

## Statistical prediction of seasonal cyclonic activity over North Indian Ocean

S. BALACHANDRAN\* and B. GEETHA

*India Meteorological Department, Meteorological Centre, Bangalore, India*

*(Received 27 April 2010, Modified 6 December 2010)*

**e mail : balaimd@gmail.com**

**सार** – उत्तरी हिंद महासागर (एन. आई. ओ.) में अक्टूबर से दिसम्बर तक की उत्तरी पूर्वी मानसून ऋतु में सबसे अधिक चक्रवातीय गतिविधियाँ होती हैं। इस ऋतु के दौरान उत्तरी हिन्द महासागर में चक्रवातीय गतिविधि औसतन लगभग 20 दिन होती है। इस शोध पत्र में 1971–2000 तक के पिछले 30 वर्षों के सुपरिचित जलवायु सूचकांकों और क्षेत्रीय परिसंचरण लक्षणों का उपयोग करके अक्टूबर से दिसम्बर तक की चक्रवाती ऋतु के दौरान उत्तरी हिन्द महासागर में मौसमी चक्रवातीय गतिविधि के लिए सांख्यिकीय पूर्वानुमान देने का प्रयास किया गया है।

सह संबंध विश्लेषण का उपयोग करके संभावित पूर्वसूचकों का पता लगाया गया है और स्क्रिनिंग समाश्रयण तकनीक का उपयोग करके अधिकतम संख्या में पूर्वसूचकों का चयन किया गया है। 1971–2000 तक के 30 वर्षों के आँकड़ों के आधार पर प्रत्येक पूर्वसूचक के विभिन्न अंतरालों के लिए उत्तरी हिन्द महासागर (एन. आई. ओ.) में अक्टूबर से दिसम्बर के दौरान चक्रवातीय विक्षोभ के दिनों की संख्या के परिस्थितिजन्य औसतों का विश्लेषण करके चक्रवातीय विक्षोभ (सी. डी.) के दिनों की संख्या का गुणवत्ता पूर्वानुमान लगाने का प्रयास किया गया है। इस स्कीम के आधार पर उसके बाद की 2001 से 2009 तक की परीक्षण अवधि के लिए पूर्वानुमान और उनके वैधीकरण पर भी चर्चा की गई है। समान पूर्वसूचकों का उपयोग करके अक्टूबर से दिसम्बर के दौरान उत्तरी हिन्द महासागर में चक्रवाती विक्षोभ के दिनों के पूर्वानुमान के लिए बहुसमाश्रयण निदर्श का विकास करके मात्रात्मक पूर्वानुमान करने का भी प्रयास किया गया है। समाश्रयण निदर्श से 70 प्रतिशत अंतर वार्षिक विविधता का पता चलता है। आकलन का वर्ग माध्य मूल त्रुटि 05 दिन और बायस त्रुटि 0.36 दिन है। विचाराधीन वर्ष को छोड़कर, 29 वर्ष के आँकड़ों का उपयोग करके प्रत्येक वर्ष के लिए जैकनाइफ पद्धति द्वारा समाश्रयण निदर्श की प्रतिपुष्टि की गई है। इस मॉडल का 2001 से 2009 तक के वर्षों के स्वतंत्र डाटासेट के लिए भी परीक्षण किया गया है। इस मॉडल के निष्पादन के विशिष्ट लक्षणों की विवेचना इस शोध-पत्र में की गई है।

**ABSTRACT.** The Northeast monsoon season of October to December (OND) is the primary season of cyclonic activity over the North Indian Ocean (NIO). The mean number of days of cyclonic activity over NIO during this season is about 20 days. In the present study, statistical prediction for seasonal cyclonic activity over the North Indian Ocean during the cyclone season of October to December is attempted using well known climate indices and regional circulation features during the recent 30 years of 1971-2000.

Potential predictors are identified using correlation analysis and optimum numbers of predictors are chosen using screening regression technique. A qualitative prediction for number of Cyclonic Disturbance (CD) days is attempted by analysing the conditional means of the number of CD days during OND over NIO for different intervals of each predictor based on the 30 year data of 1971-2000. Predictions and their validations for the subsequent test period of 2001 to 2009, based on this scheme, are discussed. An attempt for quantitative prediction is also made by developing a multiple regression model for prediction of number of CD days over the NIO during OND using the same predictors. The regression model accounts for 70% of the inter annual variance. The root mean square error of estimate is 5 days and the bias error is 0.36 days. The regression model is cross validated by Jackknife method for each individual year using the data of 29 years from the sample excluding the year under consideration. The model is also tested for independent dataset for the years 2001 to 2009. Salient features of the model performance are discussed.

**Key words** – Statistical prediction, Cyclonic activity, North Indian Ocean, Correlation, Screening regression, Conditional mean, Multiple regression, Validation, Jackknife method.

### 1. Introduction

Tropical cyclones are intense low pressure vortex systems that originate over warm oceans. They draw

energy from the ocean and the atmosphere and tend to move poleward affecting coastal areas of the world extensively. Every year about 80-90 tropical cyclones (TC) form over various oceanic basins of the world and

affect various coastal areas. Initially a TC forms as a low pressure area and then gradually intensifies through a stage of Depression(D) to Cyclonic Storm (CS) / Severe Cyclonic Storm (SCS) / Very Severe Cyclonic Storm (VSCS) / Super Cyclonic Storm (SuCS). Depending on the maximum intensity, a low pressure system is classified as a Depression / CS / SCS / VSCS / SuCS. The intensity classifications are based on associated wind speeds and pressure difference between the cyclone centre and the outer closed isobar.

