MAUSAM, **65**, 3 (July 2014), 319-332

551.515.2 : 551.509.327 (267.64)

Statistical prediction of movement of cyclonic storms and depressions over Bay of Bengal through LOESS technique

RM. A. N. RAMANATHAN and Y. E. A. RAJ

India Meteorological Department, Chennai, India

(*Received 12 April 2012, Modified 8 May 2013*)

e mail : rmanramu@gmail.com

सार – प्रचालनात्मक मौसम पूर्वानुमानकर्ता के लिए चक्रवात के मार्ग का सटीक पूर्वानुमान करना हमेशा चुनौतीपूर्ण रहा है। इस अध्ययन में तीन सांख्यिकीय तकनीकों नामतः, एनालॉग, एनालॉग-सह-समाश्रयण और लॉजिकली बेटेड स्कैटरप्लॉट स्मूदिंग (LOESS) के उपयोग से चक्रवात के मार्ग का पूर्वानुमान लगाने का प्रयास किया गया है । बंगाल की खाड़ी में 1961-2008 के दौरान बने और आगे बढ़ने वाले चक्रवातीय विक्षोभों के मार्ग के आँकड़ों का उपयोग किया गया है। इन तीन तकनीकों के माध्यम से चक्रवात के मार्ग से पूर्वानुमान की सटीकता के स्तरों की तुलना करने के लिए सांख्यिकीय निदर्श तैयार किया गया और परिणामों की चर्चा की गई। यह देखा गया कि एनालॉग और एनालॉग-सह-समाश्रयण तकनीकों की तुलना में LOESS तकनीक द्वारा चक्रवात के मार्ग के पूर्वानुमान की 147 कि. मी. औसत त्रुटि सबसे कम रही। प्रतिवक्र (रीकर्व्ड) प्रणालियों के मामले में भी LOESS के माध्यम से प्राप्त पूर्वानुमान त्रुटि कम रही। एलॉग-ट्रैक और क्रॉस-ट्रैक घटकों के लिए हेडके स्किल स्कोर, पीचर्स स्किल स्कोर तथा प्रॉपोर्शन करेक्ट का आकलन किया गया जिससे पता चला कि एनालॉग और एनालॉग-सह-समाश्रयण तकनीकों की तुलना में LOESS की तकनीक की बेहतर सटीकता रही। अन्य स्किल स्कोर सूचकांकों की भी गणना की गई है और उसके परिणाम प्रस्तुत किए गए हैं।

ABSTRACT. Accurate cyclone track prediction has always been a challenge to the operational weather forecaster. An attempt has been made in this study for prediction of the cyclone track by employing three statistical techniques, *viz*., analogue, analogue-cum-regression and Locally weighted Scatterplot Smoothing (LOESS). Track data of cyclonic disturbances which formed and moved in the Bay of Bengal during the period 1961-2008 has been used. A statistical model has been developed for comparison of the accuracy levels of track prediction through these three techniques and results have been discussed. It has been observed that the average track forecast error of 147 km calculated by LOESS technique is minimum compared to those obtained from analogue and analogue-cum-regression techniques. In the case of recurved systems also, the forecast error obtained through LOESS is minimum. Heidke Skill Score, Peirce Skill Score and Proportion Correct have been calculated for Along-Track and Cross-Track components which indicate better accuracy and superiority of LOESS technique over the analogue and analogue-cum-regression techniques. Other skill score indices have also been computed and results presented.

Key words – LOESS, Analogue regression, Tropical cyclone, Along-Track error, Cross-Track error, Heidke skill score, Peirce skill score.

1. Introduction

Tropical cyclones (TC) are large scale synoptic systems which originate over oceans and can intensify into systems of tremendous destructive potential. The disastrous features such as rough seas, heavy rains, storm surge and flooding associated with the landfall of intense TCs are major threats to the human population and livestock in coastal areas and to the shipping ports. Accurate forecasting of the movement and intensity of TCs plays a vital role in disaster management. One of the major goals in TC research is to improve the accuracy levels of prediction of the track of TC and the intensity that a TC could attain. Operational weather forecasters who face the challenge of tracking a TC on a near real-time basis have the prime responsibility of providing accurate track prediction so as to safeguard the precious lives of human beings and livestock. The local governments of states/regions which are likely to be affected by the TC have to be kept warned of the impending adverse weather so that they can gear up and plan for disaster management and preparedness activities.

A number of forecasting techniques have been developed by the meteorological community using

TABLE 1

Classification of tropical cyclonic disturbances when located over the Indian Seas

Category	Maximum sustained winds	
	Knots	Kmph
Low	< 17	< 31
Depression(D)	17 – 27	31 – 51
Deep depression(DD)	28 – 33	52 – 62
Cyclonic Storm(CS)	34 – 47	63 – 87
Severe Cyclonic Storm(SCS)	48 – 63	88 – 117
Very Severe Cyclonic Storm(VSCS)	64 – 119	118 – 221
Super Cyclonic Storm(SUCS)	120 and above	222 and above

TABLE 2

Frequency of occurrence of cyclonic disturbances in the Bay of Bengal during northeast monsoon season (October, November and December)

Period	Bay of Bengal			
	D	CS	SCS	Total
1961 - 1995	52	28	61	141
1996 - 2008	16	10	11	37

D – Depression, CS – Cyclonic Storm, SCS – Severe Cyclonic Storm
(The number includes greater intensities also)

empirical, statistical and dynamical approaches which are diverse in nature. However, due to the complexity of the problem, no single technique has proved to have outstanding performance over the others. In addition to the improvement of numerical weather prediction models in predicting the motion of TCs, there is a need as well to improve the performance of empirical models (Sievers, *et al.*, 2000).

Despite the development of TC forecast models based on numerical prediction, the very basic climatology and persistence (CLIPER) model for the prediction of TC motion is still in use by the TC forecasting centres of the world. Few of the models continue to be in the process of development and further refinement. Some of the applications of such models, not all of which refer directly to the forecast process are to (*i*) provide a convenient frame of reference upon which the performance of more sophisticated models can be assessed, (*ii*) enable the assessment of "forecast difficulty", (*iii*) provide a convenient way to generate bogus TC tracks, (*iv*) provide a "first guess" forecast and (*v*) provide a reasonable forecast in portions of basins where deviations from climatology and persistence are small (Bessafi *et al.*, 2002).

