MAUSAM, **61**, 2 (April 2010), 225-228

551.577 : 551.501.777

Rainfall models – a study over Gangtok

K. SEETHARAM

Regional Meteorological Centre, Alipore, Kolkata, India

(*Received 21 April 2008, Modified 13 October 2009*)

e mail : kovelasitaram99@rediffmail.com

सार – इस शोध पत्र में वर्ष 1957 से 2005 तक की अवधि में गंगटोक स्टेशन के 49 वर्षों के ऋतु एवं वार्षिक वर्षा के कुल 14 आँकड़ा समूहों को लेकर वर्षा वितरण की संभावनाओं को निर्धारित करने का प्रयास किया गया है जिसमें जनवरी से दिसम्बर के दौरान मासिक वर्षा के निर्धारण के लिए वक्रों की पियरसोनियन प्रणाली का उपयोग किया गया है। इस अध्ययन से यह पता चला है कि जुलाई की मासिक वर्षा और ग्रीष्म मानसून ऋतु की वर्षा के वक्र पियरसोनियन प्रणाली से मेल नहीं खातें है परन्तु गंगटों स्टेशन पर अन्य महीनों की मासिक वर्षा तथा वार्षिक वर्षा पियरसोनियन टाइप–I के वितरण से मेल खाते है जिसे दूसरे शब्दों में समान वितरण कहा जा सकता है। शून्य परिकल्पना के लिए एण्डर्सन–डार्लिंग टेस्ट का अनुप्रयोग किया गया है। यह टेस्ट शून्य परिकल्पना के अनुरूप पाया गया है। इस शोध पत्र में आँकड़ा समूहो के वितरण और उनके वितरण की संभावनाओं का अध्ययन किया गया हैं।

ABSTRACT. In this paper, the Pearsonian system of curves were fitted to the monthly rainfalls from January to December, in addition to the seasonal as well as annual rainfalls totalling to 14 data sets of the period 1957-2005 with 49 years of duration for the station Gangtok to determine the probability distribution function of these data sets. The study indicated that the monthly rainfall of July and summer monsoon seasonal rainfall did not fit in to any of the Pearsonian system of curves, but the monthly rainfalls of other months and the annual rainfalls of Gangtok station indicated to fit into Pearsonian type-I distribution which in other words is an uniform distribution. Anderson-Darling test was applied to for null hypothesis. The test indicated the acceptance of null-hypothesis. The statistics of the data sets and their probability distributions are discussed in this paper.

Key words − Pearson's system of curves, Probability distribution function, Uniform distribution, Percentiles, Quartiles, Probability.

1. Introduction

Water is most essential for survival of life on this planet. Water resource management is required in the modern societies to avoid inconvenience and loss to life and property. Excess rainfall leads to floods, less rainfall leads to droughts and heavy rainfall in certain vulnerable areas often triggers landslides and flash floods. Irregular rainfalls cause water supply and drainage management problems. As the rainfall is the source of water the study of rainfall pattern becomes at most important.

Understanding the rainfall pattern helps in better water resource management. Estimation and prediction of the amount of rainfall in time and space is a problem of fundamental importance in many applications in agriculture, hydrology, and ecology. In tropical and subtropical regions, food production and hydropower generation are a function of rainfall variability. The understanding of the rainfall process under the present context of climate variability and change is of a great value.

Earlier the rainfall data sets have been analyzed using statistical distributions for the purpose of water resource management. Generally, gamma distribution, lognormal distribution, Weibull distribution and exponential distributions are utilized in addition to many other distributions for fitting to rainfall of varying time scales like, hourly, daily, monthly and annual. The theoretical distributions fitted would help in understanding the nature of the rainfall pattern and rainfall estimation, in turn, better management of water resources.

Always a search is on for a best fitting distribution model. The estimation of parameters of the selected model is done by various methods and goodness of fit employed to see the suitability of the model.