The destructive effects of a TC over the coastal areas are due to the associated gale force winds, torrential rains and storm surge. Associated with the coastal crossing of TCs, heavy damages to life and property occur due to gale winds and heavy rains in and around the areas of landfall. Sometimes, storm surge, the most devastating feature of a TC, also occur leading to massive destruction to life and property along the coastal belt of TC landfall.

Over the North Indian Ocean (NIO), the Northeast monsoon season of October to December (OND) is the chief cyclone season even though TCs do form during the pre-monsoon months of April-May. India, having an extensive coast line, is vulnerable to destructive effects of TCs that form over NIO. Some of the coastal areas, especially in the eastern coast, have been subjected to the fury of very severe cyclonic storms time and again and so, every year, the Government, social planners and NGOs work out disaster mitigation action plans well in advance of the ensuing cyclone season.

Prediction of track of a tropical cyclone two to three days prior to landfall based on climatology and persistence have been in vogue for long [Sikka and Suryanarayana (1972), Neumann and Mandal (1978)] and this has been very useful for civic planners to work out disaster mitigation plans. However, forecasting the seasonal cyclonic activity in advance of the ensuing cyclone season, even though very challenging, would serve as a major input for disaster management activities. A number of attempts have been made earlier to predict the seasonal cyclonic activity over various oceanic basins. Gray (1984-1&2) has developed a multiple regression model for forecasting Atlantic Hurricane Frequency and activity using QBO, El Nino and sea level pressure anomalies over the Caribbean region. Landsea *et al.* (1998) attributed intra seasonal variation in Atlantic Hurricane activity during August - October 1995 to the upper tropospheric circulation anomalies over the tropical North Atlantic and changes in SST across the tropical and subtropical North Atlantic. Chan *et al.* (1998) have developed an operational statistical forecast model for seasonal tropical cyclone activity over the Western North Pacific (WNP) and the South China Sea using the climate

indices representing the ENSO phenomenon and the environmental conditions over the East Asia and the WNP. Subsequent improvements to this forecast scheme were made by Chan *et al.* (2001) by including a few more predictors in the form of temporal changes in SOI, strength of the Australian monsoon and the intensity of the subtropical high in the South Pacific. Jury *et al.* (1999) have shown the influence of SST anomalies over the central and south Indian ocean, OLR anomalies over the central Amazon and east African region, 200 hPa zonal wind anomalies in the 10-25°S band over Africa and south Atlantic and the surface wind anomalies over the central Indian ocean on the number of days tropical cyclone over the SouthWest Indian Ocean. For the Indian region, a few works have been carried out in this direction by earlier workers [Rajeevan and Butala (1990), Singh and Khan (2001), Singh *et al.* (2007), Krishna Kumar *et al.* (2007)].

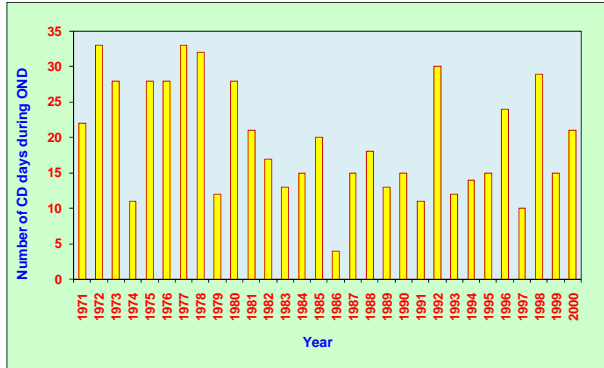
In this present study, statistical prediction of seasonal cyclonic activity over NIO during OND is attempted by identifying global / regional predictors using correlation and screening regression analysis and arriving at predictions based on these predictors using conditional mean analysis and multiple regression equation.

## 2. Tropical cyclonic activity over North Indian Ocean

The Northeast monsoon season of October to December is the main cyclone season with about 5 depressions forming over the North Indian Ocean. Out of this 5, 2 to 3 reach the intensity of CS & above and 1 to 2 reach the intensity of SCS & above. The life period of cyclonic disturbances are important from the point of view of their influences and hence the actual number of days of presence of cyclonic disturbance (CD) (which includes depression stage onwards) over the NIO during the cyclone season (OND) is termed as 'Number of CD days' or simply 'CD days' and considered to represent the seasonal cyclonic activity for the present study. The time series of number of CD days during OND over NIO for the recent 30 year period of 1971-2000 is presented in Fig. 1. The mean number of CD days based on this data is 20 days. Table 1 gives the statistical details of the number of CD days over NIO during October to December. These statistics were derived using CWCDSTAT, a software for generation of statistics on cyclones and depressions over the NIO (IMD, 2008).

## 3. Data and methodology

As mentioned in the previous section, the number of CD days over NIO during OND is taken to represent the seasonal cyclonic activity and so this serves as the predictand. Here, the term CD includes the stages of



**Fig. 1.** Time series of number of CD days during OND over NIO for the period 1971-2000

Depression, Deep Depression, Cyclonic Storm, Severe Cyclonic Storm, Very Severe Cyclonic Storm and Super Cyclonic Storm and the term ‘Number of CD days’ includes the total number of days from the day of intensification from low / well-marked low to depression to the day of weakening from depression to low / well-marked low. The number of CD days during OND for the period 1971-2000 were extracted using CWCSTAT software (IMD, 2008).

