Some mixture models for joint analysis of wind speed and wind direction for North Indian Ocean Cyclones

Debarghya Jana

MTH598A Presentation

Supervisor:

Dr. Arnab Hazra

Department Of Mathematics and Statistics

Indian Institute Of Technology, Kanpur

Date: November 11, 2023

Contents

Introduction

- The primary motivation for this work came from the shocking incident in 2020: In 2020, Cyclone Amphan made the North Indian Ocean cyclone season the costliest ever and Amphan was the strongest Bay of Bengal storm in 21 years and caused significant destruction including overturned vehicles, snapped 10,000 trees, and knocked down 4,000 electric poles in Kolkata, resulting in a 14-hour power outage and at least 19 fatalities.(Economic Times (2020),
- Indeed, not only Amphan,2020; the North Indian Ocean basin experiences a significant number of cyclones each year, particularly during the cyclone season, which typically lasts from April to November and affects countries bordering the Bay of Bengal and the Arabian Sea, including India, Bangladesh, Myanmar, Sri Lanka, Pakistan, and others.
- On an annual average, the North Indian Ocean (NIO) experiences the genesis of approximately 5.5 tropical cyclones (TCs), with four of these typically occurring in the Bay of Bengal and one and a half in the Arabian Sea. This accounts for approximately 7% of the global yearly total of TCs. (Rajasekhar et al., 2014)
- 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.

Introduction(Continued)

- In this context, joint modelling for both wind speed and wind direction in cyclones play a pivotal role in fortifying coastal areas against the potentially devastating impacts of these formidable weather systems.
- Joint modelling is important in precise hazard assessment and by considering both wind speed and direction, authorities and emergency managers can identify the specific coastal zones that are most vulnerable to cyclone-related damage.
- In storm surge prediction, joint modelling of wind speed and wind direction plays a major role in generating storm surges, which can result in coastal flooding.
- Joint modelling informs the development of mitigation strategies such as vegetation management, coastal dune restoration, and building codes that take into account the specific wind patterns associated with cyclones. These are the rough varieties of reasons why joint modelling of wind speed and wind direction is so important.

Introduction(Continued)

- This was our main motivation to propose useful statistical models which will be used for joint modelling of wind speed and wind direction of cyclones.
- There are very few statistical models where joint modelling of wind speed and wind direction has been done in **parametric** and **nonparametric** set up but there was no **semiparametric** approach for joint modelling of wind speed and wind speed. So, we explored the **semiparametric method**.
- We proposed two mixture models named **Isotropic Gaussian Mixture (IGM)** model and **Anisotropic Gaussian Mixture(AGM)** model for joint modelling of wind speed and wind direction and we applied our proposed models for joint modelling of wind speed and wind directions of cyclones in the North Indian Ocean Basin(NIO).

Background

- We have imposed an one-one transformation $f : (v, \theta') \to (\log(v), \theta' - \pi)$.Let us denote, $\log(v) = v'$ and $\theta' - \pi = \theta''$ in all of the models described below except the NP-JW model.
- In NP-JW model, we have imposed an one-one transformation $f(v, \theta') \to (v, \theta' - \pi)$ and we denote, $v = v'$ and $\theta' - \pi = \theta''$.
- We used two most popular parametric models and two nonparametric models for comparison purposes. Here is a brief discussion about those models.
- **A)Isotropic Gaussian(IG) model:**The joint pdf of V' and Θ'' will be,

$$f_{V',\Theta''}(v', \theta') = \frac{\exp(2v')}{2\pi\sigma_{y'}^2} \exp\left[-\frac{\mu_{y'}^2}{2\sigma_{y'}^2}\right] \exp\left[-\frac{\exp(2v') - 2\mu_{y'}\exp(v')\cos(\theta'' + \pi)}{2\sigma_{y'}^2}\right] \tag{1}$$

where, $-\infty < v' < \infty$ and $-\pi < \theta'' < \pi$ (McWilliams and Sprevak, 1980)

- **B)Anisotropic Gaussian(AG) model:**The joint pdf of V' and Θ'' will be

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

where, $-\infty < v' < \infty$ and $-\pi < \theta'' < 2\pi$.(Weber, 1991)

Background(Continued)