2. Data

In this paper, to study the theoretical model for monthly, annual and southwest monsoon rainfall, the available monthly, annual rainfall data for the 1957 - 2005 has been collected from the records of meteorological

TABLE 1

Showing statistical parameters in respect of monthly/annual/southwest monsoon rainfalls (1957-2005)

Month/Parameter	Mean	Median	Mode	SD	CV	ML	β_1	β_2	γ_1	γ_2	Kurtosis	Skewness	Kappa	Type
January	34.9	30	20.2	24.1	69%	21.4	2.00	5.02	1.41	2.02	Lepto	Right	-0.03346	Type 1
February	65.8	75	93.4	39.8	61%	50.3	0.13	2.34	0.36	-0.66	Platy	Left	-0.00086	Type 1
March	125	125	125	59.8	48%	123.5	0.08	2.62	0.29	-0.38	Platy	Right	-0.00034	Type 1
April	299.3	315	346.4	135.2	45%	252.2	0.45	3.05	0.67	0.05	Lepto	Left	-0.00094	Type 1
May	536.2	525	502.6	169.8	32%	478.2	0.69	3.08	0.83	0.08	Lepto	Right	-0.00122	Type 1
June	633.9	637.5	644.6	192	30%	556.0	0.77	3.77	0.87	0.77	Lepto	Left	-0.0012	Type 1
July	658.2	650	633.7	110.4	17%	676.5	0.03	2.82	0.16	-0.18	Platy	Right	-6.8E-05	Type 1
August	578.7	575	567.7	125.9	22%	594.4	0.19	2.47	0.43	-0.53	Platy	Right	-0.0004	Type 1
September	467.6	462.5	452.3	142.8	31%	422.1	0.82	3.51	0.91	0.51	Lepto	Right	-0.00176	Type 1
October	172.7	140	74.7	116.5	67%	137.1	1.20	3.27	1.10	0.27	Lepto	Right	-0.00357	Type 1
November	43.8	15	ND	46.5	106%	16.6	2.81	4.52	1.68	1.52	Lepto	Right	-0.03017	Type 1
December	24.3	22.5	18.8	20.7	85%	10.9	1.74	4	1.32	1	Platy	Left	-0.0329	Type 1
SW	2324	2375	2477	257.4	11.1%	2341.9	0.01	3.2	0.08	0.2	Lepto	Left	-9.8E-06	Type 1
Annual	3627.6	3750	3994.9	352.9	9.7%	3611.0	0.01	2.48	0.08	-0.52	Platy	Left	-7.2E-06	Type 1

Note : SD – Standard Deviation; CV – Coeficient of variation; ML – Most likely value; ND – Not defined; SW – Southwest monsoon

center, Gangtok and the same data was used in the computations.

3. Methodology

Generally, in statistics, the sample set of observations presumed to represent the entire population from which the sample set of observations were derived. However, in certain cases, the sample set of observations may not entirely represent the parent population especially when the distorted frequency distribution occurs. So as to determine the frequency curve, 'curve fitting' techniques are being utilized all over. Therefore, a generalized form of curve fitting has become necessity. One such system of frequency curves is the 'Pearson's system of frequency curves'. The shape of the curve in this system is determined by a single criterion known as 'criterion Kappa [Greek alphabet usually shown by symbol (κ)].

The Pearson's system of curves is represented by the generalized equation

$dy/dx = y(x\text{-a})/b_o + b_1x + b_2x^2$ in which the constants a, b_o, b_1 and b_2 calculated from the given data set.

The Pearson's Main Type I curve is obtained when the roots of the quadratic equation are real and of opposite sign and the Kappa (κ) is less than zero or negative.

The standard form of the Pearson's main type I curve is

$f(x) = y_o(1+x/a_1)^{m1}(1-x/a_2)^{m2}$ $-a1< x <a2$ in which the a_1 & a_2 are the roots of the quadratic equation and $m1 = a_1/b_2\,(a_1 + a_2)$ and $m2 = a_2/b_2\,(a_1 + a_2)$.

$y_o = a_1^{m1}a_2^{m2}/(a1 + a2)^{m1+m2+1}$ B (m1 + 1, m2 + 1). Where 'B' is the Beta distribution function. Finally,

$f'(x) = f(x)\,(x\text{-a})/b_0 + b_1x + b_2x^2$

The Anderson-Darling to test the null-hypothesis is done by

$A^2 = -N - S$ where 'N' is the sample size and 'S' is computed based on

$$S = \sum_{k=1}^{N} 2k\text{-}1/N[\ln F(Y_k) + \ln F(Y_{k+1} - k)] \text{ and then}$$

$$A^{*2} = A^2\,[1 + 0.75/N + 2.25/N^2]$$

If the value of A^{*2} is greater than 0.752 then the null-hypothesis is rejected.