The North Indian Ocean (NIO) comprises of two basins, *viz*., Arabian Sea (AS) and Bay of Bengal (BoB). The pre-monsoon (March, April and May) and post-monsoon seasons (October, November and December) are the seasons with high incidence of severe cyclonic storms (SCS) in NIO. The total number of SCS formed over NIO during the period 1961-2008 is 144, out of which SCS formed during pre-monsoon and post-monsoon alone accounts for 84% (Cyclone eAtlas, 2008). Season wise, the post-monsoon season is most prone to occurrence of such SCS, accounting for nearly half the annual frequency. The basin-wise break-up of percentage frequencies of SCS is 65 and 42 for BOB and AS respectively during the period 1961 to 2008.

Studies on prediction of movement of TCs over Indian seas have been attempted by several researchers. Surendra Kumar and Kanti Prasad (1973), Bansal and Datta (1974), Datta and Gupta (1975), Sikka (1975 and 2006), Ramanathan and Bansal (1977), Neumann and

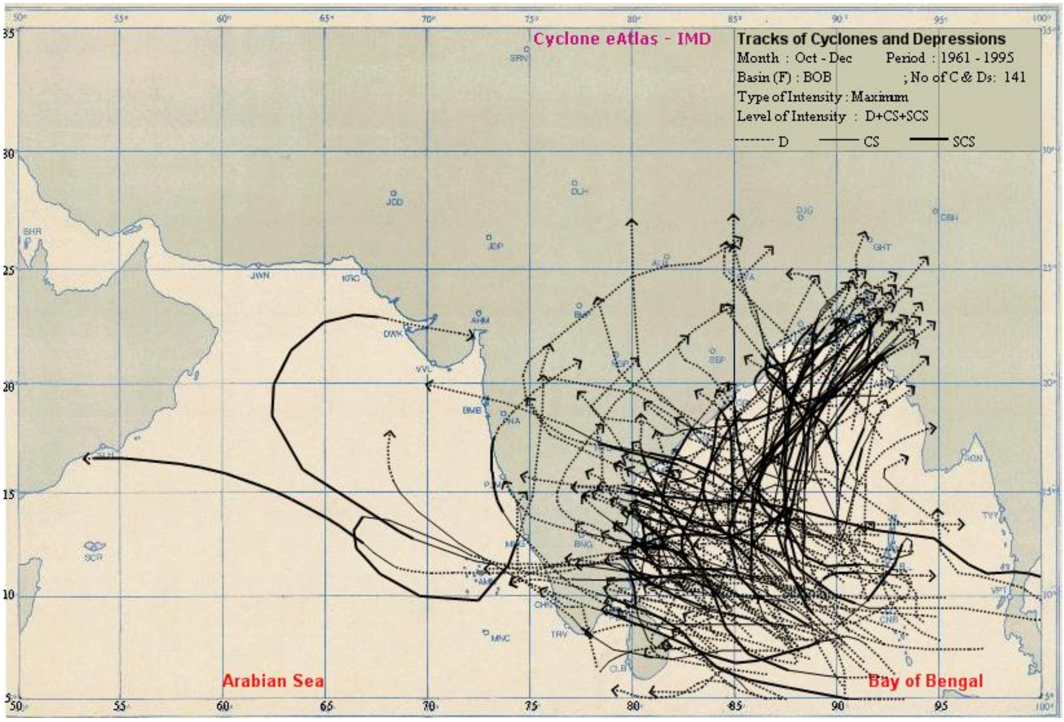

Fig. 1. Tracks of cyclones and depressions which formed over the Bay of Bengal during the period 1961 to 1995

Mandal(1978), Sen Sharma (1983), Raj *et al.* (1991), Bessafi *et al*. (2002) and Bandyopadhyay and Singh (2006) reviewed the research work in the field of TC forecasting in the Indian scenario where analogue, statistical and dynamical methods are used to a large extent in combinations.

The objective of this study is to develop a statistical model for prediction of the 24 hours future position of cyclonic disturbances (CDs) which include cyclonic storms and depressions. In this study we have considered forecasting the tracks of CDs of BoB during northeast monsoon (October, November and December) season by adopting methodologies which are somewhat different from what has been attempted in other similar studies. A technique known as Locally Weighted Scatter plot Smoothing, abbreviated as LOWESS or LOESS (Cleveland and Devling, 1988 and Montgomery *et al*., 2003) based on the principle of giving more weightage to the nearest point and less to the farthest point has been utilised. In literature, the terms LOWESS and LOESS are used inter-changeably for locally weighted linear regression analysis though they can be differentiated by the degree of the polynomial used. LOWESS uses a linear polynomial while LOESS uses a quadratic polynomial. In this study, uniformly we have used only the LOESS technique. For comparison with LOESS, two other techniques, *viz*., analogue (ANL) and analogue-cum-regression (ACR) have also been utilised. A statistical model has been developed for comparison of the accuracy levels of track prediction through these three techniques and results have been presented. As per literature survey by the authors, LOESS technique is not used by any national meteorological service in the world for the prediction of movement of TCs. For the cyclone track prediction in NIO also this technique has not yet been tried and hence this study has been attempted.

2. Data

The nomenclatures used for defining the various categories of CDs based on estimated maximum sustained wind speeds over the Indian Seas (IMD, 2003) is given in Table 1. The frequency of occurrence of CDs in BoB during the northeast monsoon season is given in Table 2.

The data set used in this study has been obtained from the electronic version of Cyclone Track Atlas published by India Meteorological Department (IMD, Cyclone eAtlas, 2008). The data set consists of the positions (Lat./Long.) at 0300 UTC of all tracks of CDs for the BoB for the period 1961-2008. For other lat/long positions, the exact UTC at which the CD was in that coordinate is not known. Hence from the track, 1500 UTC coordinates cannot be directly/readily obtained.

