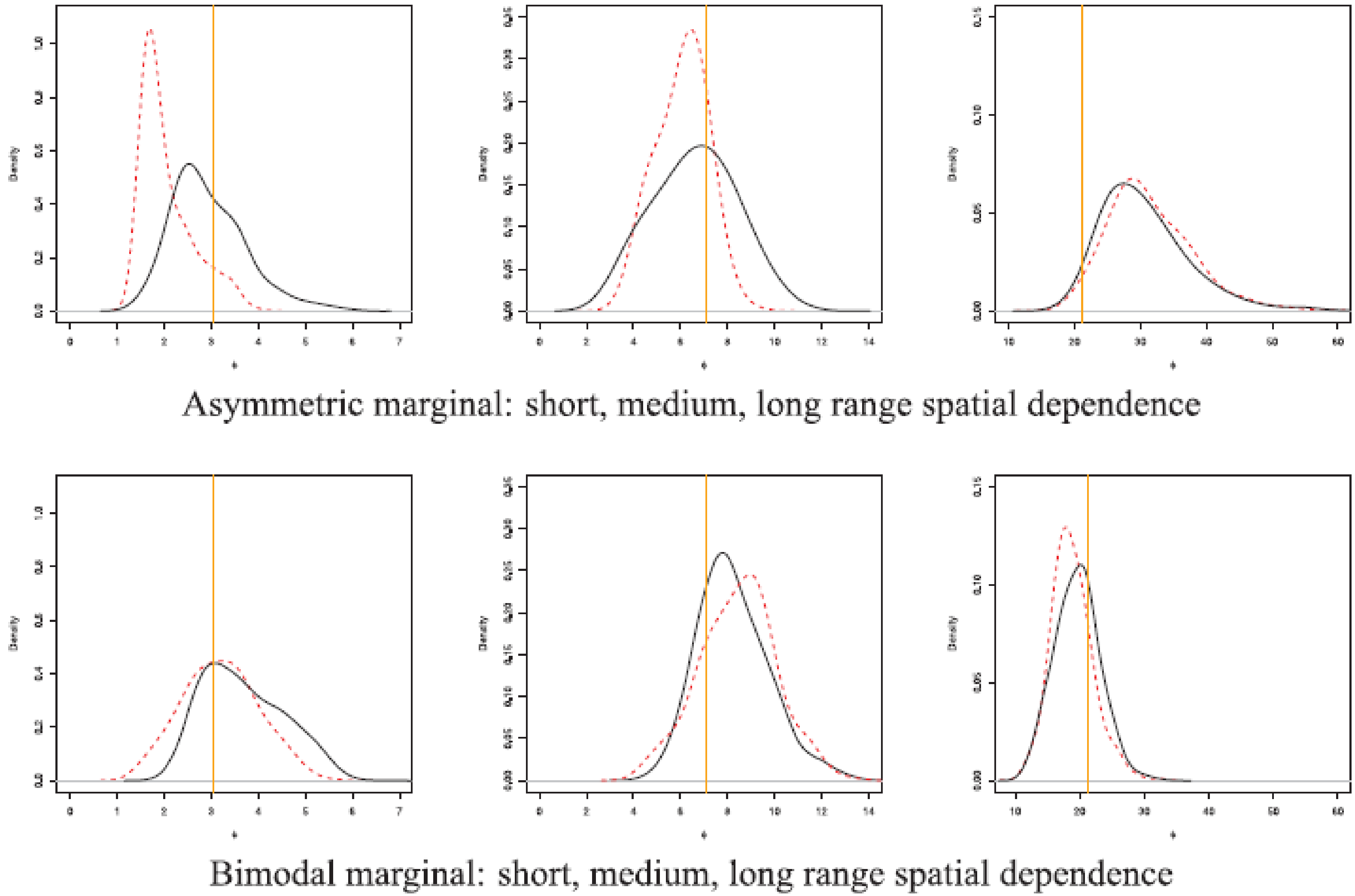
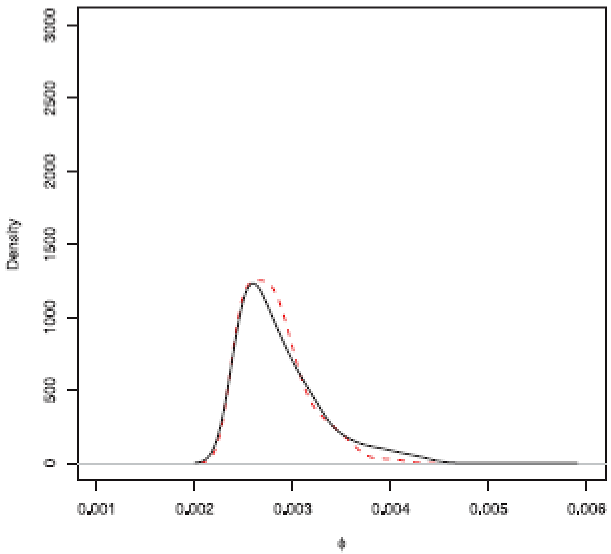
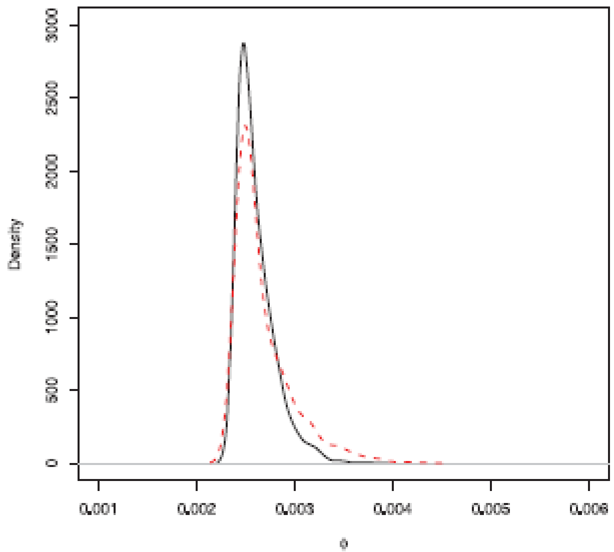


Figure 8. Posterior summaries of the decay parameter ϕ in all six scenarios under the projected Gaussian process model (solid line) and the reduced projected Gaussian process model $T = I$ (dashed line) for $n = 200$. True values are shown as vertical lines.

Real example analysis



(a) calm



(b) storm

Figure 10. Posterior summaries of the decay parameter ϕ under the projected Gaussian process model (solid line) and the reduced projected Gaussian process model $T = I$ (dashed line) for two data examples.

Summary and future work

- This paper provides a flexible class of specifications for modeling space and space-time dependence between angular measurements.
- The projected bivariate Gaussian process performs better than the commonly used exponential model.
- In future, the authors plan to enriching the models through mixture distributions. For example, how to combine the linear variable (height of waves) and the circular variable (direction)?

- Thanks for your time.