551.577 : 551.509.324.2

ARIMA MODEL FOR RAINFALL PREDICTION AT GANGTOK (SIKKIM)

1. Understanding the annual rainfall and summer monsoon seasonal rainfall behaviour is important for any place due to the dependency on rains for virtually every activity of human life. Accurate forecasts of rainfall on various temporal and spatial scales are important in view of the above. Many economic activities depend on water resources, especially, the fields of agriculture and power generation that in turn depends on seasonal rains. In this paper, summer monsoon season and annual rainfalls considered for the study over Gangtok. The city Gangtok lies at an elevation of 5500 ft above mean sea level for which this study has been taken up is the state capital of Sikkim, a tiny land locked Himalayan state, stretching

between 27° 00′ 46″ to 28° 07′ 48″ North latitude and 88° 00′ 58″ to 88° 55′ 25″ South longitude covering a distance of 112 km from North to South and 64 km from East to West and a total area of 7096 sq. km. The State climate is tropical, temperate and alpine. The station Gangtok receives an average annual (Seasonal) rainfall of 3604.6 mm (2319.6 mm) and with a variance of 11% (13%). The figures are computed on the basis of data available during the period 1979-2007 (in total 29 years). On an average, approximately 64% of the annual rainfall occurs during the southwest monsoon season over Gangtok. Of the total annual rainfall, Gangtok receives during the winter (three months period Dec, Jan & Feb), the station receives an average total rainfall of 116.4 mm, pre-monsoon (three months period Mar, Apr & May), the station receives an average total rainfall of 947.4 mm, post monsoon (two months period Oct & Nov), the station receives an average total rainfall of 215.6 mm. Earlier in research study, monthly, Seasonal and annual rainfall series for Gangtok as been analyzed using trend analysis to discuss the climate change scenario over Gangtok which to much extent covers the rainfall variability on the time scale of monthly, seasonal and annual (Seetharam, 2008). It is seen that the following regional weather systems influence the rainfall pattern and distribution in the state of Sikkim.

(*i*) Western disturbance phenomenon that originates in the Mediterranean region and travel eastwards.

(*ii*) Norwesters – a pre-monsoon thunderstorm activity.

(*iii*) Low pressure systems like, troughs, lows, depressions and cyclonic storms that form in the Bay of Bengal and cross Orissa coast and dissipate over Chattisgarh, Bihar or Uttar Pradesh or cyclonic systems that cross Bangladesh or Bengal coast and dissipate in North east region in proximity to Sikkim.

(*iv*) Monsoon trough, specially, when the axis of monsoon trough passes through foothills of Himalayas.

(*v*) Upper air cyclonic circulations that travel eastwards embedded in the westerly flow over Sikkim.

In this paper, firstly investigated, the tele-connection between ENSO phenomenon and summer monsoon seasonal rainfall as the study may throw light on variability summer monsoon rainfall over place of study. Many climatic indices were developed which provide valuable information about the coupled ocean-atmosphere phenomenon. The reanalysis data on various climatic indices provide ample opportunity for scientific investigations on annual rainfall and seasonal rainfall (Kanamitsu *et al*., 2002). The same climatic indices were utilized in this study to explore their relationship with seasonal rainfalls over Gangtok. Earlier studies indicated that Indian summer monsoon rainfall shows a large inter-annual variability (Mooley and Parthasarathy, 1984) and is being thought due to anomalous circulation patterns that develop during Indian summer monsoon. Another study indicated that variation in Sea-Surface Temperature (SST) plays a major role in long-term variability as the variations usually persist for a long term monsoon rainfall may be predicted based on the mean monthly conditions at low latitudes (Charney and Shukla, 1981). Through one earlier study it has been shown that the inverse relationship between summer monsoon sub-divisional rainfall of India and NINO 3.0 and NINO 3.4 region Sea-Surface Temperatures (SST) show influence of ENSO on Indian summer monsoon sub-divisional rainfalls of some sub-divisions and succeeding post monsoon SST anomalies of both the above regions indicate a cooling effect and the conclusions were drawn based on the data period (1951-1999) (Seetharam, 2005).

