# Markovian Approach To The Frequency Of Tropical Cyclones And Subsequent Development Of Univariate Prediction Model

**Shreya Bhowmick**
University of Calcutta Faculty Council for Post-Graduate Studies in Science
**Surajit Chattopadhyay** (surajitchatto@outlook.com)
Amity University https://orcid.org/0000-0002-5175-2873

**Research Article**

**Keywords:** Tropical Cyclone, Markov chain, Autoregressive Model, Autoregressive Neural Network.

**Posted Date:** June 21st, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-641542/v1

# MARKOVIAN APPROACH TO THE FREQUENCY OF TROPICAL CYCLONES AND SUBSEQUENT DEVELOPMENT OF UNIVARIATE PREDICTION MODEL

SHREYA BHOWMICK[1] and SURAJIT CHATTOPADHYAY[2]

[1] Department of Atmospheric Sciences, University of Calcutta, Kolkata-700019,India.

[2] Department of Mathematics, Amity University, Kolkata-700135,India.

[2]Orchid ID- 0000-0002-5175-2873, Scopus ID- 35475366000

**Keywords**: Tropical Cyclone; Markov chain; Autoregressive Model; Autoregressive Neural Network.

## ABSTRACT

Tropical cyclones is one of the most devastating meteorological events. In the recent years we faced some very severe cyclones to super cyclone successively that caused heavy damages to life and property during the helpless situations of the global pandemic. In this paper, westudied the frequency of cyclones from the year 1891 to 2019 i.e. for 129 years on the Arabian Sea Basin, Bay of Bengal Basin and land. We have categorised the cyclones according to their wind speeds: i) Cyclonic storms and Severe cyclonic storms(CS+SCS) and ii) Depressions, Cyclonic storms and Severe Cyclonic storms(D+CS+SCS) where Depressions, Cyclonic storms and Severe Cyclonic storms have wind speeds of more than equal to 17 knots, 34 knots and 48 knots respectively. We examined the Markovian dependence of the discretized time series of the two categories mentioned earlier for the first, second, third and fourth order of a two-state Markov chain model. It is found that CS+SCS represents the First Order Two State (FOTS) model of Markov chain and D+CS+SCS represents the Second Order Two State (SOTS) model of Markov chain. Thereafter we have developed autoregressive models for the two categories and checked its goodness of fit using Willmott's indices of order 1 and 2. Its is found that CS+SCS best represents the autoregressive model of order 5 whereas D+CS+SCS could not be efficiently represented by the developed autoregressive models. So we further developed autoregressive neural networks for D+CS+SCS and obtained some significant hike in the prediction yield. Nevertheless, it is found that both the categories are clearly not serially independent.

# 1. INTRODUCION

Tropical cyclones (TC) are synoptic-scale phenomenon where a large mass of air swirls around a low pressure, counter-clockwise direction in the Northern Hemisphere and clockwise in the Southern Hemisphere. As referred to as the heat engine, a tropical cyclone is mainly fuelled by the latent heat of the moist air rising from the ocean. Gray (1968,1979) identified the six parameters necessary for the formation of tropical cyclones which are widely accepted even today . It consists of cyclonic low-level relative vorticity, large value of relative humidity in mid-troposphere, conditional instability through a deep tropospheric layer, warm and deep oceanic mixed layer, weak tropospheric vertical shear of the horizontal wind and location of disturbance a few degrees poleward of the equator.

Around 80 TCs are formed globally every year (Emanuel 2003). The North Indian Ocean (NIO) contributes about 7% of the global TCs (Gray 1979). These cyclones are primarily originating in the Bay of Bengal (BOB) and the Arabian Sea (AS). The stages of cyclones are classifed as depression(D), deep depression, cyclonic storms (CS), severe cyclonic storm (SCS), very severe cyclonic storms , extremely severe cyclonic storms and super cyclones on the basis of their associated wind speeds. On average, there are about 5 cyclones every year in the NIO. For every 4 cyclones formed in the Bay of Bengal and 1 cyclone is formed in the Arabian Sea (Mohapatra et al 2017). In the recent years, we see an increase in the intensity of pre-monsoon cyclones in the AS. Rajeevan et al. (2013) suggested that epochal variation in the intensity of TCs over the AS is correlated with the epochal variation of vertical wind shear and TC heat potential. Mohanty et al (2012) further pointed out that after 1950, the frequency of severe cyclonic storms have have increased significantly by 71% in BOB and 300% in AS during the post monsoon months.

Mooley(1980) indicated that the intensification of storms into severe storms is distributed by binomial probability model whereas the formation and landfall is based on Poisson-stochastic processes. Mooley(1981) and Mooley and Mohile(1984) showed that the mean annual frequency of severe storms have increased over the NIO in the period 1965-1980 at 10% level. Srivastava Sinha De (2000) reported that in the period 1891-1997 there is a significant decreasing trend in the number of cyclonic storms in NIO at 99% level which might be due to weakening of Hadley circulation. This decreasing trend was more in BOB than AS. Although the annual frequency was recorded having a decreasing trend of nearly 15% per hundred years, there was an increase trend of cyclones in the month of November and May mainly contributed by the BOB (O.P. Singh et al, 2001). This paper also stated that during November there was an increase of 20% per hundred years in the rate intesification of cyclonic disturbances to severe cyclonic stage. Later, using regional climate model HadRM2, O.P Singh (2007) indicated that the climate change due to increase in the atmospheric greenhouse gas concentration is the cause of this incresing trend during intense cyclone months May, October and November. His simulation experiments also showed that the frequency of post-monsoon tropical disturbances in the Bay of Bengal will increase by 50% by the year 2050. It is hot topic of research whether the increase in cyclonic frequency is related to increase in SST. Eventhough Gray(1979) and Sikka (1977) showed that sea surface temperature (SST) more than 26ºC is a parameter for the genesis of cyclone, Pattnaik (2005) and Chan (2007) reported that the variability in the planetary-scale atmospheric circulation is the main cause of the interdecadal variability of cyclonic activity over the Indian region rather than the variability of SSTs over the region. Goldenberg et al (2001) related the sudden increase in hurricane activity during the period 1995-2000 with the increase in SST and decrease in vertical wind shear in the North Atlantic . They also claimed that this increased activity will continue for the next 10 to 40 years. Later, Webster et al (2005) also showed that number of tropical storms and hurricanes cannot be correlated with increasing SST. However, it was shown in a single storm simulaton that SST and its gradient played an important role in the peak intensity and track of a tropical cyclone (M. Mandal, U. C. Mohanty, P. Sinha, M. M. Ali, 2007). Nina Črnivec, Roger K. Smith and Gerard Kilroy (2016) showed that the intensity of cyclone also depends on the latitude when SST is changed. It states that intensification is more dependant SST for higher latitude (say 25°N) than a lower latitude (say 10°N). Mandke and Bhide (2003) pointed out that during 1958-1988 the frequency of cyclones over BOB decreased eventhough the SST increased. Nolan and Rappin (2008) also stated that on increase in SST in a radiative convective environment, the wind shear actually prevents cyclone genesis.