Here '$F(Y_k)$' is the cumulative distribution function.

TABLE 2

Showing Percentiles and Quartiles with coefficient of dispersion in respect of monthly rainfall/southwest monsoon/annual rainfall (1957-2005)

Month/ parameter	Percentiles										Quartiles		Coefficient of dispersion
	10	20	30	40	50	60	70	80	90	100	1	3	
January	4.9	8.9	23.6	26.7	28.3	33.7	37.4	44.1	63.1	110.2	15.0	41.9	0.47
February	13.2	24.1	38.9	48.9	62.2	70.7	85.1	96.8	110.5	173.2	34.0	90.5	0.45
March	58.9	79.7	94.7	105.6	126.4	134.5	149.8	178.9	217.4	263.9	89.7	156.9	0.27
April	148.0	181.4	212.3	237.7	265.4	327.4	343.6	410.9	471.0	688.1	204.4	378.3	0.30
May	341.4	374.8	423.1	458.1	490	548.3	592.5	637.1	764.4	950.9	400.9	620.5	0.21
June	413.7	474.4	515.7	565.4	631.6	663.5	691.4	726.1	876.3	1231.3	501.5	694.1	0.16
July	524.4	555.4	597.0	641.1	661.3	697.7	711.0	743.7	795.9	861.1	591.4	728.0	0.10
August	425.0	479.9	515.0	560.4	588.5	619.0	641.7	680.4	735.5	779	499.6	654.5	0.13
September	310.5	347.6	373.4	404.1	439.8	485.0	529.8	558.6	632.5	881.9	357.6	555.6	0.22
October	51.1	76.4	97.7	123.8	147.5	161.2	223.5	262.1	345.0	488	88.3	233.6	0.45
November	01.3	4.0	7.5	16.8	20.2	32.7	37.5	56.0	128.0	173.5	4.4	44.6	0.82
December	0.5	1.7	8.6	11.3	16.2	18.6	27.4	34.5	51.2	83.8	2.7	32.3	0.85
Annual	3137.0	3296.8	3422.8	3489.1	3641.1	3684.3	3812.3	3899.1	4052.1	4422.5	3369.6	3860.1	0.07
SW Monsoon	2016.2	2128.8	2183.8	2253.7	2315.9	2391.2	2472.2	2525.6	2658.5	2929.4	2171.4	2484.6	0.07

4. Results and discussion

In order to study the daily rainfall data a first order and second-order Markov chain model was fitted in which the transition probabilities vary with time of the year and also from place to place Coe and Stern (1982). In practice, it is seen that generalized linear models are fitting to rainfall data and Gamma distribution is most commonly used for modeling rainfalls. Das and Sharda (2005) in their study of weekly rainfall data in a sub-humid climate of India, fitted two or three parameter probability distributions and compared the same to identify the most appropriate distribution that describes weekly rainfall data and found that no single probability distribution is appropriate even though Weibull distribution has been found to fit most of the data sets. Further, the above study indicated that Gamma probability distribution was found to grossly approximate the underlying process. Saralees Nadarajah *et al*. (2007) applied the skewed Bessel distribution function to the rainfall data for Orlando, Florida and derived the estimation procedure by the method of maximum likelihood.

