International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XII **Month of publication:** December 2017

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Shift in Date of Onset of the Southwest Monsoon in India from 146 Years of Data Exploring the Change in Distribution Using Non-Parametric Polynomial Estimators

T. Unnikrishnan[1], P. Anilkumar[1] and Kadambot H. M. Siddique[2]
[1]*Dept. of Statistics, Farook College, University of Calicut*
[2]*The UWA Institute of Agriculture and School of Agriculture and Environment, The University of Western Australia,*

Abstract**: *The accurate date of onset of the southwest monsoon in India is unpredictable due to its complexity in estimation. However, it appears that in recent years the onset of the southwest monsoon has been occurring earlier than 'normal'. Therefore, we studied the change in distribution of the date of onset of the southwest monsoon in India using a recently developed, highly efficient non-parametric method, the 'non-parametric polynomial estimator'. The polynomial estimator was used to estimate the non-parametric distribution of the frequency of occurrence of the onset of the southwest monsoon in India from 1870 to 2015. The 146-year data was split in to three time periods and analyzed; removing the middle time periodshoweda clear shift in the median of the observed values towards an earlier date.***

Keywords: Onset of South West Monsoon, Distribution Fitting, Non-parametric Polynomial Estimation, Climate .

I. INTRODUCTION

Random variables are characterized by probability density functions. Hence, the estimation of probability density function has great importance. Parametric density estimation pre-assumes shape of the distribution, but not the parameters. Situations may arise where parametric density estimation will not provide an accurate estimate of a probability density function. In these circumstances, non-parametric estimation is important as it does not assume a prior distribution. Here, the data will speak for themselves. Silverman (1981) reported that density estimates are useful at all stages of statistical treatment of data. Density estimators' indicate multimodality, skewness, or dispersion of the data in the exploratorys tage. For confirmatory purposes, density estimators can be used, for instance, in non-parametric discriminant analysis (Prakash Rao, 1983). There is extensive literature on the asymptotic properties of the kernel estimate. Aweak uniform consistency is established for kernel estimates (Prakash Rao, 1983). A different and somewhat weaker type of consistency is discussed in Deveroy and Gyorfi(1985). Density estimators provide the best information—transformations of the given datafor presentation. Rudin (1976) introduced a polynomial to prove the Stone–Weierstrass theorem. The same polynomial was used todevelop the method of the polynomial estimator for fitting distributions using the nonparametric technique. The proposed estimator belongs to the general class of estimators. The appropriate degree of the polynomial (d)will be fixed based on the sample using an objective criterion. The theory of the non-parametric polynomial estimator for fitting distribution was tested on the arrival of the southwest monsoon season in India. This monsoon accounts for 80% of the rainfall in India. The coastal state of Kerala is the first Indian state to receive rain from the southwest monsoon, with the 'normal' onset being 1 June (±7 days).The southwest monsoon reaches Kolkatta and Visakhapattanam on 7 June, Mumbai on 10 June, Delhi on 3 July, and the extreme western part of Rajasthan on 15 July. The onset of the southwestmonsoon is criticalfor Indian agriculture as it irrigates more than half of India's cropping area. The earliest onset of the southwest monsoon was recordedon 11 May 1918 and the latest on 18 June 1972. Many studies have analyzed the date of onset of the southwest monsoon in India. For example, Subbaramayya and Bhanu kumar(1978) and Subbaramayya et al. (1984) discussed different features related to the onset date of the southwest monsoon in India. Based on these studies and the climatic conditions, criteria for onset of southwest monsoon are predicted. Ananthakrishnan and Soman (1988) collated and analyzed data from the Indian Meteorological Department on the date of onset of the southwest monsoon in India from 1901–1980and found that the mean onset date from1901–1940 was four days later than that from1941–1980. Unnikrishnan and Ajitha (2009) developed a time-series model for forecasting the onset of the southwest monsoon in Kerala using data from 1870–2009. All of the above studies were based on the concepts of parametric statistical analysis. Parametric tests are usually based on assumptions, which may not be valid in certain situations. In this study, we used a

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

novel approach incorporating anon-parametric method to analyze 146 years of data on the onset date of the southwest monsoon in Kerala, India.

II. MATERIALS AND METHODS

The distribution of onset date of the southwest monsoon in India from 1870 to 2015 was analyzed using non-parametric statistical tools recently developed for situations where parametric tests miss the mark. Onset dates from 1870 to 2015 were collected from the Department of Agricultural Meteorology, Kerala Agricultural University. The range of dates spanned 39 days; the earliest monsoon onset on 11Maywas given the code '0', and the latest monsoon onset on 18 June given the code'38'. The frequency of each date of onset was recorded. Nonparametric tests were applied to determine whether the median varied significantly between the observed dates of onset corresponding to the50 years of data at the beginning of the sample period and the 50 years of data at the end of the observed period. To do this, the data from 1921 to 1964were removed, and the two groups either side of these dates were analyzed to determine any change in median date of onset of the southwest monsoon in India. The Kolmogorov two-sample test was applied to determine whether the median of two samples, i.e., median of the dates of onset from 1870 to 1920 and median of the dates of onset from 1965 to 2015, varied significantly. To test whether the sample variancesdiffered significantly, Levene's test for homogeneity of variance was applied using SPSS.

