Some Mixture Models for Joint Analysis of Wind Speed and Wind Direction

Debarghya Jana[1] and Arnab Hazra[2]

Iowa State University[1]

Indian Institute of Technology Kanpur[2]

IISA conference, Nebraska

June 14, 2025

Outline

Table of Contents

Motivation

- Our primary motivation of this work came from the shocking incident of cyclone Amphan in 2020 which was the strongest Bay of Bengal storm in 21 years, made the North Indian Ocean cyclone season the costliest ever, causing massive infrastructural damage and resulting in at least 19 fatalities in Kolkata *(Economic Times, 2020)*.
- In fact, the North Indian Ocean annually generates about 5.5 tropical cyclones, mostly in the Bay of Bengal and contributing roughly 7% to the global total.
- So, effective disaster preparedness and response measures are crucial in these regions to mitigate the impact of these cyclones and protect the vulnerable populations living along the coastlines.

Why Joint Modelling?

- By jointly modelling wind speed and wind direction in this context, we can effectively capture the underlying dependency structure between wind speed and wind direction, anisotropic behavior, and rotational patterns present in cyclone systems.
- Furthermore, joint models of wind speed and wind direction for cyclones are essential for simulating wind field vector over the spatial domains.
- The displacement of the cyclone's eye (center) can be inferred which enables cyclone path tracking.

Gap in existing Literatures

- We found two widely used parametric models Isotropic Gaussian (IG) *[McWilliams and Sprevak, 1980]* and Anisotropic Gaussian (AG) *[Weber, 1991]* models and some popular nonparametric models i.e, nonparametric Kernel-density model (NP-KD) *(Han et al.,2018)* and nonparametric Johnson-wherely (NP-JW) *(Johnson and Wehrly, 1978)* model where joint modelling of wind speed and wind direction had been done already.
- But, There is a lack of semiparametric approach in joint modelling for wind speed and direction in existing literatures.

But why semiparametric approach?

- Most existing approaches are either fully parametric, which impose rigid structural assumptions, or they are nonparametric, which are lacking in interpretability and efficiency.
- But, joint models in a semiparametric framework restricts flexibility in capturing complex, non-linear dependencies while retaining interpretability and structure offered by parametric components.
- So, we explored semiparametric approach for joint modelling of wind speed and wind direction and proposed two new models:
    - **Isotropic Gaussian Mixture (IGM) model**
    - **Anisotropic Gaussian Mixture (AGM) model**

Table of Contents

Data and Exploratory analysis

- Source: International Best Track Archive for Climate Stewardship (IBTrACS, NCDC).
- Time Period: 1980–2022.
- Bay of Bengal has nearly four times cyclone frequency compared to the Arabian Sea.

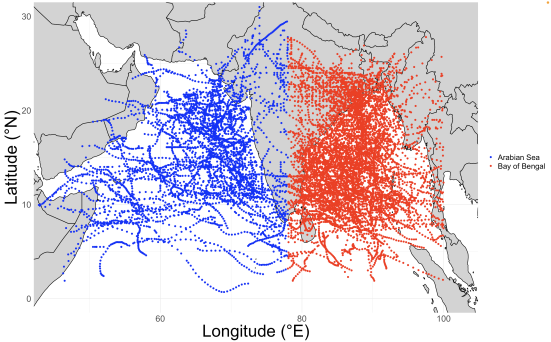

Figure: Cyclones in the North Indian Ocean basin between 1980–2022.

Data and Exploratory analysis

- Increase in NIO cyclone activity in post-1990.

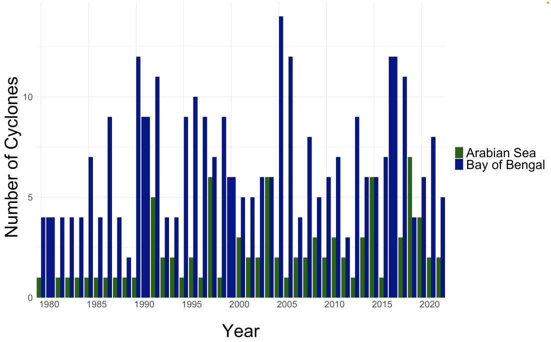

Figure: Year-wise cyclone occures in North Indian Ocean basin

Data and Exploratory analysis

- We observed a clustered point pattern that has mainly two clusters surrounding the regions of Arabian Sea and Bay of Bengal sub-basins.
- The kernel density within the Bay of Bengal basin is higher than that in the Arabian Sea on average.