The frequency of cyclones is also speculated to be affected by various other phenomenon or parameters. Gray (1984) correlated El Niño/Southern Oscillation (ENSO) with tropical cyclone activity in the North Atlantic whereas Chan (1984) observed the same in the Northwest Pacific. The influence of Madden Jullien Oscillation(MJO) over the Australian region was studied by Hall, J. D., A. J. Matthews, and D. J. Karoly (2001) showing increased cyclonic activity in the active phase of MJO which was strengthened during El Niño. Further Klotzbach (2014) also showed that TC activity is enhanced during and immediately following the active convective phase of the MJO while it is suppressed during and immediately following the convectively suppressed phase throughout the globe .B Kumar, P Suneetha and S R Rao (2011) related the decreasing trend of CS and SCS in the pre-monsoon with the increasing SST over NIO in general and BOB in particular whereas in the post-monsoon season the frequency of tropical systems are positively related with

Southern Oscillation Indices (SOI) and inversely correlated with MJO index for the period 1891-2008. This paper also mentions that there is influence of El Niño and La Niña on the frequency of tropical cyclones over BOB. Eric K. W. Ng and Johnny C. L. Chan (2012) points out that the tropical cyclone activity is less in the El Niño year and more in the La Niña year during Oct-Dec and the possibilty of ENSO and Indian Ocean Dipole(IOD) influencing tropical cyclone genesis and development due to variation in the atmospheric dynamic and thermodynamic conditions during the post monsoon season. A.A.Deo and D. W. Ganer (2013) also explained the increase in cyclones over the AS due to variability in SST, wind shear and energy metrics like Accumulated Cyclone Energy and Power Dissipation Index. They also showed that the cyclone season length i.e. number of days from start of the cyclone season to end of the cyclone season is increasing at the rate of 0.33 days per year mainly due to pre-monsoon season length. Later, Baburaj et al (2020) showed that there is epochal variability of cyclone frequency in AS whereas in BOB there is decrease in cyclone frequency in all three epoch which is related to epochal decadal variability in the equatorial Indian Ocean SST and vertical variation of the thermal profiles during the three epochs due to the warming of both atmosphere and ocean. However other factors influencing the frequency of cyclone in NIO might also be present which are yet to be discovered. Ki-Seon Choi, Do-Woo Kim, and Hi-Ryong Byun (2009) developed a multiple linear regression model that showed that the Tibetian plateau snow cover is an important influence in the formation of tropical cyclone in Korea. Gray (1975) introduced a new factor Yearly Genesis Parameter to measure tropical cyclogenesis which was further modified by Royer et al (1988) by including the impact of greenhouse gases.

The occurrence of cyclones causes havoc in the environment. Emanuel (2005) defined a term Power Dissipating Index based on the power dissipated by hurricanes in its lifetime that causes destruction and his study for the period 1970-2004 shows that increase in this index over North Atlantic plus western North Pacific is partly due to increase in SST. The high resolution climate model analysis by Sushil Gupta, et al (2019) suggests that in the 21$^{st}$ century the frequency of the most intense cyclones in BOB and AS will likely to increase due to warming while the total number of cyclonic disturbances should decrease. Using atmospheric general circulation models under the Intergovernmental Panel on Climate Change (IPCC) A1B scenario and phase 5 of the Coupled Model Intercomparison Project (CMIP5) models under the representative concentration pathway (RCP) 4.5 and 8.5 scenarios Murukami et al (2014) indicated decrease in projected frequency of cyclones in the basins of the Southern Hemisphere, Bay of Bengal, western North Pacific Ocean, eastern North Pacific, and Caribbean Sea and increases in the Arabian Sea and the subtropical central Pacific Ocean. Earlier Emanuel (2013) showed that on downscaling tropical cyclones of

CMIP5 models for the period 1950-2005 and comparing with 21st century there is a noticeable increase in cyclone activity as well as increase in intensity of cyclones in the North Pacific, North Atlantic and South Indian Oceans.

In the recent times, the track of the cyclone can be comprehended 48-72 hours in advance. However the intensity and accompanying storm surge needs more advanced models to be predicted more precisely (Sikka 2006). S. K. Dube, Indu Jain, A. D. Rao T. S. Murty (2009) showed the developments in storm surge prediction in BOB and AS. Being a major threat to life and property, its frequency, intensity and track needs properly estimated such that the required precautions can be taken and the budget for relief funds can be decided. Hence, it a major concern for the IMD to analyse the frequency as well as the intensity of cyclones over the North Indian Ocean.

## 2. DATA AND METHODOLOGY

In this section, we are going to develop the Markov chain models for the frequency of tropical cyclones over Arabian Sea Basin, Bay of Bengal Basin and land. The data has been collected from Cyclone eAtlas by the India Meteorological Department for the period 1891-2019 i.e. 129 years. We have categorised our study into two intensity levels : i) CS and SCS and ii) D, CS and SCS, where D represents depressions having wind speed of 17 knots or more, CS represents cyclonic storms with having wind speed of 34 knots or more and SCS represents severe cyclonic storms having wind speeds 48 knots or more. The data are converted to binary form and hence checked for two-state Markovian dependence using $\chi^2$ test. The order of the Markov chain(MC) is decided using minimisation of Bayesian Information Criterion (BIC). The implementation procedure is detailed in the following subsections.

*2.1 Category I*

In this subsection we are going to examine the frequency of tropical cyclones considered under category I i.e. both CS and SCS are taken into consideration. Total 668 cases are observed under this category for the study period under consideration . Each data point (x) corresponds to the total frequency of CS+SCS in one year within the period of study. When averaged over the entire study period, the mean frequency (x) comes out to be 5.178. Now, we convert the data series to a binary series using the following definition of a random variable:

$X_t$ = 1, if $x_t$ greater than or equal to the mean frequency

and = 0 otherwise.

Now, we apply the Markovian approach to this dicotomous time series to test for Markovian dependance. In order to do the same, we take the null hypothesis $H_0$ : The data are serially independent against the alternative hypothesis $H_1$ : There is a first order serial dependance. The following contingency table representing the number of 4 types of transition within the dicotomous time series is generated with $N_{ij}$ representing the transition count from state i to state j where both i and j realises two values 0 and 1.