In order to predict the seasonal cyclonic activity qualitatively, the number of CD days are categorised based on the sample mean (20 days) and half of the standard deviation (4 days). The following criteria is adopted for the categorisation of seasonal cyclonic activity (CA) :

<i>Categorisation of number of CD days</i>	<i>Qualitative definition of seasonal cyclonic activity (CA)</i>
Less than 12 days	Subdued CA
Between 12 and 16 days	Below normal CA
Between 16 and 24 days	Normal CA
Greater than 24 days	Above normal CA

Here, half of the standard deviation is considered for categorisation in order to get sufficient number of samples in each of the defined categories.

Next, a set of potential predictors comprising of global climate indices like Nino indices, SOI, QBO, Atlantic hurricane frequency as well as regional circulation features such as zonal and meridional winds at lower and upper troposphere and SST are considered to

**TABLE 1**

**Statistics of number of CD days over NIO during October to December**

Month(s)	Mean (days)	Standard deviation (days)
October	7.30	4.09
November	8.93	5.19
December	3.33	3.34
OND	19.57	7.99

(based on 30 year data of 1971-2000)

predict the CD days. The data were downloaded from the websites of NCEP ([www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysisi](http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysisi)) and Australian Bureau of Meteorology ([www.bom.gov.au](http://www.bom.gov.au)) for the period 1971 to 2000. From this set of potential predictors, an optimum set are selected based on the following method. First, the teleconnections between the number of CD days and various global climate indices, regional circulation features in antecedent modes were studied by means of correlation techniques. Depending on the significance of CCs obtained and the inter correlation amongst each other, optimum number of predictors are identified by means of screening regression methods.

Finally, for the prediction of number of CD days using the selected predictors, two methods are adopted – (i) conditional mean analysis and (ii) multiple regression methods. A qualitative prediction scheme based on conditional mean analysis and a quantitative prediction based on multiple regression equation are developed using the same datasets. The regression model is cross validated by Jackknife method. Predictions based on both the methods are validated for the subsequent independent test period of 2001 to 2009.

#### 4. Results and discussion

4.1. The potential predictors are chosen based on two factors similar to the way Chan *et al.* (1998) did for the Typhoons of West North Pacific basin. (i) Predicting parameters should be those which are likely to have physical link with the development and / or movement of TCs. (ii) Since the forecast should be made at least one month before the start of the ensuing cyclone season (OND), any predictor chosen must be available and easily accessible latest by September.

4.2. First, parameters which are readily available as indices for climate studies, such as ENSO, are considered.

TABLE 2

CCs between various climate indices during each month of the past one year and the number of CD days over NIO during OND

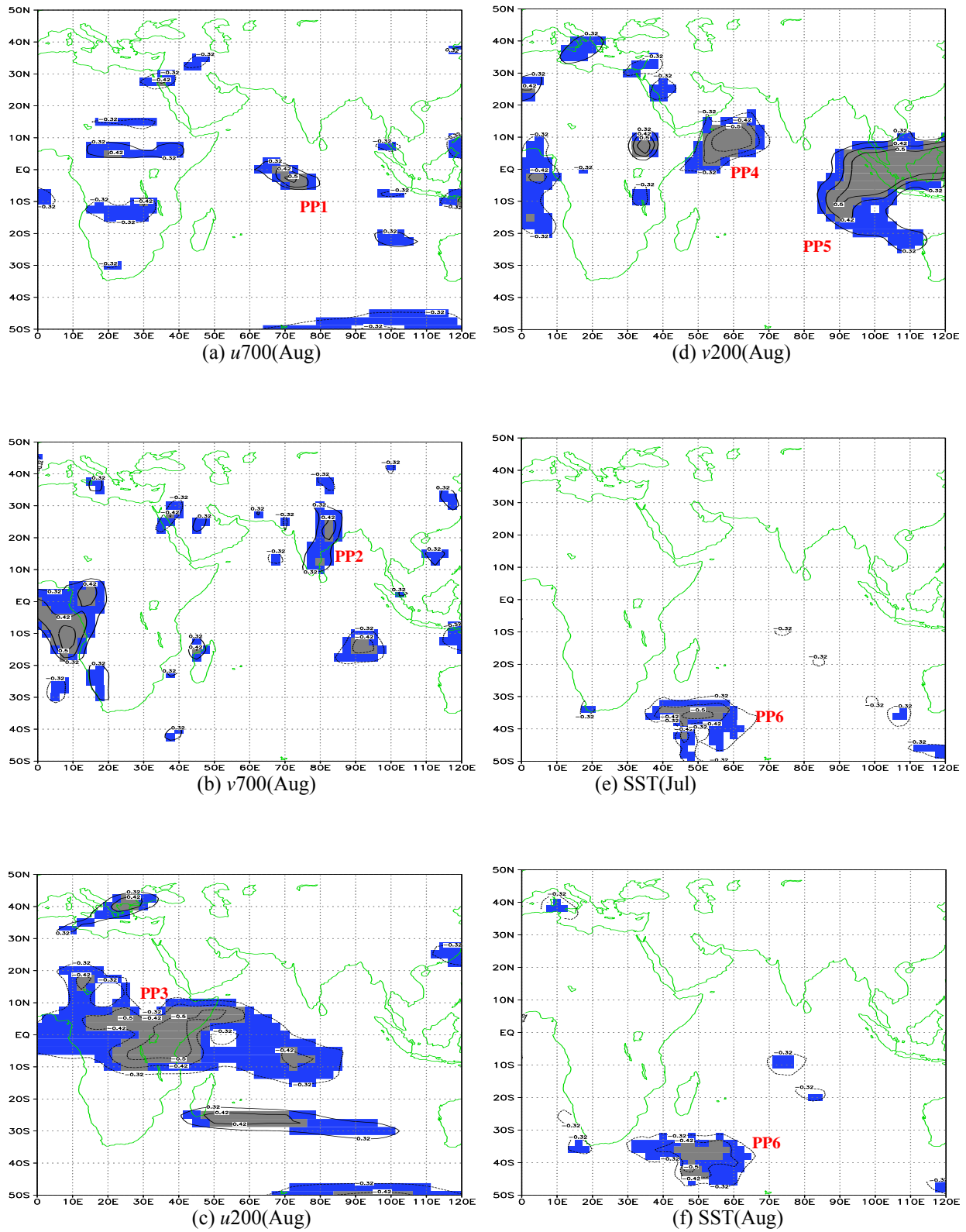
Climate Index	Month of Index											
	Sep (-1)	Oct (-1)	Nov (-1)	Dec (-1)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Nino1+2	0.28	0.22	0.19	0.13	0.19	0.12	0.09	0.14	0.07	-0.02	-0.05	-0.13
Nino3	0.24	0.22	0.22	0.19	0.24	0.17	0.11	-0.01	-0.10	-0.23	-0.21	-0.20
Nino 3.4	0.20	0.22	0.21	0.18	0.22	0.20	0.15	-0.02	-0.04	-0.23	-0.22	-0.25
Nino4	-0.07	0.05	0.07	0.02	0.05	0.12	0.03	-0.03	-0.01	-0.16	-0.19	-0.33
SOI	-0.10	0.08	-0.03	-0.17	-0.24	-0.03	-0.04	-0.03	0.11	0.10	-0.05	0.33
QBO	-0.10	-0.14	-0.21	-0.27	-0.29	-0.27	-0.26	-0.19	-0.12	-0.07	0.00	0.08
Atlantic Hurricane frequency	0.02	-0.02	-0.21	0.15	0.29	-	-	0.25	0.33	*-0.49	-0.11	0.02