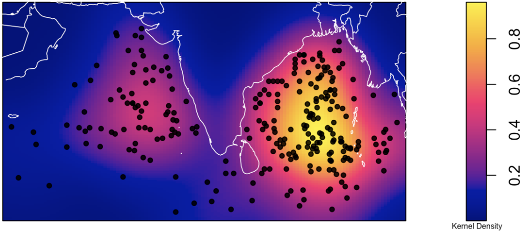

Figure: Heat map of Kernel density in NIO basin

Table of Contents

Methodology

- We proposed two semiparametric models, Isotropic Gaussian mixture (AGM) and Anisotropic Gaussian mixture(IGM) model which were developed by using the parametric models Isotropic Gaussian model *[McWilliams and Sprevak, 1980]* and Anisotropic Gaussian model *[Weber, 1991]* respectively.
- **Isotropic Gaussian model (IG)** *[McWilliams and Sprevak, 1980]*: The joint pdf of wind speed (V') and wind direction (Θ'') can be expressed as,

$$f_{V',\Theta''}(v',\theta'') = \frac{\exp(2v')}{2\pi\sigma_{y'}^2}\exp\left(-\frac{\mu_{y'}^2 + \exp(2v') - 2\mu_{y'}\exp(v')\cos(\theta''+\pi) - v'}{2\sigma_{y'}^2}\right)$$

for $v' \in \mathbb{R}$ and $-\pi < \theta'' < \pi$ and μ_y' is the mean and ${\sigma_y'}^2$ is the variance of the longitudinal component.

Methodology

- **Isotropic Gaussian Mixture (IGM) model:** We proposed the joint pdf of wind speed (V') and wind direction (Θ'') as

$$\mathrm{f}_{V',\Theta''}(v',\theta') = \sum_{j=1}^{C} \phi_j f^{(j)}_{V',\Theta''}\left(v', \theta'' \mid \mu^{(j)}_{y'}, \sigma^{(j)}_{y'}\right), v' \in \mathbb{R}, -\pi < \theta'' < \pi$$

  Here, $\phi_{i=1,2,\ldots,C}$ is the mixture weight and $f^{(j)}_{V',\Theta''}\left(v', \theta'' \mid \mu^{(j)}_{y'}, \sigma^{(j)}_{y'}\right)$ denotes the joint pdf of wind speed and wind direction in Isotropic Gaussian (IG) model for the mixture $\{j\}_{\{1,2,\ldots,C\}}$.
- Now, we have used the EM algorithm to estimate the set of parameters $\boldsymbol{\theta} = \left(\phi_1, \phi_2, \ldots, \phi_C, \mu^{(1)}_{y'}, \mu^{(2)}_{y'}, \ldots, \mu^{(C)}_{y'}, \sigma^{(1)}_{y'}, \sigma^{(2)}_{y'}, \ldots, \sigma^{(C)}_{y'}\right)$ in this proposed model.

Methodology

Algorithm 1 EM Algorithm for Mixture of IG model

[1] $\boldsymbol{\theta}_{(0)} = \left(\phi_{1,(0)}, \ldots, \phi_{C,(0)}, \mu_{y'}^{(1,(0))}, \ldots, \mu_{y'}^{(C,(0))}, \sigma_{y'}^{(1,(0))}, \ldots, \sigma_{y'}^{(C,(0))}\right)$ [Using K-means clustering method]