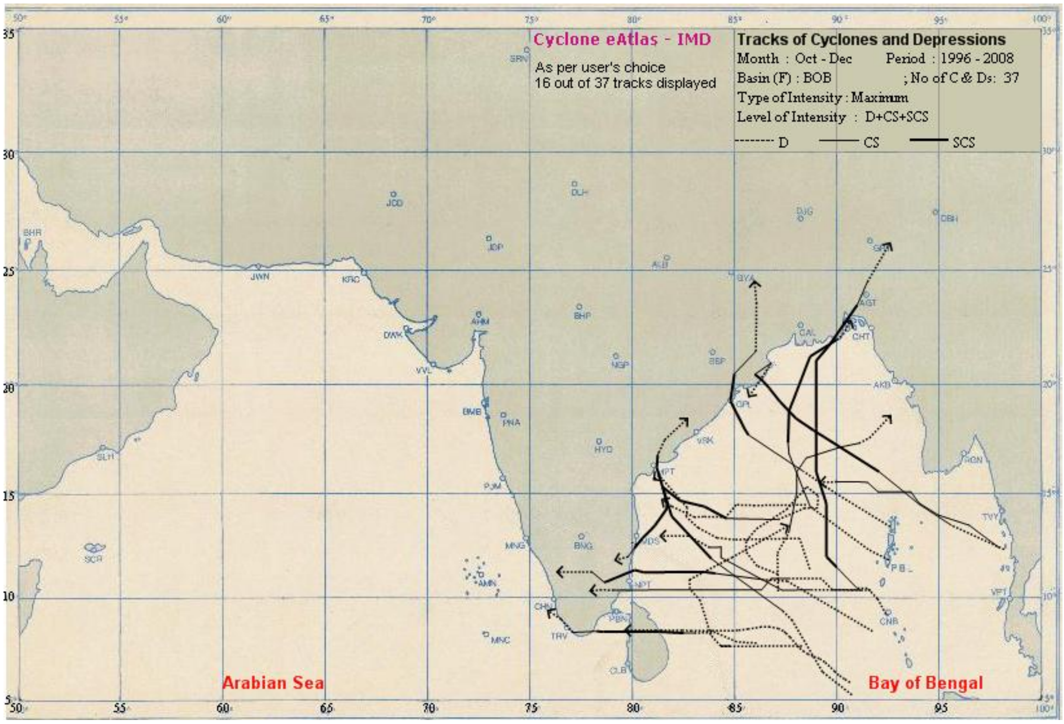

Fig. 2. Tracks of 16 CDs which formed and moved over BoB during 1996-2008, the test period

In this study, in ACR and LOESS technique four predictors are used. So, in order to get more analogue records, from the initial 0300 UTC position, the 12 hour interval position, *i.e.*, 1500 UTC position is retrieved by using an interpolation by curve fitting method as in the Hurran model (Hope and Neumann, 1970). Using the 03 and 1500 UTC positions thus obtained, further analysis is done. During the period 1961 to 2008, the number of CDs, which formed over BoB during northeast monsoon season (October, November and December), were 178 and these have been considered in this study for developing the database. Over Indian seas the CDs were tracked by IMD with the help of imageries transmitted by the weather satellites. This crucial development took place in the early 1960's and provided well-defined and accurate initial vortex positions, which are important for better track prediction (Mohapatra, *et al.*, 2013). Hence the data set of the period 1961 to 2008 which is the satellite era for monitoring and tracking CDs has been considered.

The three techniques used for the study are based on analogues taken from the historic data. The methodology followed by Bessafi (2002) has been adopted. The test period is 1996-2008. The period 1961-1995 is selected as the historic data set (*i.e.*, the development period) to search for and pick up analogues for the CDs of the year 1996. Similarly 1961-1996 is the historic data set for selecting analogues for the CDs in the year 1997 and so on for every year which increases the number of years of data in the historic data set. It has been ensured that the test period 1996-2008 has enough sample size of at least 80 individual instances of forecast positions to derive a mean forecast error and to calculate other skill scores.

The tracks of CDs during the period of study are displayed in Fig. 1 (1961 to 1995) and Fig. 2 (1996 to 2008). The tracks have been shown here for two different periods because individual forecast errors have been calculated for the tracks of CDs which occurred during 1996-2008.

3. Methodology

The techniques utilized in this study are described briefly in the following sections.

3.1. *Analogue method*

The design of the ANL model is based on the principle that classes or families of TCs exist which tend to follow similar tracks. ANL sample is a collection of CDs in the period of study which had similar track movement as that of the current CD (for which track has to be predicted). A few parameters defined as ANL parameters are identified which denote the characteristics of the collection of CDs called families of CDs. ANL parameters which have been used in this study in the selection of ANL sample include (*i*) period of formation, (*ii*) position of formation (*iii*) direction of past movement and (*iv*) speed of movement. The CDs, which met the above criteria, are then translated to a common origin and rotated to a common heading and then the mean of the future position is calculated (Anthes, 1982). The ANL method used is the same as followed in Hope & Neumann (1970), Datta & Gupta (1975) and Bessafi (2002).

3.1.1. *Definition of analogue parameters*

In this study, current CD date, location, direction, basin (BoB) and season are the primary factors used in the analogue model process. The various ANL parameters which have been defined and used to select ANL (Raj, 2010) CDs for a given CD are described below :

(*i*) *The basin of the given CD* - If the current CD (for which track has to be predicted) lies over BoB, positions of earlier CDs of BoB (excluding those over Land or AS) could be analogues.

(*ii*) *The day of occurrence* - Dates of CD positions which are within ± 30 days of the day of occurrence of the current CD could be considered. For example, if the date of the current position of the CD (whose movement is to be forecast) is 20th November, positions of CDs whose dates lie between 20th October and 20th December could be chosen. Here, the choice of ± 30 days taken in this study is obviously a bit arbitrary and could be increased or decreased according to the requirement.

(*iii*) *The position of the CD* - If the current position of the CD is (λ,φ), where "λ" is the longitude and "φ" is the latitude, all positions of CDs which occurred in the square (λ ± 4° E, φ ± 4° N) could be considered as analogues. Here also, 4° range could be varied.

(*iv*) *Direction of movement* - If the direction of movement of the CD for the past 24 hours is d, then we can consider all the CDs with past movement of d ± 30° as analogous. Here again, the increment of 30° could be varied.

The speed of the movement of the CD could be another potential parameter but has not been included/ considered in the present study.

The above ANL parameters using the so called Jack-knife method (Wilks, 1995) are utilised for track prediction.

The differences of latitude (Δφ) and longitude (Δλ), which are the displacements, *viz*.,

$$\Delta\varphi = \varphi'_{24} - \varphi'_0 \; ; \Delta\lambda = \lambda'_{24} - \lambda'_0$$

are calculated for each selected analogue where $\varphi'_{24}, \varphi'_0$ are the 24 hours and present latitude positions respectively. Similarly, $\lambda'_{24}, \lambda'_0$ are the 24 hours and present longitude positions respectively. The mean values of $\Delta\varphi$ and $\Delta\lambda$, *viz*., $\overline{\Delta\varphi}$ and $\overline{\Delta\lambda}$ of all selected analogues have to be added with the current position of CDs to be predicted, *viz*.,

$$\varphi_{24} = \varphi_0 + \overline{\Delta\varphi} \; ; \lambda_{24} = \lambda_0 + \overline{\Delta\lambda}$$

where, φ_{24}, φ_0 are respectively the future and current positions of latitude. Similarly λ_{24}, λ_0 are respectively future and current positions of longitude.