(based on 30 year data of 1971-2000; \*: significant at 1% level)

It is well known that El Niño has teleconnective influence on various atmospheric phenomena over different parts of the globe. As such, the ENSO parameters were first taken for analysis. The monthly values of SST anomalies over the four Niño regions of the equatorial Pacific (Niño1+2, Niño3, Niño3.4, Niño 4) and the SOI for all the months from September of the previous year to August of the year for which prediction is to be made during the period 1971-2000 are analysed in relation to the number of CD days over NIO during OND, 1971-2000. Further, other readily available indices, QBO and Atlantic hurricane frequency, are also considered. Table 2 presents the CCs obtained between these indices and the number of CD days during OND over NIO for the period 1971-2000. It may be noted from the Table 2 that there are no significant relationships between any of the above parameters and the number of CD days over NIO during OND excepting the Atlantic hurricane frequency of June which is negatively correlated (significant at 1% level). But the relationship is not sustained and hence, this parameter also is not taken for further analysis. As such, it is observed that there are no sustained significant CCs between various climate indices and the number of CD days over NIO during the cyclone season (OND) during the recent three decades and hence none of them are considered as potential predictors for prediction of seasonal cyclonic activity.

4.3. Next, large scale circulation features such as zonal and meridional winds at lower (700 hPa) and upper tropospheric (200 hPa) levels representing the atmospheric circulation features and SST over different regions representing forcings from oceans, in the antecedent mode, that could be associated with TC genesis are taken for analysis. These features are considered on the presumption that they reflect the background flow in the antecedent mode which are going to shift to a conducive environment later for the formation of TCs over the region of interest through spatial and temporal scale interactions and hence are likely to have physical links with the predictand.

First, the areas over which these circulation features have association with seasonal cyclonic activity over the NIO during the month(s) of July / August are identified by computing the CCs between each of these features over various grids of the globe and the number of CD days over the NIO during OND. With the aim to have an outlook in the preceding month just before the start of season, we focus our attention on the flow patterns in the month of August (/July). Figs. 2(a-f) depict spatial distribution of significant CCs obtained for seasonal cyclonic activity over NIO with the various circulation features during July / August. From the CCs obtained, the



**Figs. 2 (a-f).** Spatial distribution of correlation coefficients between number of CD days and zonal/meridional wind and SST based on 30 year data of 1971-2000. Grid boxes with CCs significant at 1%(5%) level are shaded grey (blue)