[2]**E-step:** For all $c = 1, 2 \ldots, C$ and all $i = 1, \ldots, n$, calculate

$$\gamma_{i,c,k} = \frac{f_{V',\Theta''}^{(c)}\left(v', \theta'' \mid \mu_{y'}^{(c,k)}, \sigma_{y'}^{(c,k)}\right)\phi_{c,k}}{\sum_{j=1}^{C} f_{V',\Theta''}^{(j)}\left(v', \theta'' \mid \mu_{y'}^{(j,k)}, \sigma_{y'}^{(j,k)}\right)\phi_{j,k}},.$$

and

$$q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{(\boldsymbol{k})}\right) = \sum_{i=1}^{n}\sum_{c=1}^{C} \log\left\{f_{V',\Theta''}^{(c)}\left(v', \theta'' \mid \mu_{y'}^{(c)}, \sigma_{y'}^{(c)}\right)\phi_c\right\}\gamma_{i,c,k}$$

[3] **M-step:**

$$\boldsymbol{\theta}_{(\boldsymbol{k}+\mathbf{1})} = \arg\max q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}_{(\boldsymbol{k})}\right)$$

[4] Stop when $\|\boldsymbol{\theta}_{(\boldsymbol{k}+\mathbf{1})} - \boldsymbol{\theta}_{(\boldsymbol{k})}\| < \epsilon$.

Methodology

Anisotropic Gaussian model (AG) *[Weber, 1991]:* The joint pdf of wind speed (V') and wind direction (Θ'') can be expressed as

$$f_{V',\Theta''}\left(v',\theta''\right) = \frac{\exp(2v')}{2\pi\sigma_{y'}\sigma_{x'}} \exp\left[-\frac{\left(\exp(v)\cos(\theta'') - \mu_{y'}\right)^2}{2\sigma_{y'}^2} - \frac{v^2\sin^2(\theta''+\pi)}{2\sigma_{x'}^2}\right]$$

for $v' \in \mathbb{R}, -\pi < \theta'' < 2\pi$ where, ${\sigma'_x}^2$ is the variance of the lateral component, μ'_y is the mean and ${\sigma'_y}^2$ is the variance of the longitudinal component.

Methodology

- **Anisotropic Gaussian Mixture(AGM) model:** Here, We proposed the joint pdf of wind speed (V') and wind direction (Θ'') as

$$f_{V',\Theta''}\left(v',\theta'\right)=\sum_{j=1}^{C}\phi_j f_{V',\Theta''}^{(j)}\left(v',\theta''\mid \mu_{x'}^{(j)},\sigma_{x'}^{(j)},\sigma_{y'}^{(j)}\right),v'\in\mathbb{R},-\pi<\theta''<\pi$$

  Here, $\phi_{i=1,2,\ldots,C}$ is the mixture weight and $f_{V',\Theta''}^{(j)}\left(v',\theta''\mid \mu_{x'}^{(j)},\sigma_{x'}^{(j)},\sigma_{y'}^{(j)}\right)$ denotes the joint pdf of wind speed and wind direction in Anisotropic Gaussian(AG) model for the mixture $\{j\}_{\{1,2,\ldots,C\}}$.

- Now, we have used the EM algorithm to estimate the set of parameters $\boldsymbol{\theta}=\left(\phi_1,\ldots,\phi_C,\mu_{x'}^{(1)},\ldots,\mu_{x'}^{(C)},\sigma_{x'}^{(1)},\ldots,\sigma_{x'}^{(C)},\sigma_{y'}^{(1)},\ldots,\sigma_{y'}^{(C)}\right)$ in this proposed model.