3.2. *Analogue-cum-multiple regression*

Next, in addition to considering the mean position of the analogues as detailed in the previous section, the multiple regression technique based on the same ANL data set called as analogue-cum-multiple regression (ACR) is used to predict the movement of CD. The regression equations used for ACR have the following format:

$$\lambda_{24} = a_1 + b_1\lambda_0 + c_1\varphi_0 + d_1\lambda_{-24} + e_1\varphi_{-24} \quad (1)$$

$$\varphi_{24} = a_2 + b_2\lambda_0 + c_2\varphi_0 + d_2\varphi_{-24} + e_2\varphi_{-24} \quad (2)$$

where, λ_{24}, λ_0 and λ_{-24}, are the 24 hours forecast positions, present and past 24 hours longitude positions respectively. Similarly, φ_{24}, φ_0 and φ_{-24} are the 24 hours forecast positions, present and past 24 hours latitude positions respectively. The coefficients a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2, e_1 and e_2 are the regression coefficients which are to be estimated through the method of least squares (Kendall and Stuart, 1968). The data set consists of positions of every 12 hour interval.

It may be noted that if the *N* consecutive 12 hours positions of a given CD are available then they are also included as the *N* - 4 number of data sets for obtaining the 24 hours forecast position. For example, consider a CD for which 6 (*i.e.*, *N*) consecutive 12 hour positions are available with Long./Lat. positions (λ_i, φ_i), where i = 1 to 6 with i = 1 as initial position of the CD. In this study, once the CD is 24 hours old then only forecast can be generated as the regression equation uses present and past 24 hours Long./Lat. positions as dependent variables. The first data set comprises of the three positions, *viz*., (λ_5, φ_5), (λ_3, φ_3) and (λ_1, φ_1) which are the future known 24 hours position, current and previous 24 hours position respectively taken from the historic data set considered for analysis. Similarly the three positions, *viz*., (λ_6, φ_6), (λ_4, φ_4) and (λ_2, φ_2) comprise the second data set. So, for 6 consecutive 12 hours positions, we get 2 (6-4) data sets.

In the conventional multiple regression models, once the regression equation is developed, the predictions are made using the developed regression equation only regardless of the data for the track of the TC to be predicted. In contrast, the ACR is a memory-based method that performs a regression using the database corresponding to the selected analogues.

3.3. *LOESS technique vis-à-vis regression technique*

Since the objective of the present study is to utilise the LOESS technique to statistically predict future position of CD's, a brief background theory of the same is provided below.

3.3.1. *LOESS with one predictor*

The LOESS estimation technique was independently introduced in several different fields in the late nineteenth and early twentieth centuries. LOESS technique is a method of fitting a regression equation locally. Under the theory of LOESS, suppose x_0 is an observation of x, we consider the interval (x_0 - a, x_0 + a) where a is suitably chosen so as to include the points of x that lie in the local neighbourhood of x_0. Suppose there are m such points in the data set, which is less than 'n', the total number of points. Now we build up a linear regression model based on these m points using method of weighted least squares. A weighted function is defined corresponding to each observation x, where the weight varies with the distance of x from x_0 such that more weight is assigned to closer points. The default weight function used in the many statistical software's is the tri-cube weighting function, presented as

$$W(t) = [(1-t^3)^3 \; for \; 0 \leq t < 1] \quad (3)$$

$$W(t) = 0 \quad \text{elsewhere; and} \quad t = \left[\frac{|x_i - x_0|}{\Delta(x_0)}\right]$$

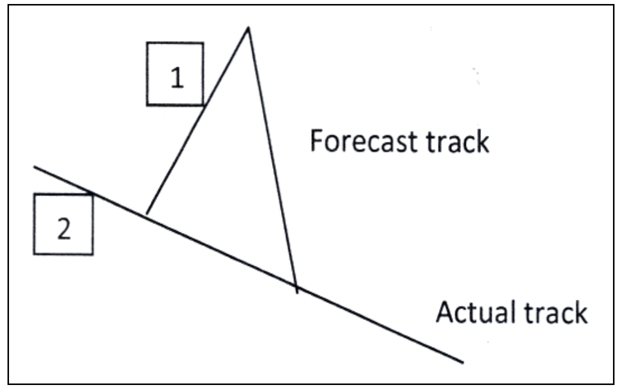

Fig. 3. Pictorial depiction of Cross-Track (1) and Along-Track (2) component

where, x_0 is the observation for which prediction is to be made, and $\Delta(x_0)$ is the distance of the farthest point in the neighbourhood of x_0. The local approximation can be fitted by locally weighted least squares. Coefficients estimates are then chosen to minimize 'S' where

$$S = \sum_{i=1}^{m} w(t)[y_i - (ax_i + b)]^2 \quad (4)$$

$$\frac{\partial S}{\partial a} = 0 \text{ and } \frac{\partial S}{\partial b} = 0 \quad (5)$$

By solving the equation (5) we obtain a and b. The regression equation y = ax + b can now be used for estimating the value of y, given $x = x_0$.

As x_0 changes, the corresponding neighbourhood also changes. The value t and the weight function $W(t)$ are different and hence the regression coefficients a and b keep changing for varying positions of x_0. This concept is unique to LOESS whereas in contrast, in the case of simple linear regression, a and b once computed, remain fixed.

3.3.2. *LOESS with more than one predictor*

In the single predictor case, the distance measure between two points is just the difference between them. When the number of predictors is greater than one, the distance measure used is generally the Euclidean distance or another such metric called Mahalanobis distance. In the case of LOESS technique used here, the number of predictors being more than one, the distance measure used is the 'Euclidean distance'.

In this study, the four predictors as in the case of ACR mentioned in equation (1) and (2) have been considered. Selecting the neighbourhood is very important to get better results using LOESS technique and