2. For the purpose of the study, annual rainfall data for the period has been collected from the meteorological records of MC Gangtok from 1979-2007, the normalized OLR equator (160° E - 160° W), other climatic indices like normalized Multivariate ENSO Index (MEI) and three more indices, namely, LI, EI and ESPI have been collected from the Climate Prediction Center, National Center for Environmental Prediction, National Oceanic and Atmospheric Agency, Washington data set archives. The climate indices were used in computation of single average annual value for the purpose of study. The warm phase years, cold phase years and the neutral phase years were obtained from ENSO Index based on Japan Meteorological Agency (JMA) Sea Surface Temperature Anomalies (SSTA). For the purpose of computation of ENSO based Precipitation Index, two regions *viz*., Maritime Continent covering the area (10° N – 10° S; 90° E – 150° E) and central to eastern Pacific covering the area (10° N – 10° S; 160° E – 100° W) selected. The maximum and minimum average precipitation anomalies are found for both regions and homogenized the record (Curtis and Adler, 2000). The ENSO Precipitation Index (ESPI) is given as ESPI = normalized EI – LI. Positive (negative) values indicate the warm (cold) phase of the ENSO cycle. Further more, the ENSO years were obtained based on the SST anomalies.

3. Firstly, the annual rainfall series of the station Gangtok has been presented in Fig. 1. The long period average (LPA) and the linear trend have been plotted superimposed on the series. The average annual rainfall line along with linear trend line indicated on the Fig. 1. The linear trend indicated below LPA (1979-2007) rainfall from 1957 to 1987 and above LPA (1979-2007) rainfall from 1987 onwards to till date and the trend likely to

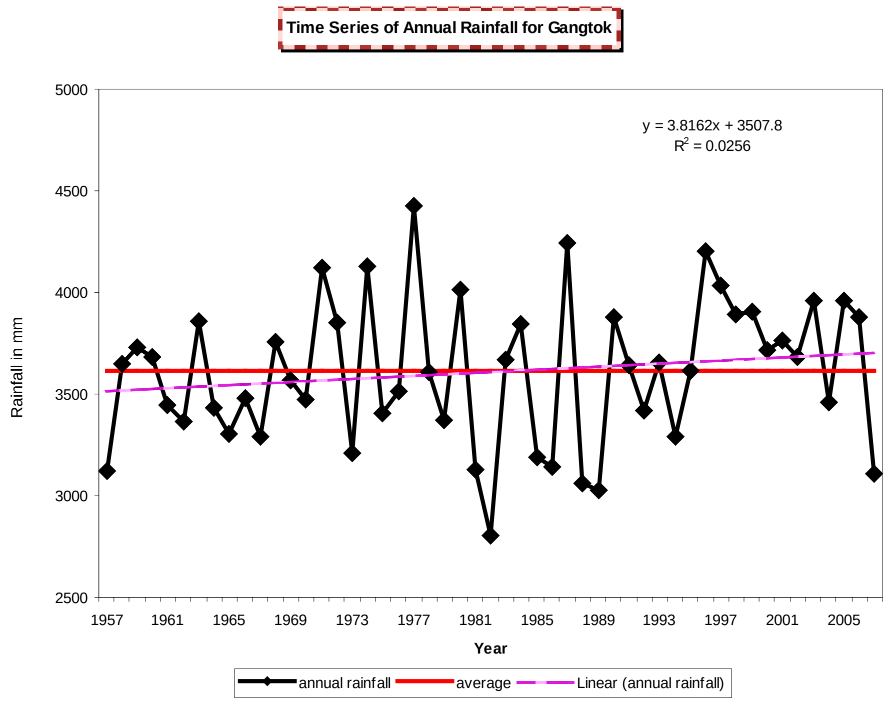

Fig. 1. Showing Annual Rainfall series of Gangtok with Long Period Average and Linear trend