Methodology

Algorithm 2 EM Algorithm for Mixture of AG model

[1] $\boldsymbol{\theta_{(0)}} = \left(\phi_{1,(0)}, \ldots, \phi_{C,(0)}, \mu_{y'}^{(1,(0))}, \ldots, \mu_{y'}^{(C,(0))}, \sigma_{x'}^{(1,(0))}, \ldots, \sigma_{x'}^{(C,(0))} \sigma_{y'}^{(1,(0))}, \ldots, \sigma_{y'}^{(C,(0))}\right)$
[Using K-means clustering method]

[2] **E-step:** For all $c = 1, 2$ and all $i = 1, \ldots, n$, calculate

$$\gamma_{i,c,k} = \frac{f_{V',\Theta''}^{(c)}\left(v_i', \theta_i'' \mid \mu_{y'}^{(c,k)}, \sigma_{x'}^{(c,k)} \sigma_{y'}^{(c,k)}\right) \phi_{c,k}}{\sum_{j=1}^{C} f_{V',\Theta''}^{(j)}\left(v_i', \theta_i'' \mid \mu_{y'}^{(j,k)}, \sigma_{x'}^{(j,k)}, \sigma_{y'}^{(j,k)}\right) \phi_{j,k}}$$

and

$$q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta_{(k)}}\right) = \sum_{i=1}^{n} \sum_{c=1}^{C} \log\left\{f_{V',\Theta''}^{(c)}\left(v_i', \theta_i'' \mid \mu_{y'}^{(c)}, \sigma_{x'}^{(c)} \sigma_{y'}^{(c)}\right) \phi_c\right\} \gamma_{i,c,k}$$

[3] **M-step:**

$$\boldsymbol{\theta_{(k+1)}} = \arg\max q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta_{(k)}}\right)$$

[4] Stop when $\|\boldsymbol{\theta_{(k+1)}} - \boldsymbol{\theta_{(k)}}\| < \epsilon$.

Table of Contents

Results

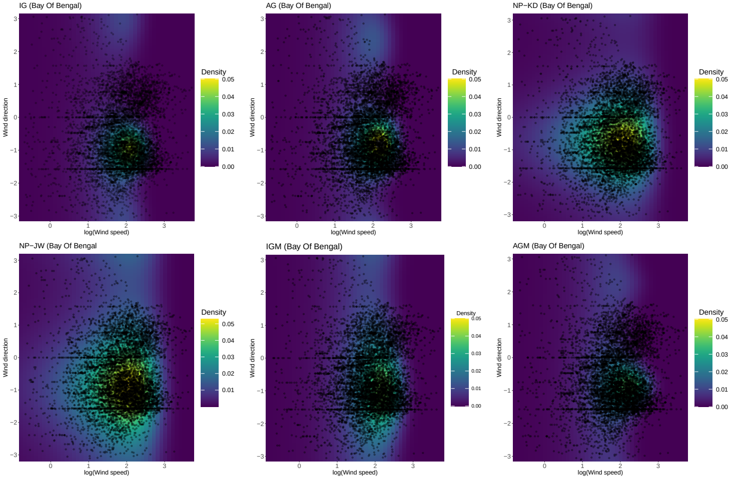

Figure: Heat maps of joint modelling of wind speed and wind direction of different models for Bay Of Bengal.

Results

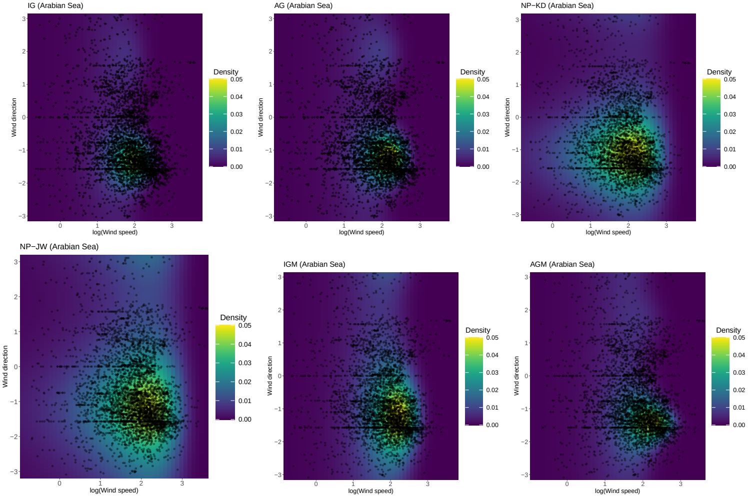

Figure: Heat maps of joint modelling of wind speed and wind direction of different models for the Arabian Sea.