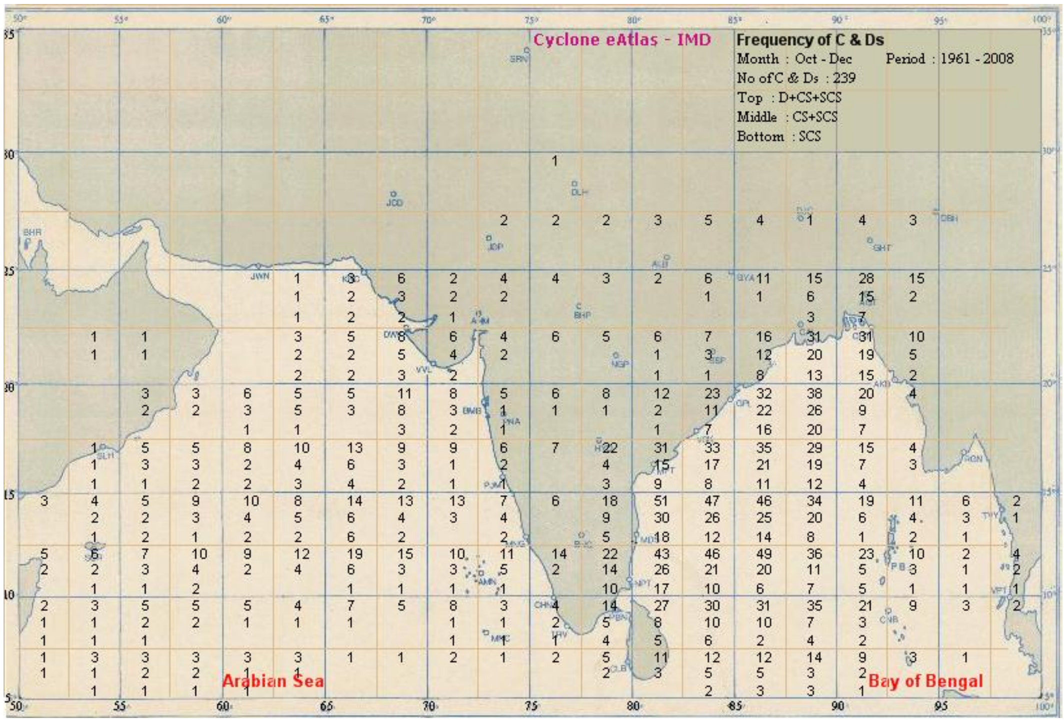

Fig. 4. Frequency of cyclonic disturbances which crossed in the 2.5° × 2.5° grid area during the period 1961 to 2008

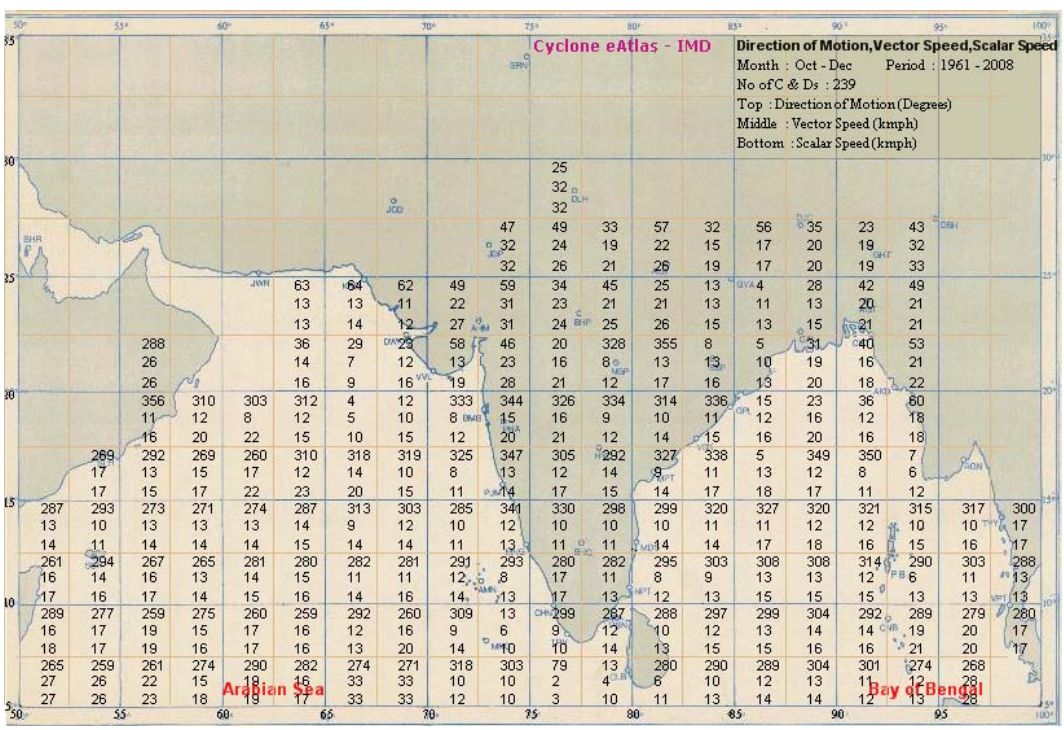

Fig. 5. Direction of motion, vector speed and scalar speed of the cyclonic disturbances which crossed in the 2.5° × 2.5° grid area during the period 1961 to 2008

the nearest neighbourhood defined for analysis with LOESS is the set of data collected from similar CD tracks based on the analogue parameter, mentioned in Section 3.1.

TABLE 3

Comparative forecast errors for individual tracks of cyclonic storms over BoB during the period 1996-2008

No.	Starting date of the Cyclonic disturbance	Max. intensity	No. of forecasts	Mean Analogue size	24 hr forecast error (in kms)		
					ANL	ACR	LOESS
1	28 Nov 1996	VSCS	11	46	174	152	142
2	04 Nov 1997	CS	3	18	251	229	212
3	26 Oct 1998	DD	3	76	176	151	157
4	19 Nov 1998	VSCS	3	61	186	195	176
5	15 Oct 1999	VSCS	4	43	138	205	188
6	25 Oct 1999	SUCS	6	31	160	145	142
7	15 Oct 2000	CS	5	44	157	104	80
8	26 Nov 2000	VSCS	5	84	124	135	118
9	23 Dec 2000	VSCS	5	35	174	89	103
10	23 Nov 2002	CS	6	63	146	178	146
11	11 Dec 2003	SCS	5	54	175	121	127
12	28 Nov 2005	CS	5	95	181	161	157
13	06 Dec 2005	CS	5	58	172	160	155
14	15 Dec 2005	DD	6	35	128	131	147
15	11 Nov 2007	VSCS	5	61	206	176	177
16	04 Dec 2008	DD	3	75	173	213	208

ANL – Analogue, ACR – Analogue-cum-Regression; LOESS – Locally weighted scatter plot smoothing;
Legend for DD, CS, SCS, VSCS, SUCS is as in Table 1

LOESS technique has its own advantages and disadvantages. The biggest advantage of LOESS over other techniques is the fact that it does not require the specification of a function to fit a model to all of the data in the sample. Instead, a smoothing parameter value and the degree of the local polynomial alone are required. Further, LOESS is very flexible and simple, making it ideal for modeling complex processes for which no theoretical models exist. Hence it is one of the modern regression techniques used for applications that fit the general framework of least squares regression but which have a complex deterministic structure. The disadvantages are that LOESS requires fairly large, densely sampled data sets in order to produce good models. Further, it does not produce a regression function that is easily represented by a mathematical formula. Keeping the advantages in mind this study has been attempted.