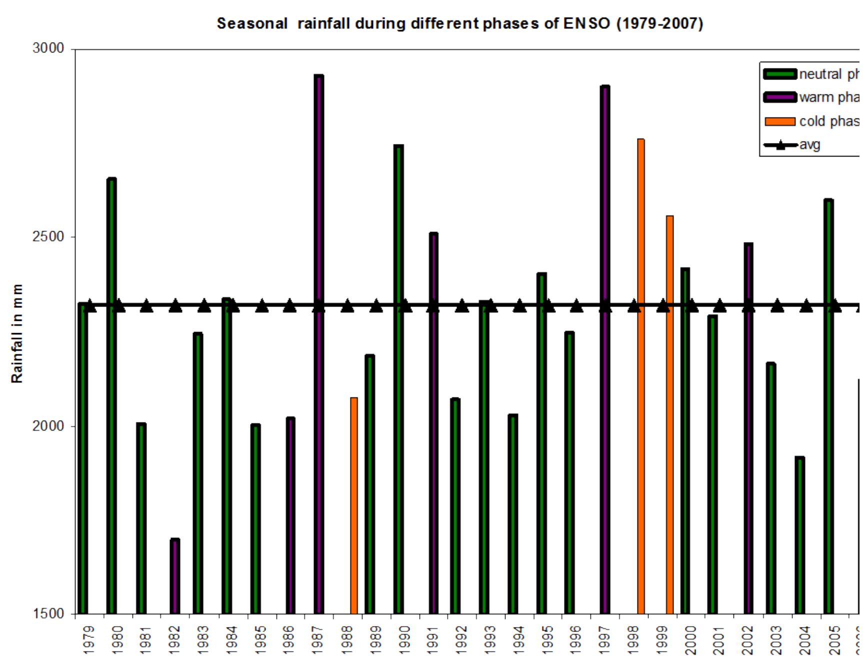

Fig. 2. Showing Seasonal Rainfall (Jun – Sep) during three phases of ENSO (Warm, Cool and Neutral) with LPA annual

TABLE 1

Showing correlation coefficients and statistics of annual rainfall series and climatic indices (1979-2007)

Parameter	Mean	Standard deviation	Coefficient of variance	CC with annual R/F	CC with seasonal R/F	Test statistic for annual R/F	Test statistic for seasonal R/F
Annual rainfall (mm)	3604.6	384.5	11%	-	0.74	-	5.65*
Seasonal rainfall (mm)	2319.6	302.3	13%	0.74		5.65*	-
MEI	0.4	0.7	195%	-	0.24	0.87	1.26
OLR	0.2	0.7	478%	-	-0.16	-0.45	-0.83
LI	0.1	0.6	642%	-	-0.03	0.93	-0.15
EI	-0.1	0.7	1177%	-	0.30	0.48	1.63
ESPI	-0.1	0.7	588%	-	0.22	0.56	1.12