- **C)Non-Parametric Kernel Density(NP-KD) model:** The estimated joint PDF of wind speed and wind direction, based on the NP-KD model (Han et al.,2018) as described in will be

$$\hat{f}_{V',\Theta''}(v',\theta'') = \frac{\exp\left(v' + \hat{\kappa}\cos\left(\theta'' - \theta_i''\right)\right)}{(2\pi)^{3/2} I_0(\hat{\kappa})\hat{\sigma}} \sum_{i=1}^{n} \exp\left(-\frac{\left(\exp(v') - \exp(v_i')\right)^2}{2\hat{\sigma}^2}\right) \tag{3}$$

- $\hat{\sigma}$ and $\hat{\kappa}$ are the optimal bandwidth of σ and κ which we got by maximizing the likelihood function by Likelihood cross-validation(LCV) method,

$$LCV(\sigma,\kappa) = \prod_{i=1}^{n} f_{-i,V,\Theta}\left(v_i,\theta_i,\sigma,\kappa\right) \tag{4}$$

where, $f_{-i,V,\Theta}$ represents the joint density after deleting the ith observation value.

Background(Continued)

- **D)Non-Parametric Johson-wehrly(NP-JW) model:** The estimated joint PDF of wind speed and wind direction, based on the NP-JW model (Han et al.,2018),(Johnson and Wehrly, 1978) as described will be

$$\hat{f}_{V',\Theta''}(v',\theta'') = \frac{1}{(2\pi)^{5/2}(I_0(\hat{\kappa}))^2\sigma}\sum_{i=1}^{n}\exp\left(-\frac{(v-v_i)^2}{2\hat{\sigma}^2}\right)\exp\left(\hat{\kappa}(\cos\left(\theta''-\theta_i''\right)+\cos\left(\zeta''-\zeta_i''\right)\right) \quad (5)$$

  where, $-\pi < \theta < \pi, -\infty < v < \infty$.

- $\hat{\sigma}$ and $\hat{\kappa}$ are the optimal bandwidth of σ and κ which we got by maximizing the likelihood function by Likelihood cross-validation(LCV) method,

$$LCV(\sigma,\kappa) = \prod_{i=1}^{n} f_{-i,V,\Theta}(v_i,\theta_i,\sigma,\kappa) \quad (6)$$

  where, $f_{-i,V,\Theta}$ represents the joint density after deleting the ith observation value.

- ζ represents the circular variable between the wind speed and direction, defined as

$$\zeta = \begin{cases} 2\pi\left[F_V(v) - F_{\Theta'}(\theta')\right], & F_V(v) \geq F_{\Theta'}(\theta') \\ 2\pi\left[F_V(v) - F_{\Theta'}(\theta')\right] + 2\pi, & F_V(v) < F_{\Theta'}(\theta'), \end{cases} \quad (7)$$

  where $F_V(v)$ and $F_{\Theta'}(\theta')$ are the distribution functions of wind speed and direction, respectively, and $g(\zeta)$ represents the PDF of ζ.

Methodology

- Initially, there were the Anisotropic Gaussian model and Isotropic Gaussian model which were used to model wind speed and wind direction jointly, but now we proposed the Anisotropic Gaussian mixture(AGM) model and Isotropic Gaussian mixture(IGM) model using Expectation-Maximization(EM) algorithm.
- **Isotropic Gaussian Mixture(IGM) model:** In this model, the joint pdf of wind speed and wind direction will be

$$f_{V',\Theta''}\left(v',\theta'\right)=\sum_{j=1}^{C}\phi_j f_{V',\Theta''}^{(j)}\left(v',\theta''\mid \mu_{y'}^{(j)},(\sigma_{y'}^{(j)})^2\right) \tag{8}$$

where, $-\infty < v' < \infty$ and $-\pi < \theta'' < \pi$.

- Here, $\phi_{i=1,2,\ldots,C}$ is the mixture weight and $f_{V',\Theta''}^{(j)}\left(v',\theta''\mid \mu_{y'}^{(j)},\sigma_{y'}^{(j)}\right)$ denotes the joint pdf of wind speed and wind direction in Isotropic Gaussian(IG) model in 8 respectively for the mixture $\{j\}_{\{1,2,\ldots,C\}}$
- Now, we have used the EM algorithm to estimate the set of parameters $\boldsymbol{\theta}=\left(\phi_1,\phi_2,\ldots,\phi_C,\mu_{y'}^{(1)},\mu_{y'}^{(2)},\ldots,\mu_{y'}^{(C)},\sigma_{y'}^{(1)},\sigma_{y'}^{(2)},\ldots,\sigma_{y'}^{(C)}\right)$ in this proposed model.