For this statistical study, in-house software has been developed to predict the future position of CD tracks using the techniques, ANL, ACR and LOESS. The output of the software is checked with standard software package like R-language which is open source and available freely over the internet and found correct. For a given CD position, once current and previous 24 hour positions are known, the future position could be predicted using these three techniques by executing this software written in FORTRAN and run using Salford compiler. As it is developed in-house, it is easy to update the database periodically.

3.4. *Calculation of forecast error*

The test period of tracks for the period 1996-2008 has been considered for comparing with the analogue, analogue-cum-regression and LOESS technique derived forecast positions. The Forecast Error (FE) is defined as the distance between actual and forecast positions. Here, the distance between the points is taken as the great circle distance (GCD) between them which is calculated by a spherical trigonometric formula.

For the same period, we have calculated FE based on pure CLIPER (Climatology and Persistence) and skill score for the FE has been calculated using the formula (Neumann and Pelissier, 1981) given by $P = 100^* (E_c - E_m)/E_c$, where P is the percentage improvement or deterioration over CLIPER FE. E_c is the CLIPER FE and E_m is the model FE (in this case, each of the three techniques).

TABLE 4

24 hrs average forecast errors resulting from three techniques for the period 1996-2008

24 hr forecast error (kms) (No of cases 80)		
ANL	ACR	LOESS
167	154	147
	Skill score in percentage (P)	
-4.4	3.8	8.1
	Mean error in three recurvature positions (kms)	
273	280	244
	Skill score in percentage (P)	
1.4	-1.1	11.9

Legend as given in TABLE 3

TABLE 5

Test of significance for the difference in 24 hour forecast errors calculated by the three techniques

Difference	Total number of forecast cases (N)	Difference in FE (kms)	SD (difference, kms)	Calculated *t* value	Significant level
ACR - LOESS	80	7	44.7	1.37	NS
ANL - ACR	80	13	56.3	2.09	5%
ANL - LOESS	80	20	58.3	3.06	1%

NS : Not Significant, Legend as given in TABLE 3

3.5. *Calculation of bias*

In order to account for the bias, Along-Track (AT) error and Cross-Track (CT) error have been computed for all the three techniques, *viz*., ANL, ACR and LOESS. In order to account for the bias, a coordinate system that is oriented along the storm track is employed. The error components in this framework are AT and CT (Elsberry and Peak 1986) which are respectively defined as the parallel and normal projections of the forecast track along the observed track. Elsberry and Peak (1986) define CT and AT error components relative to persistence or CLIPER tracks respectively which are used to normalise the forecast positions, which in turn help to focus on improving the accuracy of track prediction relative to these "no-skill" aides.

In this study, CLIPER forecast is used to normalise the forecast position. The CT component can be thought of as a turning motion which is a deviation to the right (left) side of the forecast track. The AT component is related to acceleration/deceleration of the TC. The direction of the CLIPER forecast track, actual track and forecast track obtained by this model are converted into a 16 point compass scale. With the CLIPER track as the reference track, for each forecast position, actual and forecast tracks obtained by this method are checked as to whether the forecast track is along the same direction, or right or left of the CLIPER track. Similarly, whether, for each forecast position, speeds of actual and forecast tracks are same (*i.e*., within ± 2 kmph), fast (greater than 2 kmph) or slow (less than 2 kmph) in comparison to the speed of CLIPER track is also determined. Fig. 3 is the pictorial depiction of the CT and AT components. The 3 × 3 contingency table for each of the three techniques ANL, ACR and LOESS has been developed. The three contingency parameters for CT are Correct(C), Left (L) and Right (R). Similarly for AT the three parameters are Correct, Slow(S) and Fast (F).

3.6. *Calculating skill score*

Forecast skill is usually presented as a skill score, which is interpreted as percentage improvement over the reference forecasts. One of the most frequently used skill score is HSS, which is defined as (Wilks, 1995)

$$HSS=\frac{\sum p(y_i,o_i)-\sum p(y_i)\,p(o_i)}{1-\sum p(y_i)\,p(o_i)} \qquad (6)$$

TABLE 6

Frequency distribution and percentage of forecast error of cyclonic disturbances over BoB during the period 1996 to 2008

Forecast error	Frequency			Percentage		
(in km)	ANL	ACR	LOESS	ANL	ACR	LOESS
0-50	6	5	7	7.5	6.3	8.8
51-100	8	17	17	10.0	21.3	21.3
101-150	24	22	18	30.0	27.5	22.5
151-200	11	17	23	13.8	21.3	28.8
201-250	21	7	7	26.3	8.8	8.8
251-300	6	8	3	7.5	10.0	3.8
300-350	2	2	2	2.5	2.5	2.5
350-400	2	2	3	2.5	2.5	3.8
TOTAL	80	80	80	100	100	100

Legend as given in TABLE 3

TABLE 7

Contingency Table for Along-Track (AT) and Cross-Track (CT) components

(a) ANL

		Observed		
		C	S	F
Forecast	C	4	12	5
	S	5	14	8
	F	4	18	10

(b) ACR

		Observed		
		C	S	F
Forecast	C	3	17	15
	S	8	20	4
	F	2	7	4

(c) LOESS

		Observed		
		C	S	F
Forecast	C	6	16	12
	S	6	19	4
	F	1	9	7

Along-Track (AT) speed component

Contingency Table

C – Correct; S – Slow; F – Fast

(a) ANL

		Observed		
		C	L	R
Forecast	C	8	24	10
	L	0	12	1
	R	7	3	15

(b) ACR

		Observed		
		C	L	R
Forecast	C	9	14	12
	L	1	21	3
	R	5	4	11

(c) LOESS

		Observed		
		C	L	R
Forecast	C	7	9	7
	L	2	26	3
	R	6	4	16

Cross-Track(CT) direction component

Contingency Table

C – Correct; L – Left; R – Right

Legend as given in TABLE 3

where, $p(y_i, o_i)$ is the joint distributions of forecast and observed tracks, $p(y_i)$ is the marginal distribution of the forecast and $p(o_i)$ is the marginal distribution of the observed respectively. HSS is based on summarising the results as a square contingency table by indicating the proportion which is correct as the basic accuracy measure.

Thus, perfect forecasts receive HSS = 1, forecast equivalent to the reference forecasts receive zero scores, and forecasts worse than the reference forecasts receive negative scores. Other skill scores (Wilks, 1995) which have been calculated are Peirce Skill Score (PSS) and Proportion Correct (PC). Other skill score indices like Critical Success Index (CSI), Theta (θ) which is called Odds Ratio, Bias (B), False Alarm Ratio (FAR), Probability Of Detection (POD) and Probability Of False Detection (POFD) in the form of 2 × 2 contingency table have been calculated for all parameters C, L, R, C, S and F as mentioned in Section 3.5 above.