* Significant at 5% confidence level

continue for another 9 years as suggested by the trend line. The regression equation of the trend line and the square of the residual are presented on Fig. 1. The regression equation can itself be utilized as tool for forecasting annual rainfall. The analysis of the normalized OLR data chosen for this study indicated a linear trend with an increasing trend suggesting an average warm pool in the region along the equator for the period of study of 28 years. The analysis of MEI index indicated a decreasing trend. As per the decreasing trend the extent of warming in the Pacific region on the average decreasing and further continuation of the same trend indicate cold phase in Pacific Ocean during coming years. The analysis of ESPI indicated a linear trend that is decreasing. The EI and LI were opposite and alternate as per the formula and ESPI suggesting an increasing in the intensity of cold phase in the Pacific Ocean. The MEI and ESPI showed contrasting phase in the Pacific Ocean but linear trends in both suggest a favorable conditions for cold phase only. As per the classification the La Nina years were 1988, 1998, 1999 and 2007. The El Nino years were 1982, 1986, 1987, 1991, 1997, 2002 and 2006. The neutral years were 1979, 1980, 1981, 1983, 1984, 1985, 1989, 1990, 1992, 1993, 1994, 1995, 1996, 2000, 2001, 2003, 2004 and 2005. In total there were seven (7) warm phase years, four (4) cold phase years and eighteen (18) neutral years during the period of study. On seasonal scale, *i.e.*, during summer monsoon season, 7 years (39%) received above average rainfall and remaining 11 years (61%) received average or below average rainfalls out of 18 neutral phase years selected. During 7 warm phase years, 4 years (57%) received above average rainfalls and remaining 3 years (43%) received below or average rainfalls. Similarly, out of 4 cold phase years, 2 years (50%) received above average, 2 years (50%) received below average rainfalls. In Fig. 2, the summer monsoon seasonal realized rainfalls during warm, cold and neutral phase events are presented. The correlation coefficients found between the summer monsoon seasonal rainfall series of Gangtok and chosen climatic indices (OLR, MEI, LI, EI, ESPI) were low, weak and insignificant (Table 1). From Table 1, it can be seen that the normalized OLR equator, MEI and other ENSO indices are highly variable on annual scale. The figures indicate that the OLR equator was positive on the average suggesting the idea of warm pool. Similarly, positive annual MEI indicates that again on average the warm pool in Pacific Ocean during last 28 years considered in this study. Some of the earlier studies indicated that the MEI better reveals the nature of the coupled ocean-atmosphere system better (Wolter and Timlin, 1998). The other three ENSO indices also suggest an average condition of cold phase of ENSO events. The found correlation coefficient rules out the possibility of any strong tele-connection with the pacific oceanic warming or cooling.

Subsequently, the ARIMA model was utilized as so far no correlation studies for Gangtok has been carried out indicating significant correlation between regional or global circulation parameters and summer monsoon rainfall. The development of multi regression technique to forecast annual rainfall/seasonal rainfall is thought not appropriate because the indices chosen indicate the same phenomenon measured based on different methodologies. Here, the attempt is only to identify the index that bears good correlation with annual/seasonal rainfalls over Gangtok and explains the inter-annual variability of monsoon over Gangtok. Moreover, to understand the spatial and temporal variability of the rainfall in 1-day, 2-day and extreme rainfall events, various distribution methods were suggested and study to find appropriate distribution out of many statistical distributions available would be worthy. Furthermore, the time series analysis is one of the important methods for understanding the observational data that has been produced by various processes. The time series analysis provides a way for

TABLE 2

Showing Auto correlation and Partial Auto Correlation coefficient for the annual rainfall series (1979-2007)

Lag	Auto correlation coefficient	95% confidence interval	Partial auto correlation coefficients
1	0.016	0.364	0.16
2	-0.23		-0.23
3	0.41*		0.44*
4	0.16		0.06
5	-0.17		0.14
6	0.16		0.06
7	0.22		0.10
8	-0.18		-0.14
9	-0.01		0.02
10	-0.0		-0.28

* Significant at 5% level.

TABLE 3

Showing ARIMA (3,0,3) parameters

Predictor	Coefficient	Standard error	*t*-ratio	*p*-value
AR (1)	0.51	0.07	7.21	0.00
AR (2)	-0.22	0.07	-2.98	0.01
AR (3)	0.72	0.07	9.96	0.00
MA (1)	0.48	0.09	5.27	0.00
MA (2)	0.20	0.09	2.18	0.04
MA (3)	0.24	0.09	2.70	0.01

forecasting, monitoring. For the purpose of prediction, the ARIMA model requires only past data to predict the future data. The Box-Jenkins mathematical model involves identification of appropriate model, fitting the model to the data set and then using the model in forecasting. One basic requirement of the model is that the time series to be stationary and the non-stationary series need to be reduced to stationary series before application of this methodology.