Methodology(Continued)

Algorithm 1 EM Algorithm for Mixture of Isotropic Gaussians

1: $\boldsymbol{\theta_{(0)}} = \left(\phi_{1,(0)}, \ldots, \phi_{C,(0)}, \mu_{y'}^{(1,(0))}, \ldots, \mu_{y'}^{(C,(0))}, \sigma_{y'}^{(1,(0))}, \ldots, \sigma_{y'}^{(C,(0))}\right)$

2: **E-step:**

For all $c = 1, 2 \ldots, C$ and all $i = 1, \ldots, n$, calculate

$$\gamma_{i,c,k} = \frac{f_{V',\Theta''}^{(c)}\left(v', \theta'' \mid \mu_{y'}^{(c,k)}, (\sigma_{y'}^{(c,k)})^2\right)\phi_{c,k}}{\sum_{j=1}^{2} f_{V',\Theta''}^{(j)}\left(v', \theta'' \mid \mu_{y'}^{(j,k)}, (\sigma_{y'}^{(j,k)})^2\right)\phi_{j,k}}, .$$

and

$$q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta_{(k)}}\right) = \sum_{i=1}^{n}\sum_{c=1}^{C} \log\left\{f_{V',\Theta''}^{(c)}\left(v', \theta'' \mid \mu_{y'}^{(c)}, (\sigma_{y'}^{(c)})^2\right)\phi_c\right\}\gamma_{i,c,k}$$

3: **M-step:**

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

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

Methodology(Continued)

- **Anisotropic Gaussian Mixture(AGM) model:** In this model, the joint pdf of wind speed and wind direction will be

$$f_{V',\Theta''}\left(v',\theta'\right)=\sum_{j=1}^{2}\pi_j f_{V',\Theta''}^{(j)}\left(v',\theta'' \mid \mu_{x'}^{(j)},\sigma_{x'}^{(j)},\sigma_{y'}^{(j)}\right) \tag{9}$$

where, $-\infty < v' < \infty$ and $-\pi < \theta'' < \pi$.

- 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 in 2 respectively 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(Continued)

Algorithm 2 EM Algorithm for Mixture of Anisotropic Gaussians

1: $\boldsymbol{\theta_{(0)}} = \left(\phi_{1,(0)}, \ldots, \phi_{C,k}, \mu_{y'}^{(1,(0))}, \ldots, \mu_{y'}^{(C,(0))}, \sigma_{x'}^{(1,(0))}, \ldots, \sigma_{x'}^{(C,(0))} \sigma_{y'}^{(1,(0))}, \ldots, \sigma_{y'}^{(C,(0))}\right)$

2: **E-step:**

For all $c = 1, 2$ and all $i = 1, \ldots, n$, calculate

$$\gamma_{i,c,k} = \frac{f_{V',\Theta''}^{(c)}\left(v', \theta'' \mid \mu_{y'}^{(c,k)}, \sigma_{x'}^{(c,k)} \sigma_{y'}^{(c,k)}\right) \phi_{c,k}}{\sum_{j=1}^{2} f_{V',\Theta''}^{(j)}\left(v', \theta'' \mid \mu_{y'}^{(j,k)}, \sigma_{x'}^{(j,k)}\right) \sigma_{y'}^{(j,k)}} \phi_{j,k}, .$$

and

$$q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta_{(k)}}\right) = \sum_{i=1}^{n} \sum_{c=1}^{2} \log\left\{f_{V',\Theta''}^{(c)}\left(v', \theta'' \mid \mu_{y'}^{(c)}, \sigma_{x'}^{(c)} \sigma_{y'}^{(c)}\right) \phi_c\right\} \gamma_{i,c,k}$$

3: **M-step:**

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

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

Data and exploratory analysis

- We obtained this data from the International Best Track Archive for Climate Stewardship (IBTrACS), hosted by the National Climatic Data Center (NCDC) at www.ncdc.noaa.gov/oa/ibtracs/.
- The study period of the cyclones in NIO basin for our analysis spanned from 1980 to 2022.
- To visualize the cyclone trajectories in the North Indian Ocean region during this period, we created a figure 1 that displays the paths followed by the cyclones.
- We observed that the Arabian Sea experiences a lower frequency of cyclones compared to the Bay of Bengal.
- The approximate ratio between the two regions is 1:4, indicating that the Bay of Bengal has a significantly higher occurrence of cyclones.
- Furthermore, by examining 2, we identified a notable increase in the occurrence of cyclones generated in the North Indian Ocean region since 1990.