Table I: Contingency table for observed transition count for first order Markovian dependance

| | $X_{t+1}=0$ | $X_{t+1}=1$ | Total |
|---|---|---|---|
| $X_t=0$ | $n_{00}=49$ | $n_{01}=22$ | $n_{0.}=71$ |
| $X_t=1$ | $n_{10}=21$ | $n_{11}=36$ | $n_{1.}=57$ |
| | $n_{.0}=70$ | $n_{0.}=58$ | |

Using table I, the transition probabilities are computed in table II as follows-

Table II: Transition probabilities for the first order Markovian dependance

| | $X_{t+1}=0$ | $X_{t+1}=1$ |
|---|---|---|
| $X_t=0$ | $p_{00}=0.690$ | $p_{01}=0.310$ |
| $X_t=1$ | $p_{10}=0.368$ | $p_{11}=0.632$ |

In table II, $p_{ij}$ represents the transitional probability from state i to state j. From table II, we can compute the stationary probability as $\pi_1 = \frac{p_{01}}{1-p_{01}+p_{11}} = 0.457$. Since $p_{01} < \pi_1 < p_{11}$ , positive serial correlation exists. Furthermore, we compute the persistance parameter $r_1 = p_{11} - p_{01} = 0.322 \neq 0$. The above computation shows that the time series is expected to have serial correlation. To further establish the above fact and to check for first order Markovian dependance we carry out $\chi^2$ test based on the null hypothesis presented above. In this particular case, the $\chi^2$ is computed using the formula $\chi^2 = \sum_i \sum_j \frac{(n_{ij}-e_{ij})^2}{e_{ij}}$ where $n_{ij}$ is the transition count as already explained and $e_{ij} = \frac{i^{th} rowtotal * j^{th} columntotal}{Totalfrequency}$

For the binary time series under consideration $\chi^2 = 13.206$ with degrees of freedom v=1 and null hypothesis is not accepted at 5% level.

Now we go to test for the second order Markovian dependance. In order to do the same, the transition counts and transition probabilities are computed and presented in table III and IV respectively.

Table III: Contingency table for observed transition count for second order Markovian dependance

| | $X_{t+1}=0$ | $X_{t+1}=1$ | Total |
|---|---|---|---|
| $X_t=00$ | $n_{000}=36$ | $n_{001}=13$ | $n_{00.}=49$ |
| $X_t=01$ | $n_{010}=9$ | $n_{011}=13$ | $n_{01.}=22$ |
| $X_t=10$ | $n_{100}=13$ | $n_{101}=8$ | $n_{10.}=21$ |
| $X_t=11$ | $n_{110}=12$ | $n_{111}=23$ | $n_{11.}=35$ |
| | $n_{..0}=70$ | $n_{..1}=57$ | |

Table IV: Transition probabilities for the second order Markovian dependance

| | $X_{t+1}=0$ | $X_{t+1}=1$ |
|---|---|---|
| $X_t=00$ | $p_{000}=0.735$ | $p_{001}=0.265$ |
| $X_t=01$ | $p_{010}=0.409$ | $p_{011}=0.591$ |
| $X_t=10$ | $p_{100}=0.619$ | $p_{101}=0.381$ |
| $X_t=11$ | $p_{110}=0.343$ | $P_{111}=0.657$ |

In this case, $\chi^2 = 14.997$ with degrees of freedom v = 3 and null hypothesis is not accepted at 5% level. Repeating similar procedure for Markovian dependance upto fourth order, the $\chi^2$ values are computed and presented in table V.

Table V: Outcomes of $\chi^2$ test for different orders of Markovian dependence for CS+SCS

| Order of Markov Chain | Value of $\chi^2$ | Degrees of freedom | Conclusion |
|---|---|---|---|
| MC(1) | 13.206 | 1 | $H_0$ is not accepted at 5% level |
| MC(2) | 14.997 | 3 | $H_0$ is not accepted at 5% level |

| | | | |
|---|---|---|---|
| MC(3) | 30.046 | 7 | $H_0$ is not accepted at 5% level |
| MC(4) | 37.969 | 15 | $H_0$ is not accepted at 5% level |

Since, it has been observed that all the four orders of Markovian dependance are acceptable, we need to decide the best representative order of Markov chain. For this purpose, we calculate BIC for all the orders.

$BIC(m) = -2\ L_m + s^m(\ln n)$ where $L_m$ is the log likelihood for order m.

Table VI: Computation of BIC for all the 4 orders of Markov chain for CS+SCS

| Order of Markov Chain | Log-likelihood | BIC |
|---|---|---|
| MC(1) | -81.460 | 172.625 |
| MC(2) | -79.689 | 178.754 |
| MC(3) | -69.337 | 177.364 |
| MC(4) | -63.804 | 204.861 |

From table VI, it is apparant that BIC is minimised for the first order Markovian dependance and hence first order two-state (FOTS) model of MC is the best representative for the CS+SCS time series converted to a binary time series.

*2.2 Category II*

In this subsection we apply similar procedure as in category I. In this case, each data point represents the total frequency of D+CS+SCS in a given year. In the present case, the mean of all frequency is 12.295. In this case, the available time series is converted to a binary time series using the following definition of a random variable:

$Y_t = 1$, if $y_t$ greater than or equal to the mean frequency

and $\quad = 0$ otherwise.

Hence, the stationary probability is computed as $\pi_1 = \frac{p_{01}}{1-p_{01}+p_{11}} = 0.453$. Since $p_{01} < \pi_1 < p_{11}$, positive serial correlation exists. We also compute the persistance parameter $r_1 = p_{11} - p_{01} = 0.513 \neq 0$. The above computation also shows that the time series is expected to have serial correlation. Furthermore, we calculate the Markovian dependance upto fourth order, the $\chi^2$ values are computed and presented in table VII.

Table VII: Outcomes of $\chi^2$ test for different orders of Markovian dependence for D+CS+SCS

| Order of Markov Chain | Value of $\chi^2$ | Degrees of freedom | Conclusion |
|---|---|---|---|
| MC(1) | 33.726 | 1 | $H_0$ is not accepted at 5% level |
| MC(2) | 42.800 | 3 | $H_0$ is not accepted at 5% level |
| MC(3) | 50.122 | 7 | $H_0$ is not accepted at 5% level |
| MC(4) | 58.494 | 15 | $H_0$ is not accepted at 5% level |

Since, it has been observed that all the four orders of Markovian dependance are acceptable, we need to decide the best representative order of Markov chain. Hence, BIC for all the orders is calculated again.