4. *ARIMA (p, d, q) Model* - For the purpose of forecasting, a perfect model fitting well the time series can be searched. However, in this paper, only a model with $p = 3$, $d = 0$ and $q = 3$ has been selected in this paper and the results discussed. If the given series is non-stationary, we take a first-difference of the series so that the resultant series becomes stationary.

$$D = X_t - X_{t-1}.$$

The new series $X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_i X_{t-i} + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \ldots + \beta_i u_{t-i} + E$. The term 'E' is a white noise process. 'a_0' is the constant term or mean of the process. Selection of a model so that the residuals are as small as possible and replicates the patterns in the series as closely as possible produces more or less accurate forecasts. As a first step in the analysis of the time series, auto correlation function and partial correlation function are determined. The series is stationary or not can be decided by considering the Auto Correlation Coefficients. If the Auto Correlation coefficients of the time series values dies down fairly quickly, then the time series values should be considered stationary and if not the time series is considered non-stationary. Auto-regressive integrated moving average (ARIMA) models were also used to forecast the All India Summer Monsoon Rainfall (AISMR) as well as the monsoon rainfall over Northwest-India and Peninsular India, which were reported to have shown marginally better forecasting skill over the multiple regression models (Thapliyal, 1981 & 1990). However, the autocorrelations (-0.12, 0.04, 0.08, -0.09 and -0.02 for lags 1 to 5 respectively) in All India Summer Monsoon Rainfall during the period 1871-1990 are statistically insignificant (Parthasarathy *et al.*, 1994). In view of this, the applicability of ARIMA models for monsoon rainfall forecasting is doubtful. In this paper, an ARIMA model has been fitted to the time series and further, the reliability of the forecasts generated by the ARIMA model has been checked using verification procedures. As a part of the study, first 10 lag auto-correlation coefficients and partial auto-correlation coefficients are determined from the time series. The auto-correlation coefficients and partial correlation coefficients suggest a significant 3-lag coefficient in both the cases (Table 2). Therefore, a 3-years periodicity has been assumed and an ARIMA (3,0,3) model was fitted to the time series for the repetition of observed pattern of annual rainfall and also to generate one-year lead forecast from the annual rainfall time series. The observed 3-years periodicity can be linked to ocean-atmospheric phenomenon other than ENSO. However, no study has been made to investigate the influence of other than ENSO phenomenon in this paper and the study has been restricted only to ENSO related one. The resultant equation obtained by applying the ARIMA (3,0,3) model is given below.

$$X_t = 0.51\ X_{t-1} - 0.22\ X_{t-2} + 0.72\ X_{t-3} - 0.48\ u_{(t-1)} - 0.2\ u_{(t-2)} - 0.24\ u_{(t-3)} + E\ .$$

Where 'X's are past values and u's are past errors and 'E' is the white noise. The forecast for one-year lead period is as follows based on the above equation. The annual rainfall forecast for the next one-year is found to be 3848.3 mm. The 95% confidence limits can be applied to provide the higher and lower limits of the forecast with