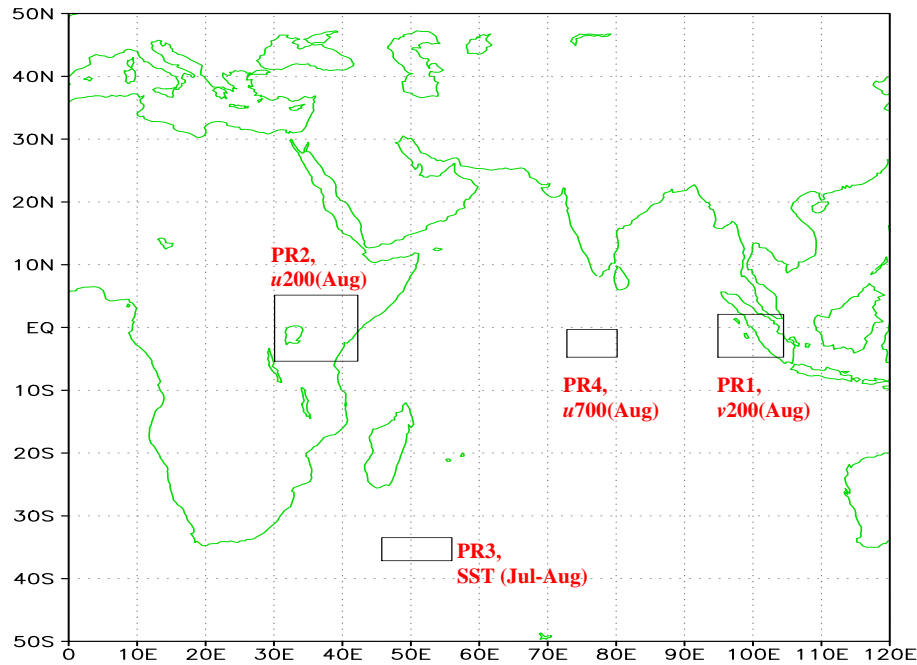


Fig. 3. Regions over which the four chosen predictors PR1, PR2, PR3 and PR4 are selected

TABLE 3(a)

Statistical parameters of the six potential predictors and their CCs with the number of CD days over NIO during OND for the 30 year period of 1971-2000

Parameter	Mean	Std. deviation	CC
PP1, $u_{700}$ (Aug)	4.74 m/s	1.84 m/s	0.40*
PP2, $v_{700}$ (Aug)	-0.36 m/s	0.8 m/s	0.50**
PP3, $u_{200}$ (Aug)	-7.98 m/s	2.61 m/s	-0.59**
PP4, $v_{1200}$ (Aug)	1.22 m/s	1.08 m/s	-0.56**
PP5, $v_{200}$ (Aug)	-7.31 m/s	1.35 m/s	0.65**
PP6, SST (Jul-Aug)	16.68°C	0.29 °C	-0.57**

\*\* : significant at 1% level; \* : significant at 5% level

TABLE 3(b)

Inter correlation coefficients amongst the six potential predictors

	PP1, $u_{700}$ (Aug)	PP2, $v_{700}$ (Aug)	PP3, $u_{200}$ (Aug)	PP4, $v_{1200}$ (Aug)	PP5, $v_{200}$ (Aug)	PP6, SST (Jul-Aug)
PP1, $u_{700}$ (Aug)	1	0.178	-0.1417	-0.3879*	0.3945*	-0.1182
PP2, $v_{700}$ (Aug)		1	-0.3309	-0.3694*	0.3946*	-0.4650**
PP3, $u_{200}$ (Aug)			1	0.5161**	-0.5033**	0.3195
PP4, $v_{1200}$ (Aug)				1	-0.6083**	0.5411**
PP5, $v_{200}$ (Aug)					1	-0.5217**
PP6, SST (Jul-Aug)						1

\*\* : significant at 1% level; \* : significant at 5% level (based on 30 year data of 1971-2000)

**TABLE 4 (a)****Test of significance for difference in mean values of predictors during extreme years of CD days**

Extreme class of CD days*	Sample size ( <i>n</i> )	Predictor	Mean value of predictor	<i>t</i> value
CD days $\geq$ 28 days	9	PR1 : $v200(\text{Aug})$	-6.22 m/s	4.22
CD days $\leq$ 12 days	8		-8.32m/s	
CD days $\geq$ 28 days	9	PR2 : $u200(\text{Aug})$	-9.54 m/s	4.62
CD days $\leq$ 12 days	8		-5.47 m/s	
CD days $\geq$ 28 days	9	PR3 : SST(Jul-Aug)	16.39 °C	4.05
CD days $\leq$ 12 days	8		16.8 °C	
CD days $\geq$ 28 days	9	PR4 : $u700(\text{Aug})$	5.85 m/s	2.33
CD days $\leq$ 12 days	8		3.85 m/s	

\*: Years with number of CD days deviating by one standard deviation or more

**TABLE 4 (b)****Conditional means of number of CD days over NIO during OND for the development period (1971-2000) for various intervals of the four selected predictors**

S. No.	Predictor*	Interval**	Conditional Mean (days)	No. of years ( <i>n</i> )
1.	PR1 : $v200$ (Aug)	< -8.5 m/s	12.00	6
		-8.5 to -7.0	17.67	12
		-7.0 to -5.5	24.10	10
		> -5.5	31.00	2
2.	PR2 : $u200$ (Aug)	< -10.5 m/s	27.7	3
		-10.5 to -8.0	21.3	15
		-8.0 to -5.5	22.5	4
		> -5.5	11.9	8
3.	PR3 : SST (Jul-Aug)	< 16.4 °C	28.5	6
		16.4 to 16.7	21.3	10
		16.7 to 17.1	14.4	14
		> 17.1	---	0
4.	PR4: $u700(\text{Aug})$	<3.0 m/s	15.7	7
		3.0-5.0	17.9	9
		5.0-7.0	21.8	11
		>7.0	25.3	3

\* PR1 [ $v200$  (Aug)], PR2 [ $u200$  (Aug)] and PR4 [ $u700(\text{Aug})$ ] in m/sec; PR3 : SST (Jul-Aug) in °C;