4. Results and discussion

The frequency of CDs crossing through the square grid of 2.5° × 2.5° for the NIO during the period 1961-2008 is depicted in Fig. 4 from which it is evident that the frequency is more than 15 in all the grids of BoB except in the grids to the west of 95° E and south of 7.5° N. The normal direction of movement, scalar speed and vector speed of the CDs of NIO during 1961-2008 are depicted in Fig. 5 for the various grids. The direction of the track varies mainly from west to northeast.

In the present study, it is observed that when the CD is moving towards southwest, it is hard to get sufficient number of analogues. To maintain the quality of the forecast reliability of this model, those situations when there are less than 15 analogue cases have been rejected. Such rejected cases constitute just 15% of the total number. Provisions are available in the model to get more analogues by relaxing the analogue parameter conditions. But, in order to maintain homogeneity in the analysis of forecast error and to compare the results with other cases in the model, criteria for selection of analogue parameters have been maintained as the same for all situations. For a CD, at least 3 or more forecast positions are considered (Bessafi, 2002) in this analysis to calculate mean FE and hence with this essential condition initially formulated for the analysis, the CDs which have been taken for verification (Fig. 2) amount to 16 during the period 1996-2008. 80 individual instances of forecast positions are available for the 16 CDs. Mean forecast error for each of the 16 CDs were calculated by the three techniques and presented in Table 3. As a whole, mean FE was calculated for the 16 CDs (Total : 80 positions) and are presented in Table 4 along with the skill score in percentage (P) as mentioned in Section 3.4 above.

4.1. *Track forecast error and skill*

The number of analogues selected to find the prediction position is the analogue size. The sum of all analogue sizes divided by the number of forecast positions in a CD is the mean analogue size for a given CD. The mean analogue sizes for each of 16 CDs are given in Table 3. This mean analogue size is more than 35 in all the cases except for two. In the case when the mean analogue size was 18, the track forecast errors were high in all the three methods and were greater than 200 km. No definite conclusion about higher track FEs when analogue size is small compared with the rest of the analogue sizes of the 16 CDs could be drawn from this analysis. Correlation coefficient calculated between analogue size and track FE did not yield reliable linkages. The comparative FE for 16 CDs which occurred in BoB is also given in Table 3. For each of the CDs considered, the mean FE varies from 80 to 212 km for LOESS technique. FE of 13 (81%) out of 16 CDs is less for LOESS when compared with ANL. Similarly, FE of 12 (75%) out of 16 CDs is less for LOESS when compared with ACR.

TABLE 8

Skill Scores for the three techniques for Along-Track and Cross-Track components

	PC	HSS	PSS
		AT	
ANL	0.35	0.01	0.01
ACR	0.34	0.00	0.00
LOESS	0.40	0.11	0.12
		CT	
ANL	0.44	0.22	0.26
ACR	0.51	0.29	0.32
LOESS	0.61	0.41	0.43

AT : Along-Track; CT : Cross-Track; PC : Proportion Correct; HSS : Heidke Skill Score; PSS : Peirce Skill Score.
Legend as given in TABLE 3

It is seen that the 24 hours FE (in kms) in prediction of track positions is minimum through LOESS when compared to the other two techniques. The average FEs calculated by the three techniques ANL, ACR and LOESS are 167, 154 and 147 kms respectively as indicated in Table 4. The FE obtained by employing CLIPER technique is 160 km. In terms of skill score based on CLIPER, LOESS and ACR show improvement of 8.1% and 3.8% respectively whereas ANL shows deterioration by 4.4% as shown in Table 4. Similar results in the case of ANL over CLIPER have been observed by the National Hurricane Centre, Florida, USA also (Neumann and Pelissier, 1981).

FE at the recurvature position is normally higher than that of the other prediction points. For the period of study, there were three occurrences of recurvature of the CDs (years 1996, 2002 and 2005). FE calculated for such recurvature cases is 277 km by CLIPER. For the three

TABLE 9

Skill Score indices for the three techniques for Along-Track and Cross-Track components

		AT					
		CSI	THETA	B	FAR	POD	POFD
Correct	ANL	0.13	1.31	1.62	**0.81**	0.31	**0.25**
	ACR	0.07	0.33	2.69	0.91	0.23	0.48
	LOESS	**0.15**	**1.19**	2.62	0.82	**0.46**	0.42
Slow	ANL	0.25	0.83	0.61	0.48	0.32	0.36
	ACR	**0.36**	1.67	0.73	0.38	**0.46**	0.33
	LOESS	0.35	**1.98**	0.66	**0.35**	0.43	**0.28**
Fast	ANL	**0.22**	1.22	1.39	0.69	**0.44**	0.39
	ACR	0.13	1.12	0.57	0.69	0.17	**0.16**
	LOESS	0.21	**2.06**	0.74	**0.59**	0.30	0.18
		CT					
Correct	ANL	0.16	1.04	2.80	0.81	0.53	0.52
	ACR	0.22	2.25	2.33	0.74	**0.60**	0.40
	LOESS	**0.23**	**2.68**	1.53	**0.70**	0.47	**0.25**
Left	ANL	0.30	17.78	0.33	**0.08**	0.31	**0.02**
	ACR	0.49	10.79	0.64	0.16	0.54	0.10
	LOESS	**0.59**	**14.40**	0.80	0.16	**0.67**	0.12
Right	ANL	0.42	6.00	0.96	0.40	0.58	0.19
	ACR	0.31	3.67	0.77	0.45	0.42	**0.17**
	LOESS	**0.44**	**7.04**	1.00	**0.39**	**0.62**	0.19

CSI: Critical Success Index; THETA: Odds ratio; B: Bias; FAR: False Alarm Ratio;
POD: Probability Of Detection; POFD: Probability Of False Detection.
Legend as given in TABLE 3

recurved systems, the FE is also minimum through LOESS technique as is shown in Table 4. In terms of skill score based on CLIPER for recurvature cases, LOESS and ANL show improvement of 11.9% and 1.4% whereas ACR shows deterioration of 1.1%. The difference in the FE obtained through the ANL and ACR techniques is statistically significant at 5% level (Table 5). Similarly the difference in the FE obtained by ANL and LOESS techniques is statistically significant at 1% whereas the difference in FE calculated by ACR and LOESS is not significant at all. The frequency distributions of FE for the three techniques are given in Table 6. The FE of LOESS technique is below 200 kms in 81% of the total number of cases whereas it is 76% and 61% respectively for ACR and ANL techniques. The ACR and LOESS techniques are good and reliable in the FE range of 201 to 250 kms compared to ANL. In case of 150 km FE and above, ACR is better than the other two methods.