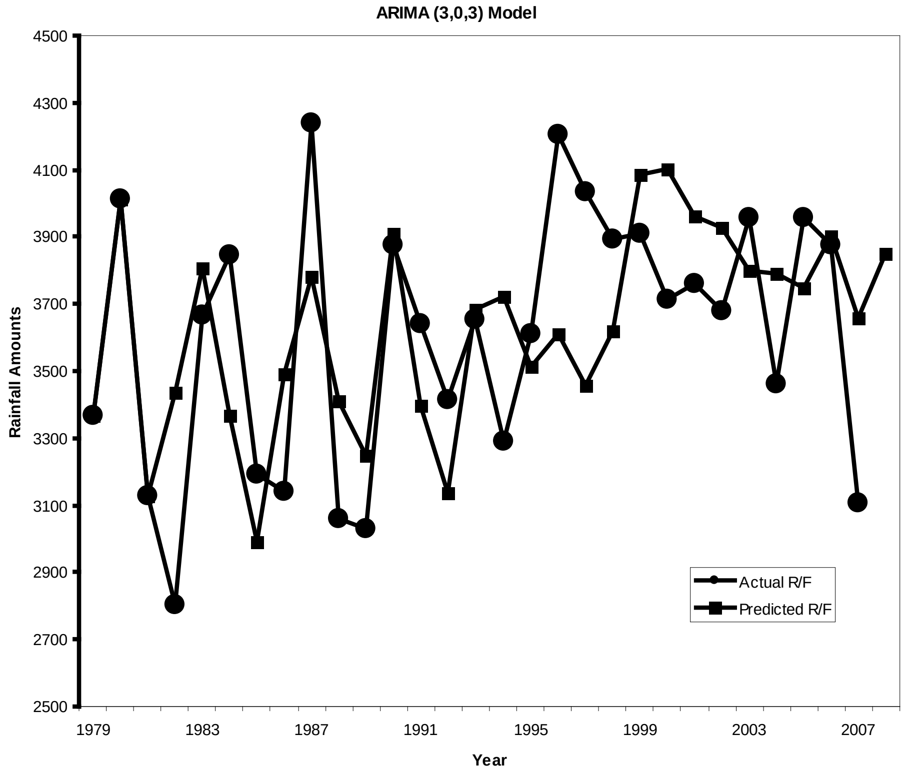

Fig. 3. Showing actual and model predicted annual rainfalls

TABLE 4

Showing deterministic/categorical forecast verification

Forecast bias	18.5
Root Mean Square Error	324.7
Correlation Coefficient	0.56
Ratio Score	0.7
Threat Score	0.6
Heide Skill Score	0.4
Hansen & Kuiper Score	0.4
Probability of Detection	0.7
False alarm	0.2
Missing rate	0.3
Correct non-occurrence	0.7
Bias for occurrence	0.8
Percentage correct	69%

standard error. The ARIMA coefficients given above are presented in the Table 3. The actual and model predicted annual rainfall amounts were shown in Fig. 3. The model forecasts are verified using the following methods in addition to probability of detection, false alarm rate,

TABLE 5

Showing the 2 × 2 contingency table for categorical verification of the ARIMA model Forecast

		Predicted		
		Above LPA	Below LPA	Total
Observed	Above LPA	12	6	18
	Below LPA	3	8	11
Total		15	14	29

missing rate, correct non-occurrence, bias for occurrence, percentage correct and results depicted in Table 4. The Forecast bias for the model is 18.5, the root mean square error is 324.7 and the correlation coefficient is 0.56. The result indicates a good prediction from the model. For categorical verification of the forecast, the predicted and actual annual rainfall figures have been placed in a 2 × 2 contingent table as shown in Table 5. The various scores like, Probability of detection, False alarm rate, Missing rate, correct non-occurrence, Bias for occurrence, percentage correct, Forecast accuracy (Hit Score or Ratio Score), Critical Success Index (Threat Score), Heide Skill

Score and Hanssen & Kuipers Score (True Skill Score) evaluated based on the above contingent table have been indicated in the Table 4. For this model, Probability of detection is 0.7, False alarm rate is 0.2, Missing rate is 0.3, correct non-occurrence is 0.7, Bias for occurrence is 0.8, percentage correct is 69%, the Hit score was 0.7, the Threat Score was 0.6, Heide Skill Score was 0.4 and the Hanssen & Kuiper Score was 0.4 for above long period average rainfall predictions. For best and perfect forecast the Probability of detection is 1.0, false alarm rate is 0.0, Missing rate is 0.0, correct non-occurrence is 1.0, Bias for occurrence 1.0, percentage correct, Forecast accuracy (Hit Score or Ratio Score) is 1.0, Critical Success Index (Threat Score) is 1.0, Heide Skill Score is 1.0. The Hit Score indicated that 70% above Long Period average rainfall predictions are correct. The threat score indicated that relative forecasting accuracy indicating lesser false alarms. The Heide Skill score indicated that this model has a skill compared with chance forecast. The Hanssen & Kuiper Score indicated an economic saving over climatology due to the predictions. The verification of model forecast is taken in comparison with LPA (1979-2007) if for normal southwest monsoon normal criteria of +/- 10% is taken then the forecast accuracy will further improve. The model predictions are very good.