\*\*: classifications based on one SD of PR\*

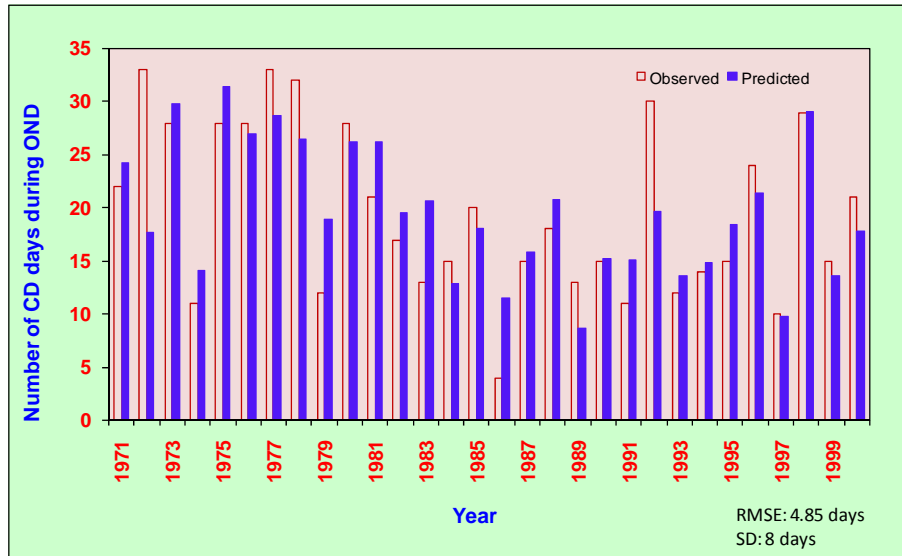


Fig. 4. Predicted number of CD days based on multiple regression models and the observed number of CD days during OND for the model development period of 1971-2000

following 6 parameters which are significantly associated with cyclonic activity over the NIO during OND are identified as potential predictors (PP\*):

Predictor	Details	Symbol
PP1	Areal average of zonal wind at 700 hPa level over 73-80° E & 5° S to Equator during August	$u_{700}$ (Aug)
PP2	Areal average of meridional wind at 700 hPa level over 80-85° E & 20-27° N during August	$v_{700}$ (Aug)
PP3	Areal average of zonal wind at 200 hPa level over 30-42° E & 7° S to 5° N during August	$u_{200}$ (Aug)
PP4	Areal average of meridional wind at 200 hPa level over 55-65° E & 8 to 15° N during August	$v_{1200}$ (Aug)
PP5	Areal average of meridional wind at 200 hPa level over 95-105° E & 5° S to 2° N during August	$v_{2200}$ (Aug)
PP6	Areal average of SST over 46-56° E & 38-34° S during July & August	SST(Jul-Aug)

Statistics of these potential predictors and significant CCs between these parameters and the number of CD days over NIO during OND are given in Table 3(a).

4.4. Generally, meteorological parameters are often inter-correlated [Table 3(b)] and hence, selection of parameters as predictors has to be done with care to avoid redundancy. Hence, out of the six potential predictors, an optimum number of four are selected by applying screening regression techniques (Wilks, 1995). In this iterative selection procedure, the parameter having the highest correlation with the predictand [Table 3(a)] is selected (PP5) as the first predictor (PR1). 41.9% of the predictand's variance is explained by this parameter. Then, based on the additional variance explained by adding the other parameters, other predictors are selected successively (PP3 (9.63%), PP6 (5.96%) & PP1 (4.4%)) as predictors PR2, PR3 and PR4. The additional variance explained by the remaining two parameters (PP2 & PP4) is less than 2% and hence are not selected. Thus, the following are the four parameters (PR1, PR2, PR3 and PR4) finally chosen as predictors for further analysis:

Predictors	Details	Symbol
PR1	Meridional wind at 200 hPa level over 95-105° E & 5° S to 2° N during Aug	( $v_{200}$ , Aug)
PR2	Zonal wind at 200 hPa level over 30-42° E & 7° S to 5° N during Aug	( $u_{200}$ , Aug)
PR3	SST over 46-56° E & 38-34° S during July & Aug	(SST, Jul-Aug)
PR4	Zonal wind at 700 hPa level over 73-80° E & 5° S to Equator during Aug	( $u_{700}$ , Aug)





**Figs. 5(a&b).** Validation of predictions based on multiple regression models (a) during the development period (1971-2000) using Jackknife method and (b) during the independent test period (2001-2009)

The location maps of the selected parameters are shown in Fig. 3.

4.5. Next, to test the skill of the predictors, the difference between the mean values of the predictors during two extreme categories of years with number of CD days deviating by (i) S.D from mean, *i.e.*, CD days  $\geq 28$  days and (ii) CD days  $\leq 12$  days, is tested for statistical significance using *t* test. The obtained *t* values for all the four predictors are statistically significant [Table 4 (a)]. Then, the conditional mean analysis is carried out by classifying each predictor based on one S.D. The mean values of CD days for various class intervals of each predictor were computed for the

developmental period (1971-2000) and the results are shown in Table 4(b). Prediction based on each predictor using the above classifications [Table 4(b)] is then determined for the subsequent test period of 2001-2009 and these are presented in Table 5. The predictions are qualitatively categorized as per the categorization of cyclonic activity defined in Sec.3.1. Finally, a single qualitative prediction for seasonal cyclonic activity is worked out for each year of the test period (2001-2009). These predictions and their validation are given in Table 5. It can be seen from Table 5 that predictions for cyclonic activity during the years 2001 to 2009, based on conditional mean analysis, agree fairly well with the observed activity except for the year 2006.