Data and exploratory analysis(Continued)

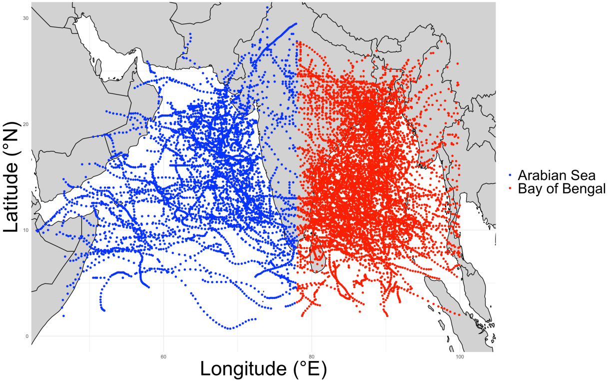

Figure: Cyclone paths in NIO basin from the year 1980 to 2022

Data and exploratory analysis(Continued)

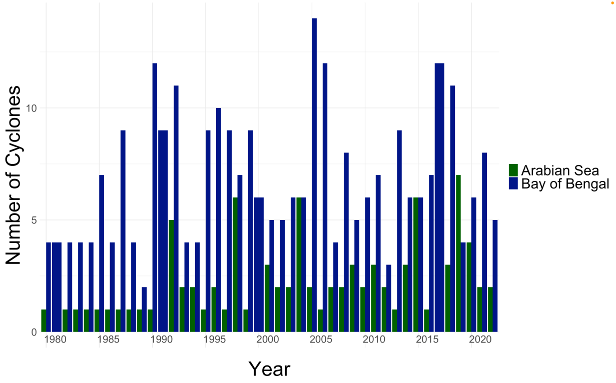

Figure: Bar plot of cyclones occur in NIO basin from the year 1980 to 2022

Model comparison

- Four metric has been used i.e RMSE,MAE, Index of Agreement(IA),χ^2 Statistic to compare the proposed models with others.
- Lower value of RMSE, MAE and χ^2 Statistic metrics indicate a higher level of accuracy for the model.
- But, IA is a metric that falls within the range of 0 to 1. When the value of IA is closer to 1, it signifies that the model is more accurate.

Model Comparison(Continued)

- From the table of Goodness of fit, we can observe that our semiparametric models IGM and AGM overperform the parametric models IG and AG respectively. But nonparametric models have less RMSE, and MAE values than our model. But, it sometimes indicates that it overfitted the scatter plot.

Goodness of fit for joint PDFs (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 for joint PDFs (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

Data Application

- We applied all of our work to the data of cyclones in the North Indian Ocean basin. Now, the joint density from parametric models i.e. IG, AG model does not fit very well with the original scatter plot of wind speed of wind direction of the cyclones in the North Indian basin.
- Then, we saw that the NP-KD and NP-JW models approached this issue in a non-parametric approach, which genuinely captured the scatter plot of wind speed and wind direction to some good extent but somehow these joint modelling seemed to overfit the scatter plot of wind speed and wind direction.
- Due to the lack of any semiparametric point of view of this problem, we made two new semiparametric models i.e. IGM and AGM model which genuinely captures the scatter plot of wind speed and direction in a better manner. We used the K-means clustering method to find the initial values of the mixture weights and other initial values of the parameters and we have seen there were two clusters, so we have taken $C = 2$ in the EM algorithm.

Data Application(Continued)

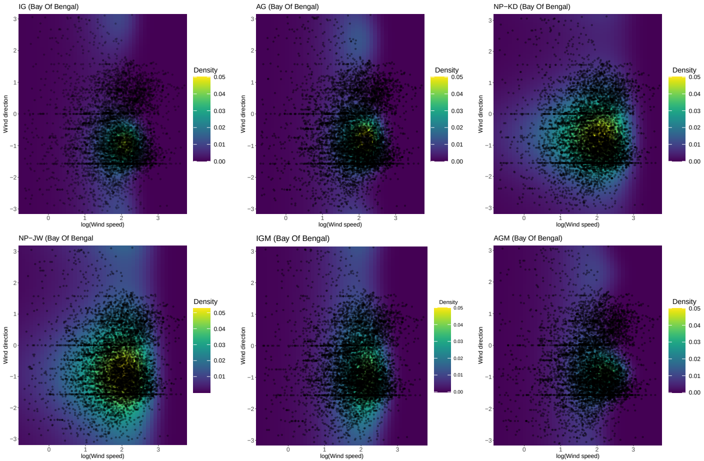

Figure: Heat maps of joint modelling of wind speed and wind direction of different models for Bay Of Bengal.