5. Conclusions

The study of the climatic indices chosen firstly indicated high variance in the values on the annual time scale and further more, analysis of climatic indices indicated linear trends indicating a trend towards cooling of Pacific Ocean and over cooling of Indian Ocean. Moreover, the correlation study carried out between the climatic indices and summer monsoon seasonal rainfalls of Gangtok indicated no statistically significant correlations between summer monsoon seasonal rainfalls and climatic ENSO indices chosen. The absence of any significant correlation between climatic indices and summer monsoon seasonal rainfall indicate no relationship between ENSO phenomenon and summer monsoon seasonal rainfalls over Gangtok.

The computation of auto-regression coefficients and also partial correlation coefficients indicated a three years quasi-periodicity in the annual rainfall series. Further studies are required to understand the 3-years quasi-periodicity in annual rainfall series over Gangtok. However, to forecast the annual rainfall a suitable model is required that can be give reliable and accurate forecast for varied purposes. Therefore, based on the significant lag-3 auto correlation coefficient and partial correlation coefficient for the purpose of forecasting ARIMA model has been selected and the analysis showed that ARIMA (3,0,3) with initial guess value of 0.1 fits well the data set. The forecast accuracy can be further improved by fixing 10% margin either side of the long period average (1979-2007). The model forecast verification using various skill scores yielded very promising results with an accuracy of over 69%.

6. The author sincerely thanks the referee for his valuable comments and suggestions for making this research paper appealing and effective text and also making necessary improvements.

References

Charney, J. C. and Shukla, J., 1981, "Predictability of monsoons", Lighthill and Pearce (eds), Monsoon dynamics, Oxford, Cambridge University Press, 99-109.

Curtis, S. and Adler, R., 2000, "ENSO indexes based on patterns of satellite-derived precipitation", *Journal of Climate*, **13**, 2786-2793.

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S. K., Hnilo, J. J., Fiorino, M. and Potter, G. L., 2002, "NCEP-DEO AMIP-II Reanalysis (R-2)", *Bulletin of the Atmospheric and Meteorological Society*, 1631-1643.

Mooley, D. A. and Parthasarathy, B., 1984, "Fluctuations in all Indian summer monsoon rainfall during 1871-1978", *Climate Change*, **6**, 287-301.

Parthasarathy, B., Munot, A. A. and Kothawale, D. R., 1994, "All-India monthly and seasonal rainfall series: 1871-1993", *Theoretical and Applied Climatology*, **49**, 217-224.

Seetharam, K., 2008, "Climate Change Scenario over Gangtok", *Mausam*, **59**, 3, 361-366.

Seetharam, K., 2005, "Influence of Equatorial Pacific SST Anomalies of NINO 3.0 and 3.4 on Indian sub-divisional rainfalls", *Mausam*, **57**, 3, 517-523.

Thapliyal, V., 1990, "Long range prediction of summer monsoon rainfall over India: Evolution and development of new models", *Mausam,* **41**, 339-346.

Thapliyal, V., 1981, "ARIMA model for long range prediction of monsoon rainfall in peninsular India", Meteorological Monograph, Climatology, India Meteorological Department, **12**, 1-21.

Wolter, K. and Timlin, M. S., 1998, "Measuring the strength of ENSO - how does 1997/98 rank?", *Weather*, **53**, 315-324.

K. SEETHARAM

Meteorological Center, Gangtok, Sikkim – 737 103, India
(*27 August 2008, Modified 1 April 2009*)
e mail : kovelasitaram99@rediffmail.com