**TABLE 5**  
**Independent predictions for the test period (2001-2009) based on conditional mean analysis**

Year	Predictor *	Predicted number of CD days	Qualitative prediction	Expected cyclonic activity	Observed cyclonic activity
2001	PR1: v200 (Aug)	12	Below normal		7 days;
	PR2: u200 (Aug)	23	Normal	Below normal to Normal	Below Normal
	PR3: SST (Jul-Aug)	14	Below normal		
	PR4: u700(Aug)	18	Normal		
2002	PR1: v200 (Aug)	18	Normal	Below normal to Normal	16 days;
	PR2: u200 (Aug)	23	Normal		Normal
	PR3: SST (Jul-Aug)	14	Below normal		
	PR4: u700(Aug)	22	Normal		
2003	PR1: v200 (Aug)	18	Normal	Normal	17 days;
	PR2: u200 (Aug)	21	Normal		Normal
	PR3: SST (Jul-Aug)	21	Normal		
	PR4: u700(Aug)	18	Normal		
2004	PR1: v200 (Aug)	12	Below normal	Below normal to Normal	16 days;
	PR2: u200 (Aug)	23	Normal		Normal
	PR3: SST (Jul-Aug)	14	Below normal		
	PR4: u700(Aug)	16	Normal		
2005	PR1: v200 (Aug)	24	Normal	Normal	24 days;
	PR2: u200 (Aug)	23	Normal		Normal
	PR3: SST (Jul-Aug)	21	Normal		
	PR4: u700(Aug)	18	Normal		
2006	PR1: v200 (Aug)	18	Normal	Normal	2 days;
	PR2: u200 (Aug)	23	Normal		Subdued
	PR3: SST (Jul-Aug)	-	-		
	PR4: u700(Aug)	18	Normal		
2007	PR1: v200 (Aug)	18	Normal	Normal	13 days;
	PR2: u200 (Aug)	28	Above normal		Below normal
	PR3: SST (Jul-Aug)	14	Below normal		
	PR4: u700(Aug)	18	Normal		
2008	PR1: v200 (Aug)	18	Normal	Normal	18 days;
	PR2: u200 (Aug)	21	Normal		Normal
	PR3: SST (Jul-Aug)	21	Normal		
	PR4: u700(Aug)	18	Normal		
2009	PR1: v200 (Aug)	18	Normal	Below normal	10 days;
	PR2: u200 (Aug)	12	Below normal		Subdued
	PR3: SST (Jul-Aug)	16	Below normal		
	PR4: u700(Aug)	15	Below normal		

\* PR1 [v200 (Aug)], PR2 [u200 (Aug)] and PR4 [u700(Aug)] in m/sec;  
 PR3 : SST (Jul-Aug) in °C

4.6. Next, a multiple regression equation for the number of CD days during OND is developed with the same four chosen predictors. The equation is

$$\text{No. of CD days} = 157.747 + 1.277*v200 \text{ (Aug)} \\ -1.081*u200 \text{ (Aug)} + 8.524*SST \\ \text{(Jul, Aug)} + 1.001*u700 \text{ (Aug)}$$

Fig. 4 depicts a plot of number of CD days predicted by the equation for the model developmental period of 1971-2000 as well as that observed. A multiple CC of 0.787 is obtained. The root mean square error (RSME) of the predicted number of CD days is 4.85 days ( $\approx 5$  days) and is less than the standard deviation of the number of

CD days (8 days) which is indicative of good forecast skill of the model. The model bias error is quite small at 0.36 days.

The model is cross validated using Jackknife technique (Chan *et al.*, 1998) to test its performance. According to this method, prediction for each year within the development period of 1971-2000 is tested by developing separate regression equations for each year using the data of 29 years of the sample excluding the year under consideration and the results are shown in Fig. 5(a). Following are the salient features of the performance of the model during the development period:

(i) Out of the 30 years, the error in prediction is within the RMS error of the model (5 days) during 23 years.

(ii) During the 23 years for which the prediction error is within the RMS error, there have been six occasions when the actual number of CD days deviated from the normal by one standard deviation (8 days) or more (1974, 77, 78, 91, 97 and 98). Thus the model shows some promise of predicting extreme cases of cyclonic activity over NIO.

(iii) But the model does not predict the extreme cyclonic activity of 33 & 30 days during 1972 & 92 and 12, 13 & 4 during 1979, 83 and 86 correctly. Predictions for these years, 1972, 79, 83, 86 and 92 are beyond the standard error of the model.

Predictions for an independent sample of nine year data (2001 to 2009) based on the multiple regression model are shown in Fig. 5(b). It can be seen that there is an over predictions for the year 2006 when the actual cyclonic activity over the NIO was somewhat subdued.

Precise quantitative prediction of CA is limited by the fact that the standard error of the multiple regression model is 5 days which is greater than half the standard deviation of the number of CD days over NIO during OND. However, taking into account the ability of the model in predicting about 70% of inter annual variance and also indications of some of the extreme cases of seasonal cyclonic activity the model can be considered to give an outlook on CA at least in qualitative terms. Hence, instead of taking the single value predicted by the equation, a range of values is considered and the prediction, qualitatively categorised in the same manner as is done for conditional mean analysis.

4.7. Based on the inputs from both the schemes of prediction, an overall outlook on the Seasonal cyclonic activity during October to December over the North Indian Ocean could be inferred in the month of September.