Data Application(Continued)

- In Figure 3 and 4 it is apparent that the proposed model, i.e. the IGM and AGM models offer a better fit to the scatter plot of wind speed and wind direction for cyclones in the Bay of Bengal as well as in Arabian sea when compared to the IG and AG models. Additionally, the estimated joint density of the IGM and AGM models effectively captures the mode of the data.
- Furthermore, these models prove to be adept at identifying both low-density and high-density areas within the distribution. This suggests that the proposed models are more suitable for characterizing the wind patterns of cyclones in the Bay of Bengal and Arabian Sea.

Data Application(Continued)

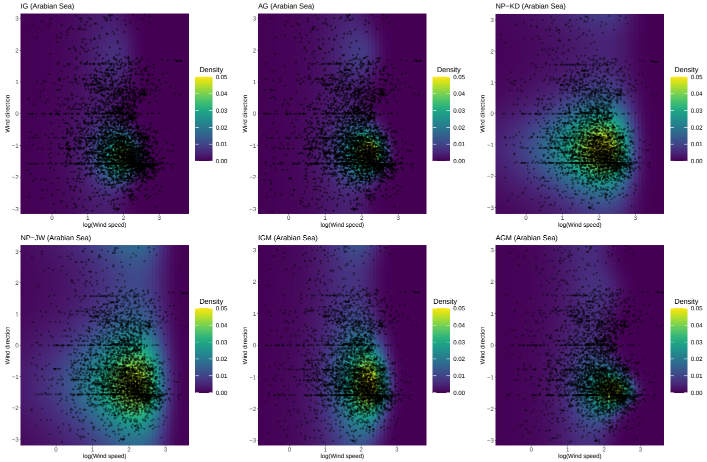

Figure: Heat maps of joint modelling of wind speed and wind direction of different models for the Arabian Sea.

Conclusion

- The primary objective of this study was to delve into the intricacies of joint modelling techniques for wind speed and wind direction within cyclones in the North Indian Ocean (NIO) Basin.
- **a)Importance of Joint Modeling:** The research underscores the critical significance of joint modelling for both wind speed and wind direction in cyclones occurring in the North Indian Ocean. This modelling approach plays a pivotal role in fortifying coastal areas against the potentially devastating impacts of these formidable weather systems.
- **b)Existing Models:** The study acknowledges the existence of previously published parametric and non-parametric models that have attempted to address the challenge of joint modelling for wind speed and wind direction in the North Indian Ocean Basin. However, it notes that parametric models struggle to provide an optimal fit for the joint relationship between wind speed and wind direction data. On the other hand, non-parametric models tend to overfit the scatter plot, which can lead to less accurate results.

Conclusion(Continued)

- **c)Proposed Models - IGM and AGM**: In response to the limitations of existing models, the research introduces two novel models, referred to as IGM and AGM. These models are designed to provide an almost perfect fit for the joint modelling of wind speed and wind direction within cyclones in the North Indian Ocean. They excel in capturing the nuances and complexities of the scatter plot, resulting in a more accurate representation of the data.
- **d)Robustness and Superiority:** The proposed IGM and AGM models are found to offer a more robust outcome compared to their parametric and non-parametric counterparts. Their ability to better fit the data and accurately capture the joint relationship between wind speed and wind direction positions them as superior choices for modelling cyclones in the North Indian Ocean Basin.

Future study

- We have done fitting RMSE but we have not explored the prediction RMSE, as the related literature has done this.So, naturally, RMSE of nonparametric models is quite low rather to our proposed model.
- But prediction RMSE will help to detect the overfitting criterion of nonparametric models, which we will explore in future.

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). Non-parametric 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] Rajasekhar, MM., Kishtawal, CMC., Prasad, MYSM., Rao, V SeshagiriV. S. Rajeevan, MM. (2014). Extended range tropical cyclone predictions for East Coast of India. In Monitoring and prediction of tropical cyclones in the Indian Ocean and climate change 137–148. Springer.
[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.

Thank you!