Results

- From the heatmaps, it is clear that IG and AG models do not seem to capture the joint behaviour of wind speed and wind direction adequately.
- Nonparametric models NP-KD, NP-JW fit the data better but they exhibited indications of overfitting.
- But our proposed models IGM and AGM performed better than IG and AG models, respectively.

Results

Goodness of Fit (Bay of Bengal)						
	IG	AG	NP-KD	NP-JW	IGM	AGM
RMSE	1.76829e-02	1.8639e-02	8.05386e-03	8.1648e-03	1.02568e-02	1.24185e-02
MAE	5.34500e-03	6.3328e-03	4.08876e-03	3.78397e-03	3.25786e-03	4.28794e-03
IA	2.49659e-01	3.58482e-01	8.65130e-01	8.86474e-01	4.52784e-01	5.73958e-01
χ^2	5.85657e+02	4.44873e+02	2.66578e+03	3.84669e+03	3.63687e+02	2.61529e+02

Goodness of Fit (Arabian sea)						
	IG	AG	NP-KD	NP-JW	IGM	AGM
RMSE	1.74823e-02	2.47760e-02	8.80077e-03	8.85967e-03	1.30803e-02	1.5696e-02
MAE	6.02321e-03	7.01313e-03	4.06935e-03	3.99924e-03	5.02041e-03	5.68204e-03
IA	4.10525e-01	5.27164e-01	7.92153e-01	8.75362e-01	5.83937e-01	6.77624e-01
χ^2	6.43673e+03	3.12971e+03	3.61440e+03	4.40314e+03	4.52839e+03	2.02516e+03

Results

- Both IG and AG consistently show the highest RMSE and MAE, and lowest IA, indicating poor model fit.
- Both NP-KD and NP-JW have extremely low RMSE, which gives the strong indication of overfitting even though very high IA values (close to 1), indicating apparently excellent fit.
- However, the extremely high χ^2 values [over 4000 for NP-JW in Arabian sea] suggest overfitting again.

Results

- For IGM and AGM model, we get low RMSE and MAE (slightly higher than nonparametric models but much lower than parametric ones).
- IGM and AGM models have higher IA (though slightly lower than NP models) values than parametric models.
- So, our semiparametric models (IGM and AGM) outperform parametric models and are less prone to overfitting than nonparametric models. Among them, AGM consistently offers the most balanced performance across all metrics in both sub-basins.

Table of Contents

Discussion and Future Study

- Our main goal was to shed light on the semiparametric approach in joint modelling of wind speed and wind direction as it was not an explored area and most of the work was either in parametric or nonparametric setup and also to provide a reliable model which can be used for cyclone data.

- While we have evaluated the in-sample fitting RMSE, we have not yet assessed the out-of-sample prediction RMSE. As expected, nonparametric models yield lower fitting RMSE due to their flexibility, but this does not necessarily reflect better predictive performance.

- We will also focus on finding Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) to give a better justification of the overfitting of those nonparametric models.

References

[1] Economic Times (2020), EconomicE. Cyclone Amphan News: 72 people dead in West Bengal, PM Modi to conduct aerial survey Friday. The Economic Times.

[2] Han, QinkaiQ., Hao, ZhuolinZ., Hu, TaoT. Chu, FuleiF. (2018). nonparametric models for joint probabilistic distributions of wind speed and direction data. Renewable Energy 126 1032–1042.

[3] Johnson, Richard AR. A. Wehrly, Thomas ET. E. (1978). Some angular-linear distributions and related regression models. Journal of the American Statistical Association 73 602–606.

[4] B McWilliams and D Sprevak. The estimation of the parameters of the distribution of wind speed and direction. Wind Engineering, pages 227–238, 1980.

[5]Weber, RudolfR. (1991). Estimator for the standard deviation of wind direction based on moments of the Cartesian components. Journal of Applied Meteorology and Climatology 30 1341–1353.

Please email me to djana@iastate.edu if you have any questions, suggestions or comments!

Thank you!