## 5. Summary

Despite the known limitations of the statistical models such as secular variations of correlation, choice of optimum number of predictors, test period etc., statistical models are quite commonly used to get a likelihood scenario of the future weather events until more precise and accurate dynamical models are developed to predict the same.

Over the North Indian Ocean, the Northeast monsoon season of October to December is the primary cyclone season, the mean number of days of cyclonic

activity being 20 days. In the present study, a statistical prediction for the seasonal cyclonic activity over the North Indian Ocean during the cyclone season of October to December is attempted by identifying predictors from amongst well known climate indices and regional circulation features using recent 30 year data of 1971-2000. Potential predictors are identified using correlation analysis and an optimum number of predictors are chosen using screening regression technique.

The following are the four predictors selected:

- (i) PR1 : Meridional wind at 200 hPa level over 95-105° E & 5° S to 2° N during August ( $v_{200}$ , Aug)
- (ii) PR2 : Zonal wind at 200 hPa level over 30-42° E & 7° S to 5° N during August ( $u_{200}$ , Aug)
- (iii) PR3 : SST over 46-56° E & 38-34° S during July & August (SST, Jul-Aug)
- (iv) PR4 : Zonal wind at 700 hPa level over 73-80° E & 5° S to Equator during August ( $u_{700}$ , Aug)

Two prediction schemes are developed –

- (i) A qualitative prediction for number of Cyclonic Disturbance days based on conditional mean analysis of the number of CD days for different intervals of each predictor and
- (ii) A quantitative prediction based on multiple regression technique using the same predictors. The regression equation is

$$\begin{aligned} \text{No. of CD days} = & 157.747 + 1.277*v_{200} \text{ (Aug)} \\ & - 1.081*u_{200} \text{ (Aug)} - 8.524* \\ & \text{SST (Jul, Aug)} + 1.001*u_{700} \\ & \text{(Aug)} \end{aligned}$$

The multiple regression model accounts for 70% of the inter annual variance. The root mean square error of the model prediction is 5 days and the bias error is 0.36 days. The model is cross validated by Jackknife method for the period 1971-2000.

Independent predictions for the subsequent years 2001-2009 are prepared using both the schemes. The results indicate that a qualitative prediction based on inputs from both the methods is possible. The prediction model, is able to give an idea on the extent of cyclonic activity over the NIO even though it has some limitations like predicting the cyclonic activity of some of the

extreme years. Inclusion of other atmospheric and ocean circulation anomalies (*e.g.*, Madden-Julian Oscillation) may improve the model though non-stationary correlations could again limit such models.

#### *Acknowledgements*

We thank the Deputy Director General of Meteorology, Regional Meteorological Centre, Chennai for his guidance, encouragement and support throughout this study.

#### **References**

- Chan, J. C. L., Shi, J. E and Lam, C. M., 1998, "Seasonal Forecasting of Tropical Cyclone activity over the Western North Pacific and the South China Sea", *Weather Forecasting*, **13**, 997-1004.
- Chan, J. C. L., Shi, J. E. and Liu, K. S., 2001, "Improvements in the seasonal forecasting of Tropical Cyclone activity over the Western North Pacific", *Weather Forecasting*, **16**, 491-498.
- Gray, W. M., 1984, "Atlantic seasonal hurricane frequency. Part I: El Nino and 30 mb QBO influences", *Mon. Wea. Rev.*, **112**, 1649-1668.
- Gray, W. M., 1984, "Atlantic seasonal hurricane frequency. Part II: Forecasting its variability", *Mon. Wea. Rev.*, **112**, 1669-1683.
- India Met. Dept., 2008, "CWCDSTAT - A Fortran based software for generation of statistics on cyclones and depressions over the North Indian Ocean", Met. Monograph No. Climatology 23/2008.
- Jury, M. R., Pathack, B. and Parker, B., 1999, "Climatic determinants and Statistical prediction of tropical cyclone days in the southwest Indian Ocean", *Journal of Climate*, **12**, 1738-1746.
- Krishna Kumar, G., Rajeevan, M. and Gadgil, A. S., 2007, "Use of wind and temperature anomalies over the Indian Ocean in the estimation of cyclonic disturbance activity", *Vayu Mandal*, **33**, 1-4, 26-33.
- Landsea, C. W., Bell, G. D. and Gray, W. M., "The extremely active 1995 Atlantic hurricane season: Environmental conditions and verification of seasonal forecasts", *Mon. Wea. Rev.*, **126**, 1174-1191.
- Neumann, C. J. and Mandal, G. S., 1978, "Statistical prediction of tropical storm motion over the Bay of Bengal and Arabian Sea", *IJMHG*, **29**, 1&2, 487-500.
- Rajeevan, M. and Butala, P. P., 1990, "A preliminary study on the variability of post-monsoon tropical cyclone activity over the north Indian ocean", *Mausam*, **41**, 3, 409-414.
- Sikka, D. R. and Suryanarayana, R, 1972, "Forecasting the movement of tropical storms/depressions in the Indian region by computer oriented technique using climatology and persistence", *IJMG*, **23**, 1, 35-40.
- Singh O.P. and Khan T.M.A., 2001, "Tropical cyclone frequency in the north Indian ocean in relation to southern oscillation phenomenon", *Mausam*, **52**, 3, 511-514.
- Singh, S., Giri, R. K. and Attri, S. D., 2007, "The study of cyclonic disturbances over Indian seas during 1991-2004", *Mausam*, **58**, 1, 113-118.
- Wilks, D. S., 1995, 'Statistical methods in the Atmospheric sciences', Academic Press, New York.