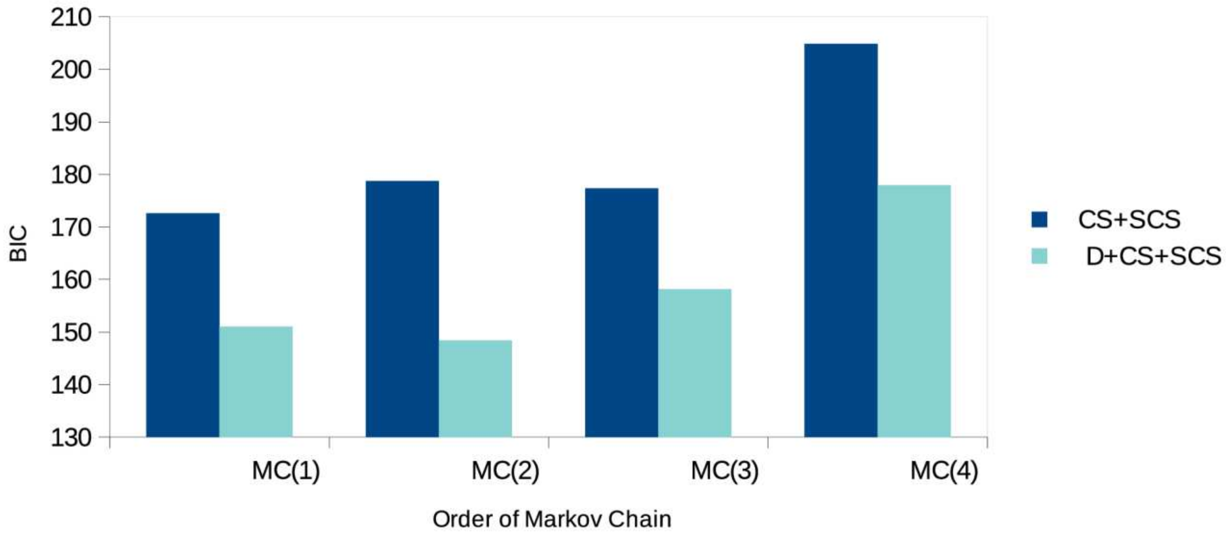
Table VIII: Computation of BIC for all the 4 orders of Markov chain for D+CS+SCS

| Order of Markov Chain | Log-likelihood | BIC |
|---|---|---|
| MC(1) | -70.675 | 151.054 |
| MC(2) | -64.516 | 148.408 |
| MC(3) | -59.742 | 158.175 |
| MC(4) | -50.350 | 177.954 |

From table VIII, it is apparent that BIC is minimised for the second order Markovian dependance and hence second order two-state (SOTS) model of MC is the best representative for the D+CS+SCS time series converted to a binary time series.

The column graph plotting the BIC values for corresponding order of Markov Chain for Category I and Category II is shown in Fig 1.

**Fig 1.** BIC values for 4 orders of Markov chain for Category I and Category II




## 3. FITTING AUTOREGRESSIVE MODELS

In the previous section, we have demonstrated Markov chain models for C+SCS and D+CS+SCS. In the case of CS+SCS, we have observed that the time series is characterised by first order Markov chain model. Also, for D+CS+SCS we have observed the second order Markov chain to characterise the time series. Now we divide the dataset having 129 datapoints into training set having 96 datapoints from year 1891 to 1986 and test set having 33 datapoints from year 1987 to 2019. Based on the outcomes presented in the previous section we apply autoregressive approach to the time series under consideration, the time series characterised by first order Markov chain leads us to interpret that the state at a given time point depends on the immediate previous time point and not on the long way it has traversed to reach upto that state. Similarly, the second order Markovian process represents a scenario where the state at a given time point depends upon the two immediately previous time points. The general auto-regressive process of order K can be mathematically presented as:

$$x_{t+1} - \mu = \sum_{k=1}^{K} [\phi_k (x_{t-k+1} - \mu)] + \epsilon_{t+1}$$

where $\mu$ is the mean of the time series, $\phi$ is the autoregressive parameter, and $\epsilon_{t+1}$ is a random shock or innovation which has $\mu_\epsilon = 0$ and variance $\sigma_\epsilon^2$ and corresponds to the residual in ordinary regression. The predictand $x_{t+1}$ is the value of the time series at time t + 1, and the predictor is the current value of the time series $x_t$. Using the training dataset, we calculate the autoregressive coeffients and with the help of the test dataset we check the goodness of fit of the above mentioned models using Willmott's index given by:

$$d = 1 - \frac{\sum_i |P_i - O_i|^\alpha}{\sum_i (|P_i - \bar{O}| + |O_i - \bar{O}|)^\alpha}$$

where $P_i$ = Predicted values

$O_i$ =Observed values

$\overline{O}$= Mean of observed values

α = 1 and 2

Table IX: Autoregressive coefficients and Willmott's index for CS+SCS

| AR(p) | Intercept | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_4$ | $\phi_5$ | Willmott(1) | Willmott(2) |
|---|---|---|---|---|---|---|---|---|
| AR(1) | 4.192 | 0.250 | - | - | - | - | 0.253 | 0.415 |
| AR(2) | 3.737 | 0.226 | 0.104 | - | - | - | 0.246 | 0.412 |
| AR(3) | 2.669 | 0.200 | 0.063 | 0.247 | - | - | 0.178 | 0.340 |
| AR(4) | 2.034 | 0.154 | 0.060 | 0.199 | 0.208 | - | 0.140 | 0.281 |
| AR(5) | 2.236 | 0.172 | 0.085 | 0.198 | 0.220 | -0.088 | 0.650 | 0.833 |

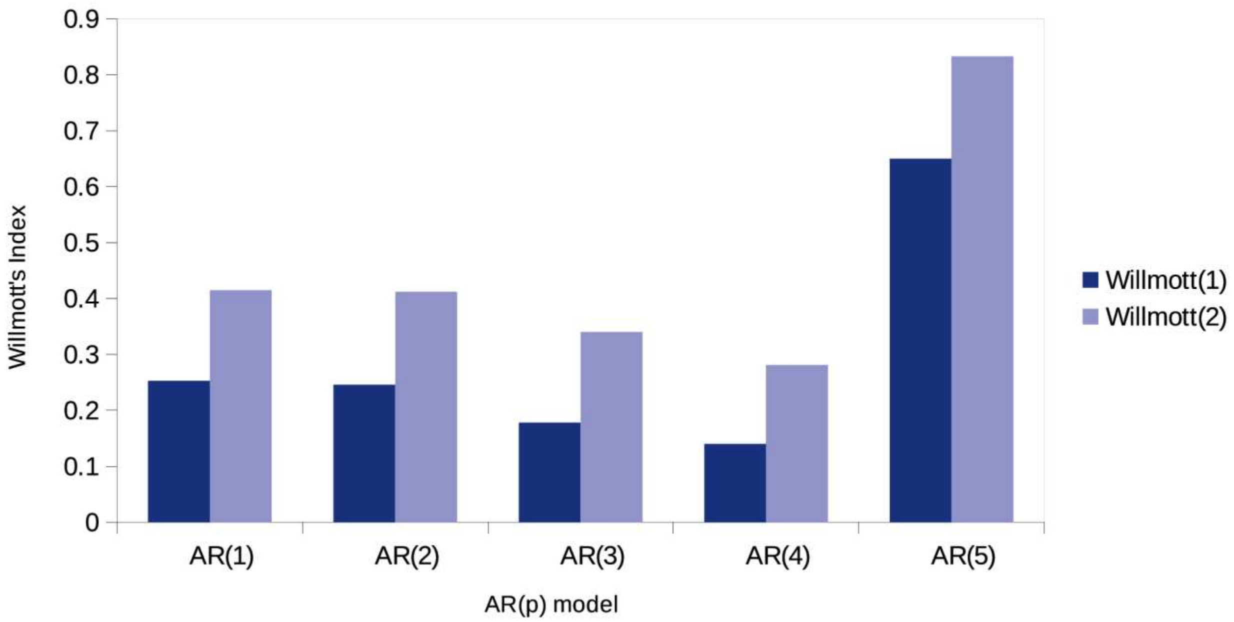
From the above table it can be clearly stated that there is strong aggreability of the AR(5) model with the observed value of frequency of CS+SCS. For this AR(5) model, the mean of the residuals $\mu_\epsilon$is found to be 7.81E-16 which is approximately equal to be0. Using Ljung-Box test, the randomness of the residuals is tested. The null hypothesis is taken that the residual is independently distributed. The test statistic is calculated using the formula:

$$Q(m) = n(n+2)\sum_{j=1}^{m} \frac{r_j^2}{n-j}$$

where n is the sample size, $r_j$ is the sample autocorrelation at lag j and m is the number of lags being tested. The test statistic for sample size 96 for lag 20 is calculated to be 8.384 with corresponding p value as 0.936 > 0.05. Thus we do not reject the null hypothesis and conclude that the residual is independently and and identically distributed i.e. white noise is present. Thus, the white noise variance $\sigma_\epsilon^2$ is calculated to be 3.034.

The column graph plotting the Willmott's Index of order 1 and 2 for the above stated autoregressive models for CS+SCS is shown in Fig 2.

**Fig 2.** Column graph to the Willmott's Index for AR(p) models for CS+SCS




The observed and predicted values of the AR(5) model for CS+SCS is plotted in Fig 3.

**Fig 3.** Line graph plotting the observed and predicted values of CS+SCS using AR(5) model

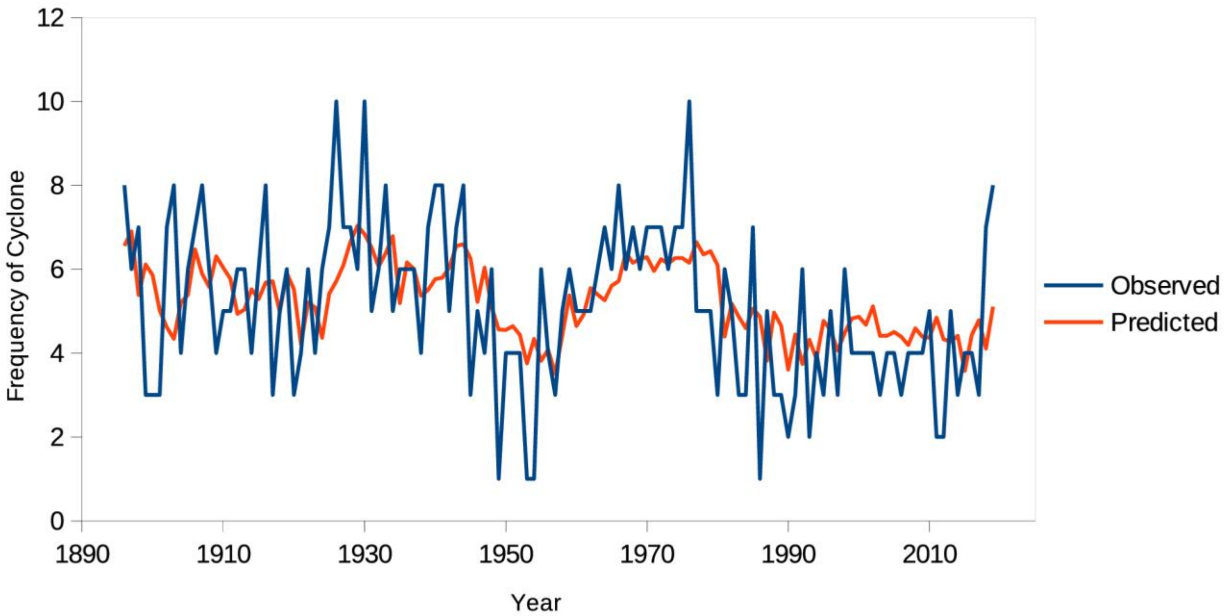



Table X: Autoregressive coefficients and Willmott's index for D+CS+SCS

| AR(p) | Intercept | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_4$ | $\phi_5$ | Willmott(1) | Willmott(2) |
|---|---|---|---|---|---|---|---|---|
| AR(1) | 8.732 | 0.343 | - | - | - | - | 0.376 | 0.475 |
| AR(2) | 5.551 | 0.213 | 0.369 | - | - | - | 0.393 | 0.496 |
| AR(3) | 4.307 | 0.140 | 0.315 | 0.220 | - | - | 0.343 | 0.429 |
| AR(4) | 3.578 | 0.110 | 0.268 | 0.168 | 0.153 | - | 0.288 | 0.373 |
| AR(5) | 2.921 | 0.072 | 0.243 | 0.131 | 0.119 | 0.214 | 0.238 | 0.268 |

From the above table we observe that the Willmott's index of the AR(p) models are low. Thus, these models have a lack of fit and is not good for prediction. Hence, we use neural network approach to fit an AR(p) model to the dataset of D+CS+SCS.

## 4. COMPARISON OF AR MODEL WITH A NON-LINEAR PREDICTIVE METHODOLOGY

In this section, a non-linear predictive model is designed using the univariate time series of frequency of cyclone for D+CS+SCS. An Autoregressive Neural Network (AR-NN) is designed for different lags. We divide the data into 75% training dataset and 25% test datatest. The input matrix of an AR-NN of order p consists of p columns for p previous states and one column for the current state. The input matrix is fed into the AR-NN and activated using logistic function to give the predicted results for the test dataset. The AR-NN models are designed upto lag 5 and the Willmott's index is computed for each case. Further, we compute the prediction yield of both AR model and AR-NN model for 5% error, 10% error and 15% error for each lag. The prediction yield is given by the formula :$PredictionYield = \frac{Totalno.ofcaseswithinx\%error}{Totalno.oftestcases}$where x = 5,10 and 15.