The contingency tables for AT and CT errors for 80 cases of forecast are given in Table 7. PC, HSS and PSS are calculated for all methods for BoB and shown in Table 8. The results for AT component show that for BoB, the PC values for ANL, ACR and LOESS are 0.35, 0.34 and 0.40 respectively. HSS for the three techniques are 0.01, 0, 0.11 and PSS are 0.01, 0.0 and 0.12 respectively indicating that LOESS is the best among all the techniques used. Similarly for CT component, PC is 0.44, 0.51 and 0.61 for ANL, ACR and LOESS whereas HSS is 0.22, 0.29 and 0.41 respectively. PSS values are 0.26, 0.32 and 0.43 for the three techniques which also indicates that LOESS is superior than the others. The difference between HSS and PSS for AT and CT for all the three methods is very small.

The forecast exhibits little bias. Table 9 presents the skill score indices like CSI, θ, B, FAR, POD and

POFD for the three techniques for AT and CT components for C, S, F, C, L and R. It is observed that figures indicated in 'bold' are the best results and the corresponding technique is superior for each of the categories C, S, F, C, L and R. In most of the cases, the result for LOESS as seen in Table 9 is the best and reliable compared to ANL and ACR.

5. Conclusions

The major results of the study where in track prediction of cyclones of Bay of Bengal were attempted based on three different statistical techniques ANL, ACR and LOESS are summarised here in below:

(*i*) The 24 hours average FE from LOESS is 147 km which is less than the 167 and 154 km FEs by ANL and ACR respectively. The difference in mean FE calculated through ACR and LOESS is not significant whereas the 13 km difference in the mean FE obtained through the ANL and ACR techniques is statistically significant at 5% level. Similarly the difference in the FE obtained by ANL and LOESS techniques is 20 km which significant at 1% which testify to the superiority of the latter. While ANL technique shows no skill, ACR and LOESS techniques show skill of about 4% and 8% respectively as compared to CLIPER model track forecast for 24 hours.

(*ii*) PC, HSS and PSS for LOESS values are 0.40, 0.11 and 0.12 respectively for AT component. Similarly for CT, PC, HSS and PSS are 0.61, 0.41 and 0.43 which indicate better accuracy and superiority of LOESS technique over the ANL and ACR techniques. When skill score of AT and CT is considered, LOESS shows good result.

(*iii*) The results indicate that LOESS shows more promise than the other statistical technique in the prediction of track positions of cyclones and depressions over the BoB.

References

Anthes, R. A., 1982, "Tropical cyclones, their evolution, structure and effects", *American Met. Society*, p208.

Bandyopadhyay, B. K. and Singh, Charan, 2006, "Cyclone track forecast by combining persistence climatology and synoptic method", *Mausam*, **57**, 619-628.

Bansal, R. K. and Datta, R. K., 1974, "A statistical method of forecasting the movement of cyclonic storms in the Bay of Bengal", *Indian J. Met. Geophys*., **25**, 391-396.

Bessafi, M., Lasserre-Bigorry, A., Neumann, C. J., Picnolet-Tardan, F., Payet, D. and Lee-Ching, Ken, 2002, "Statistical Prediction of Tropical Cyclone Motion: An Analog - CLIPER Approach", *Weather and forecasting*, **17**, 821-831.

Cleveland, W. S. and Devling, S. J., 1988, "Locally weighted regression: An approach to regression analysis by local fitting", *Journal of the American Statistical Association*, **83**, 596-610.

IMD, 2003, "Cyclone Manual", India Meteorological Department, New Delhi.

Cyclone eAtlas, 2008, "Tracks of storms and depressions in the Arabian Sea and the Bay of Bengal", India Meteorological Department.

Datta, R. K. and Gupta, R. N., 1975, "Tracking tropical storm in the Bay of Bengal and calculation of probabilities of its striking east coast by storm analogue technique", *Indian J. Met. Geophys*., **26**, 345-348.

Elsberry, R. L. and Peak, J. E., 1986, "An Evaluation of Tropical Cyclone Forecast Aids Basedon Cross-Track and Along-Track Components", *Mon. Wea. Rev*., **114**, 147-155.

Hope, J. R. and Neumann, C. J., 1970, "An operation technique for relating the movement of existing tropical cyclones to past tracks", *Mon. Wea. Rev*., **98**, 925-933.

Kendall, M. G. and Stuart, A., 1968, "The Advanced Theory of Statistics", Griffin, London, Vol. II, 278-374.

Mohapatra, M., Sikka, D. R., Bandyopadhyay, B. K. and Tyagi, Ajit, 2013, "Outcomes and challenges of Forecast Demonstration Project (FDP) on landfalling cyclones over the Bay of Bengal", *Mausam*, **64**, 1-12.

Montgomery, D. C., Peck, E. A. and Vining G. G, 2003, "Introduction to Linear Regression Analysis", John Wiley and Sons (Asia), Pvt. Ltd., Singapore, 239-241.

Neumann, C. J. and Mandal, G. S., 1978, "Statistical prediction of tropical storm motion over Bay of Bengal and Arabian Sea", *Indian J. Met. Geophys*., **29**, 487-500.

Neumann, C. J. and Pelissier, J. M., 1981, "Models for the prediction of tropical cyclone motion over the North Atlantic : An operational evaluation", *Mon. Wea. Rev*., **109**, 522-538.

Raj, Y. E. A., 2010, "Forecasting Manual on Tropical Cyclones of Indian Seas", IMD, Ch. 5.2.

Raj, Y. E. A., Pillai, P. V., Vaidya, D. V. and Natu, J. C., 1991, "Prediction of movement of post monsoon depressions and cyclonic storms over Arabian Sea", *Vayu Mandal*, **21**, 120-130.

Ramanathan, Y. and Bansal, R. K., 1977, "On the application of a primitive equation barotropic model for the prediction of storm tracks in Indian region", *Indian J. Met. Geophys*., **28**, 169-176.

Sen Sharma, A. K., 1983, "Tropical cyclone movement prediction problem, recent Indian work", *Vayu Mandal*, **12**, 30-34

Sievers, O., Fraedrich, K. and Raible, C. C., 2000, "Self-Adapting analog ensemble prediction of tropical cyclone tracks", *Weather and Forecasting*, **15**, 623-629.

Sikka, D. R., 1975, "Forecasting movement of tropical cyclones", *Indian J. Met. Geophys*., **26**, 323-325.

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

Kumar, Surendra and Prasad, Kanti, 1973, "Prediction of tropical storms in Indian Seas", *Indian J. Met. Geophys*., **24**, 31-34.

Wilks, D. S, 1995, "Statistical Methods in the Atmospheric Sciences", Academic Press, San Diego, 233-258.