Generally, in statistics many often encounter distributions other than normal distribution in which the data is distributed evenly around the mean. The types of distributions that are seen are skewed in nature. The shapes of the curves can be 'U' type or even 'J' type. The skewed distributions can be right or left. The shape of the distribution is mostly described by two parameters known as 'Skewness' and 'Kurtosis'. In this paper, various statistical parameters *viz*., mean, median, mode, coefficient of skew ness based on moments, kurtosis, most likely value has been derived from the monthly, southwest monsoon season and annual rainfall data sets. The results are presented in Table 1. From Table 1, the mostly likely rainfall values from the months of January to December, annual and summer monsoon season are 21.4 mm, 50.3 mm, 123.5 mm, 252.2 mm, 478.2 mm, 556 mm, 676.5 mm, 594.4 mm, 422.1 mm, 137.1 mm, 16.6 mm, 10.9 mm, 3611.0 mm and 2341.9 mm respectively. Further, it is seen that that the mean and median values are different and mode is ill defined. From the figures of standard deviation (Coefficient of variation in percentage) it can be said that highest percent of coefficient of variation was observed for the months of November and December *i.e*., winter rainfall. Lowest percent of coefficient is seen for annual followed by southwest monsoon season. The four required moments were calculated and Charlier's checks were applied to see the accuracy of the moments. The results indicated that the calculated moments were accurate and Sheppard's corrections were not applied. The Pearson's four coefficient based on first four moments (β_1, β_2, γ_1, γ_2) &

'β_2' are shown in Table 2. The 'β_2' values are greater than '3' for the months of January, April, May, June, September to December and for other months it is less than three. In case of southwest monsoon season and annual, the values are greater than three and less than three respectively. The 'γ_2' values are negative for the months of February, March, July, August and annual season. The same was positive for other months and southwest monsoon season and for all other months the values are positive. The monthly rainfall data sets for the months of January, April, June and from September to December indicated 'Leptokurtic' type of distribution and the remaining monthly rainfall data sets for months of February, March, May, July, August, southwest monsoon seasonal rainfall data set and annual rainfall data set indicated 'Platykurtic' type of distribution. The Skewness indicated that distribution is left skewed for February, April, June, December, summer monsoon season and annual season. The Skew ness, further, indicated that the distribution is right skewed for January, March, May, July, August, September and October.

Further, to describe the distribution of monthly rainfalls from January to December and also for southwest monsoon season and annual rainfalls, 10 different percentiles, that is, 10^{th}, 20^{th}, 30^{th}, 40^{th}, 50^{th}, 60^{th}, 70^{th}, 80^{th}, 90^{th} and 100^{th} percentiles computed and depicted in Table 2. Similarly, the first and third quartiles and coefficient of dispersion based on quartiles are also presented. The coefficient of dispersion indicates the rainfall variation is less for southwest monsoon and annual rainfalls and highest for the months of November & December.

Further, Karl Pearson's generalized frequency curves has been fitted to all the monthly, annual and southwest monsoon rainfall data sets to see which type of curve fits the best to the observational data sets. The four coefficients of the frequency curves namely, a_0, b_0, b_1 and b_2 were evaluated from the set of observations. As the frequency curve depends on the roots of the quadratic equation in terms of b_o, b_1 and b_2, the roots of the quadratic equation were obtained. The roots were real, unequal and opposite in sign. Moreover, the type of the curve in Pearson's system of curves can be determined by a single parameter known as 'Kappa' (Greek alphabet κ) indicating that they are all negative. This clearly shows that the Pearson's Main Type I curve fits for all the data sets except for the months of July, summer monsoon and annual seasons.

5. Conclussions

Application of Pearson's system of curves indicated uniform distribution (Pearson's type-I distribution) a best fit for the monthly rainfalls.

(*i*) Equal probabilities are indicated by the study in case of monthly rainfalls.

(*ii*) The monthly rainfall of July as well as the summer monsoon seasonal rainfall did not fit in to any of the Pearson's system of curves and normal distribution for these two is presumed.

(*iii*) The Anderson-Darling test indicated acceptance of null-hypothesis for all months from January to June and August to December as well as for the annual rainfalls.

(*iv*) The probability distribution can be applied to make computation as indicated in this paper.

Acknowledgements

The author expresses his deep gratitude for the referee comments on this research paper that enabled suitable modifications and improvements in the paper.

References

Coe, R. and Stern, R. D., 1982, "Fitting models to daily rainfall data", *Applied Meteorology*, **2**, 1024-1031.

Das, P. K. and Sharda, V. N., 2005, "Modelling weekly rainfall data for crop planning in a sub-humid climate of India", *Agricultural Water Management*, **75**, 2, 120-138.

Saralees, Nadarajah, Srivastava, H. M. and Gupta, A. K., 2007, "Skewed Bessel function distribution with application to rainfall data", *Statistics*, **41**, 4, 333-344.