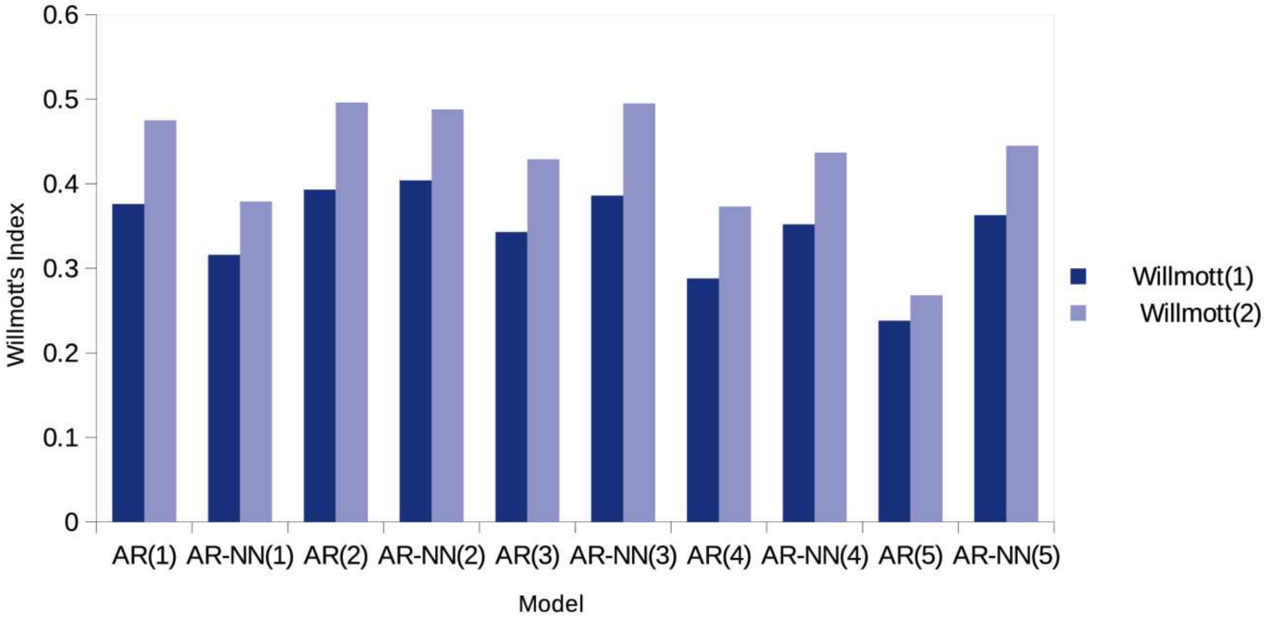
Finally the results are compared in table XI.

Table XI: Comparison of AR and AR-NN model for D+CS+SCS

| Lag | Model | Prediction Yield ( in %) | | | Willmott(1) | Willmott(2) |
|---|---|---|---|---|---|---|
| | | 5% error | 10% error | 15% error | | |
| 1 | AR(1) | 15.15 | 21.21 | 36.36 | 0.376 | 0.475 |
| | AR-NN(1) | 15.15 | 33.33 | 36.36 | 0.316 | 0.379 |
| 2 | AR(2) | 18.18 | 33.33 | 45.45 | 0.393 | 0.496 |
| | AR-NN(2) | 21.21 | 36.36 | 45.45 | 0.404 | 0.488 |
| 3 | AR(3) | 21.21 | 33.33 | 39.39 | 0.343 | 0.429 |
| | AR-NN(3) | 18.18 | 30.30 | 39.39 | 0.386 | 0.495 |
| 4 | AR(4) | 15.15 | 30.30 | 36.36 | 0.288 | 0.373 |
| | AR-NN(4) | 24.24 | 30.30 | 39.39 | 0.352 | 0.437 |
| 5 | AR(5) | 18.18 | 30.30 | 45.45 | 0.238 | 0.268 |
| | AR-NN(5) | 18.18 | 30.30 | 45.45 | 0.363 | 0.445 |

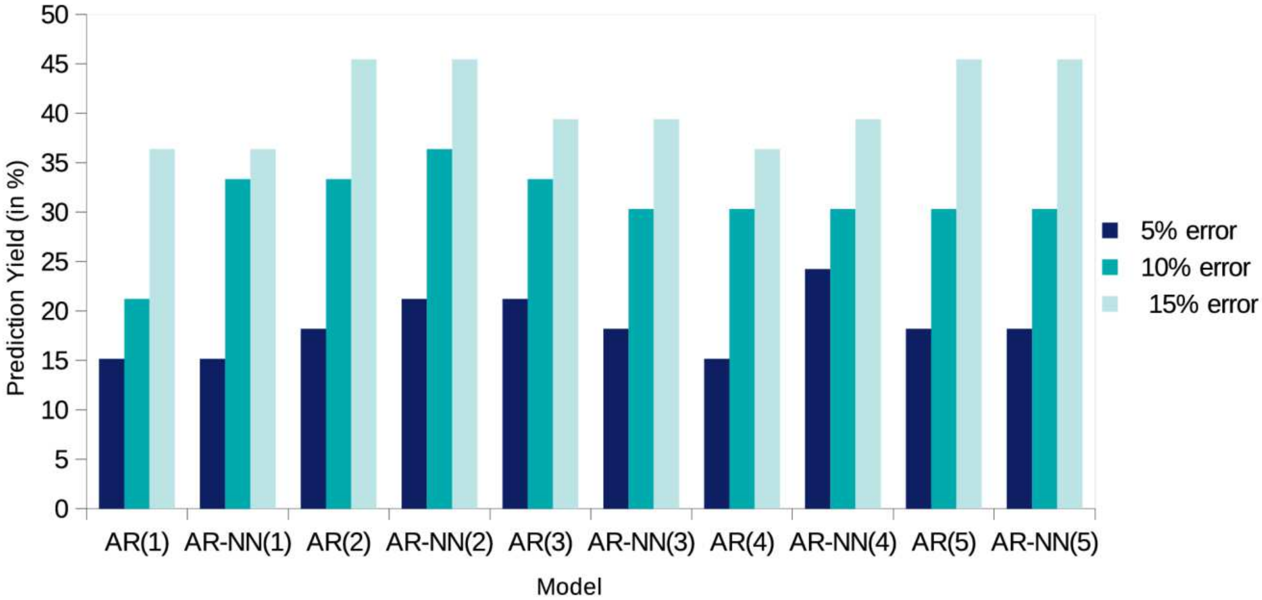
The column graph comparing the AR models and AR-NN models with respect to their Willmott's Index for D+CS+SCS is shown in Fig 4.

**Fig 4.** Comparison of AR model and AR-NN model using Willmott's Index for D+CS+SCS




The column graph comparing prediction yield of different AR models and AR-NN models for D+CS+SCS is shown in Fig 5.

**Fig 5**. Prediction yields of AR model and AR-NN models of different orders for D+CS+SCS




## 5. CONCLUSION

In the rigorous study presented in the previous sections, we have reported a Markov chain model and univariate prediction of tropical cyclones over Arabian Sea Basin, Bay of Bengal Basin and land collected fot the period 1891-2019. The data have been categorised as per their intensity levels into two categories: i) CS+SCS and ii) D+CS+SCS. Here D stands for depressions with wind speed greater than equal to 17 knots. Further details have been presented in section 2 of the present paper. Since the data corresponds to continuous random variables, we have discretized them for the application of Markov chain, which is a discrete time series approach. While discretizing we have obtained a dicotomous time series. This methodology has been adopted to each category mentioned above and accordingly the transition probabilities have been computed for each category to find the stationary probability. This computation has been presented in table II and table IV. It has been observed that in each case of CS+SCS, $p_{01} < \pi_1 < p_{11}$ and hence it has been interpreted that positive serial correlation exists. This has further been supplemented by a non-zero persistance parameter. In order to further consolidate the outcomes we have carried out $\chi^2$ test to check for first order two-state Markovian dependance. We have also carried out similar test for Markovian dependance for second, third and fourth order respectively and in each case, it has been observed that the computed $\chi^2$ is exceeding the corresponding tabular value with appropriate degrees of freedom and 5% level of significance. The results are displayed in table V where it has been clearly shown that for every competing order of two-state Markov chain model, the null hypothesis $H_O$ assuming serial independence is not accepted at 5% level. In order to choose the best order of Markov chain among the 4 orders, we have implementented BIC minimisation procedure and the results for CS+SCS are presented in table VI. This table shows that BIC gets its minimum for FOTS model of Markov chain and hence FOTS is considered to be the best representative Markovian process for CS+SCS. The outcomes are presented in table VI. A similar computational procedure, when carried out for D+CS+SCS, table VII shows that the null hypothesis of serial independance is not acceptable at 5% level (see table VII). However, in the case of D+CS+SCS, the BIC minimisation establishes SOTS as the best representative Markov chain model (see table VIII). To have a comparative view we have depicted the results in figure 1. In the subsequent phase of the study we have developed autoregressive models for the two categories under consideration. In table IX we have presented the autoregressive coefficients for CS+SCS and it is clear from this table that we have checked for autoregressive processes upto order 5. In order to check for goodness of fit of the autoregressive model of a given order we have computed the Willmott's indices of order 1 and 2. The values of Willmott's indices are also presented in table IX and it is clearly visible that for AR5 both the Willmott's indices are above 0.5 and the second order Willmott's index is above 0.8. This strongly leads us to conclude that AR5 is the best fit autoregressive model for CS+SCS. In order to test the randomness of the residuals, we have implemented the Ljung-Box test where it is has been observed that the residuals are independently and identically distributed and as a consequence we have concluded the presence of white noise within this process. We have also pictorially represented the values of Willmott's index for CS+SCS in figure 2. Also in figure 3, we have displayed the observed and predicted frequencies in the test cases, which shows that there is a significant degree of closeness in the patterns of CS+SCS frequencies in the observed and predicted cases. Contrary to what happened in CS+SCS, the autoregressive models could not perform so efficiently in case of D+CS+SCS. Table X shows that an increase in order of autoregression above 2 has resulted in decay in the values of Willmott's index. Although AR1 and AR2 have second order Willmott's index close to 0.5, they cannot be interpreted as good univariate predictive model. Considering the failure of conventional autoregressive procedure in case of D+CS+SCS we have developed autoregressive neural networks (AR-NN) for D+CS+SCS. It has been observed that applying logistic activation function for AR-NN we have observed that for the test cases the Willmott's index is not getting any significant improvement from the conventional autoregression process. However, we have observed some significant hike in prediction yield for the first and second orders with respect to acceptable prediction error of 10%. In that sense for D+CS+SCS implementation of neuro-computing methodology in autoregressive manner has given some advantage over the conventional autoregression procedure. Considering the prediction yield associated with 5% error we have observed a significant hike in the case of fourth order AR-NN. Hence in general we can say that AR-NN is a better predictive tool than conventional AR in case of D+CS+SCS.

While concluding, we would like to note that neither CS+SCS nor D+CS+SCS are characterised by serial independance. This means that somehow the frequency of cyclonic storms, severe cyclonic storms and depressions of a given year has some influence on the subsequent year. As we incorporate depressions, the prediction of frequencies becomes more difficult. This indicates towards incorporation of some degree of complexity to the system by the depressions that did not develop into cyclonic storms or severe syclonic storms. In view of the same we propose to carry a study on the fractal behaviour of the time series as our future study.

# 6. ABBRIEVIATIONS

AR-NN, Autoregressive Neural Network; AS, Arabian Sea; BIC, Bayesian Information Criterion; BOB, Bay of Bengal; CS, Cyclonic Storm; D, Depression; MC, Markov Chain; NIO, North Indian Ocean; SCS, Severe Cyclonic Storm; TC, Tropical cyclone;

# 7. DECLARATIONS

**Competing Interests**

The authors declare that they have no competing interests.

**Consent for Publication**

Not applicable.

**Ethics Approval and Consent to participate**

Not applicable.

**Funding**

Not applicable.

**Availibility of Data and Materials**

All the dataset used for this study is available in Cyclone eAtlas by Indian Meteorological Department: http://14.139.191.203/AboutEAtlas.aspx.

**Authors' Contribution**

Surajit Chattopadhyay has conceived and designed the analysis. Also, he has supervised the research. Shreya Bhowmick has performed the analysis and prepared the manuscript.

## 8. ACKNOWLEDGEMENT

Shreya Bhowmick acknowledges scientific discussion with Prof. Subrata Kumar Midya and Dr. Goutami Chattopadhyay, University of Calcutta, for suggestions and discussions during the research.

## 9. REFERENCES

1. A.A.Deo and D. W. Ganer, Variability in Tropical Cyclone Activity Over Indian Seas in Changing Climate: International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 (2013)

2. B. Kumar, P. Suneetha, S. R. Rao : Impact of rising sea temperatures on frequency of tropical storms and their relationaships over North Indian Ocean. Canadian Journal of Pure and Applied Sciences ,Vol. 5, No. 3, pp. 1675-1683 (2011)

3. Baburaj P.P., Abhilash S., Mohankumar K. & Sahai A.K. : On the Epochal Variability in the Frequency of Cyclones during the Pre-Onset and Onset Phases of the Monsoon over the North Indian Ocean. Advances in Atmospheric Sciences volume 37, pages 634–651(2020)

4. Chan JCL. 1984. An observational study ofthe physical processes responsible for tropi-cal cyclone motion.J.Atmos. Sci.41:1036–48

5. Chan, J. C. L., 2007: Interannual variations of intense typhoon activity. Tellus A, 59, 455–460, https://doi.org/10.1111/j.1600-0870.2007.00241.x.

6. Črnivec, N., Smith, R.K., Kilroy, G.: Dependence of tropical cyclone intensification rate on sea-surface temperature. Q. J. R. Meteorol. Soc. 142, 1618–1627 (2016)

7. Emanuel, K., 2003: Tropical cyclones. Annual Review of Earth and Planetary Sciences, 32, 75–104, https://doi.org/10.1146/annurev.earth.31.100901.141259.

8. Emanuel, K. 2005. Increasing destructiveness of tropical cyclones over the past 30 years. Nature. 436:686-688.

9. Emanuel, K. A., 2013: Downscaling CMIP5 climate models show increased tropical cyclone activity over the 21st century. Proc. Natl. Acad. Sci. USA, 110, 12 219–12 224, doi:10.1073/pnas.1301293110

10. Eric K. W. Ng and Johnny C. L. Chan, Interannual variations of tropical cyclone activity over the north Indian Ocean: International Journal of Climatology 32: 819–830 (2012)

11. Gray WM. 1968. Global view of the origin of tropical disturbances and storms. Monthly Weather Review 96: 669–700.

12. Gray, W. M.: 1975, Tropical Cyclone Genesis, Dept. of Atmospheric Science Paper, No. 234, Colorado State University, Fort Collins, CO, p. 121.

13. Gray WM. 1979. Hurricanes: Their formation, structure, and likely role in the tropical circulation. In Meteorology over the Tropical Oceans, Shaw DB (ed.), Royal Meteorological Society: Bracknell, UK; 155–218.

14. Gray, W. M. 1984. Atlantic seasonal hurricane frequency. Part I: El Niño and 30 mb Quasi-Biennial Oscillation influences. Monthly Weather Review 112:1649–68.

15. Goldenberg, S. B., C. W. Landsea, A. M. Mestas-Nunez, and W. M. Gray, 2001: The recent increase in Atlantic hurricane activity: Causes and implications. Science, 293, 474–479.

16. Hall, J. D., A. J. Matthews, and D. J. Karoly, 2001: The modulation of tropical cyclone activity in the Australian region by the Madden–Julian oscillation. Mon. Wea. Rev., 129, 2970–2982.

17. Klotzbach, P. J. (2014), The Madden-Julian oscillation's impacts on worldwide tropical cyclone activity, J. Clim., 27, 2317– 2330, doi:10.1175/JCLI-D-13-00483.1.

18. Ki-Seon Choi, Do-Woo Kim, and Hi-Ryong Byun, Statistical Model for Seasonal Prediction of Tropical Cyclone Frequency around Korea, Asia-Pacific Journal of Atmospheric Sciences, 45, 1, p. 21-32 (2009)

19. Mandke, S., and Bhide, U.V.: A study of decreasing storm frequency over Bay of Bengal. J. Ind. Geophys. Union. 7, 53–58 (2003)

20. Mandal M. , Mohanty U. C. , Sinha P. , Ali M. M. , Impact of sea surface temperature in modulating movement and intensity of tropical cyclones, Nat Hazards (2007) 41:413–427, DOI 10.1007/s11069-006-9051-8

21. Mohapatra M, Bandyopadhyay B.K., Rathore. L.S. , Tropical Cyclone Activity over the North Indian Ocean, Springer International Publishing (2017)

22. Mohanty, U.C., Osuri, K.K., Pattanayak, S., Sinha, P.: An observational perspective on tropical cyclone activity over Indian seas in a warming environment. Nat. Hazards. 63, 1319–1335 (2012)

23.Mooley, D.A.: Severe cyclonic storms in the Bay of Bengal, 1877–1977. Monsoon Weather Review 108, 1647–1655 (1980)

24.Mooley, D.A.: Increase in the frequency of the severe cyclonic storms of the Bay after 1964—possible causes. Mausam 32, 35–40 (1981)

25. Mooley, DA. and Mohile, CM.1984. Cyclonic storms of the Arabian Sea, 1877–1980. Mausam. 35:127-134.

26. Murukami H., Hsu P., Arakawa O., Li T. Influence of Model Biases on Projected Future Changes in Tropical Cyclone Frequency of Occurrence: DOI: 10.1175/JCLI-D-13-00436.1 (2014)

27. Nolan, D.S., Rappin, E.D.: Increased sensitivity of tropical cyclogenesis to wind shear in higher SST environments. Geophys. Res. Lett. 35, L14805 (2008). https://doi.org/10.1029/2008GL034147

28. O.P. Singh, ―Long-term trends in the frequency ofsevere cyclones of Bay of Bengal: observations and simulations‖, Mausam, vol. 58, no.1, pp. 59–66, 2007.

29. Pattanaik DR. 2005. Variability of oceanic and atmospheric conditions during active and inactive periods of storms over the Indian region. International Journal of Climatology 25: 1523–1530.

30. Rajeevan M., Srinivasan J., Niranjan Kumar, K.,Gnanaseelan C. and Ali M. M.. On the epochal variation of intensity of tropical cyclones in the Arabian Sea. Atmospheric Science Letters 14: 249–255 (2013)

31. Royer J.-F. , Chauvin F. , Timbal B. , Araspin P. and Grimal D., A GCM study of the impact of greenhouse gas increase on the frequency of occurance of tropical cyclone, Climatic Change 38: 307–343, 1998

32. Sikka D.R. 1977. Some aspects of the life history, structure and movement of monsoon depressions. Pure and Applied Geophysics 115: 1501–1529.

33. Sikka D.R. Major advances in understanding and prediction of tropical cyclones over north Indian Ocean : A Perspective. Mausam, 57, 1, 165-196 January 2006.

34. Singh O.P., Khan T.M.A., Rahman S.: Has the frequency of intense tropical cyclones increased in the north Indian Ocean? Curr. Sci. 80, 575–580 (2001)

35. S. K. Dube, Indu Jain, A. D. Rao, T. S. Murty, Storm surge modelling for the Bay of Bengal and Arabian Sea, Nat Hazards (2009) 51:3–27, DOI 10.1007/s11069-009-9397-9

36. Srivastav A.K., SinhaRay K.C., De U.S.: Trends in the frequency of cyclonic disturbances and their intensification over Indian Seas. Mausam 51, 113–118 (2000)

37. Sushil Gupta, Indu Jain, Pushpendra Johari and Murari Lal.: Impact of climate change on tropical cyclones frequency and intensity on Indian coasts(2019)

38. Webster P.J., Holland G.J., Curry J.A and Chang H.R. (2005). Changes in tropical cyclone number, duration and intensity in a warming environment. Science. 309:1844-1846.

39. Willmott CJ, Matsuura K (1995) Smart interpolation of annually aver-aged air temperature in the United States. J Appl Meteorol 34(12):2577–2586

40. Wilks, D. S., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press.