he newly developed tool for finding the probability density function using a nonparametric polynomial estimator (Hamza, 2009) was tested using the following formula:

$$f_n(x) = \frac{1}{n}\sum_{n=1}^{n} c_d\left[1-(X_i - x)^2\right]^d, \text{ for } x \in [0,1],\ X_i \text{ the } i^{th} \text{ observation.}$$

The c_d values were calculated for d=1, 2, 3… using:

$$c_d = \frac{1}{2}\frac{(2d+1)!}{2^{2d}(d!)^2}$$

where d is the degree of the polynomial and, in fact, the smoothing parameter.The estimation process is similar to that of the kernel bandwidth parameter of a kernel density estimator (Silverman, 1986). In kernel density estimation, the bandwidthparameter is determined using two procedures: (i) least square cross-validation, and (ii) likelihood cross-validation. In kernel estimation, the bandwidth decreases with increasing sample size,whereas in polynomial estimation, the degree of the polynomial increases as the sample size increases. Polynomial estimation is more robust than kernel density estimation, as it is more precise as the sample size increases. Moreover, bandwidth is highly sensitive, but the degree of the polynomial did not affect the estimation of smaller errors in d values.

In polynomial estimation, likelihood cross-validationis used to maximize CV(d) alone, where $CV(d) = \frac{1}{n}\sum_{i=1}^{n}\log\left(\hat{f}_{-i}(x_i)\right)$. Since the calculation of c_d values can be difficult due to factorials in the expression, a recurrence formula was determined, where $c_d = c_{d-1}*(2d+1)/2d$. The choice of d was based on the maximization of CV(d) values, calculated using maximum likelihood estimation of $\hat{f}_{-i}(x_i)$ by leaving the i^{th} variable in the frequency data. The d–CV(d) curves were plotted and the maximum CV(d) point was selected. The value of d, corresponding to the unique maximum, is the optimum parameter for the distribution. Using the d and c_d values, the probability density function was determined. This process is much easier than kernel band width estimation and can be estimated simply in Microsoft Excel.

If the domain lies in the interval [a,b], then the formula will be:

$$Q(x) = \frac{1}{n}\sum_{i=1}^{n} c_d\left[1 - \left(\frac{X_i - x}{b-a}\right)^2\right]^d$$

III.RESULTS AND DISCUSSION

The time-series plot showed a cyclical pattern in the data (Fig. 1) as did the 30-year moving average plot (Fig. 2). The cyclical behavior of the time-series data shifted the median value. At times, the periodic behavior can repeat at regular intervals. Also, there may be an upward or downward trend in the data, and hence the distribution may vary over the long run. Here, we aimed to test whether there was any shift in the distribution of the onset of the southwestmonsoon in India.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

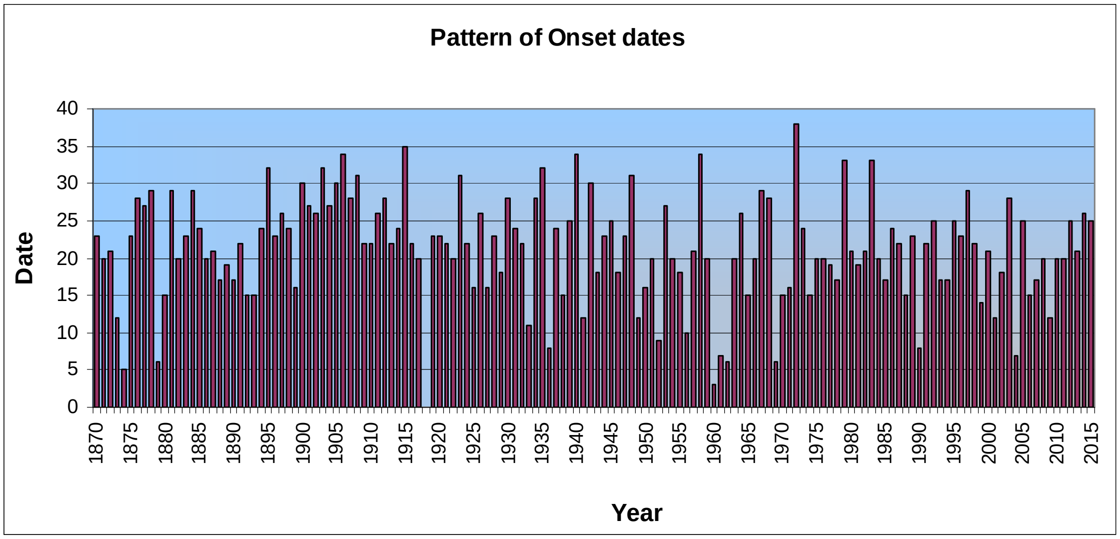

Figure1. Time-series pattern for the date of onset of the southwest monsoon in India.

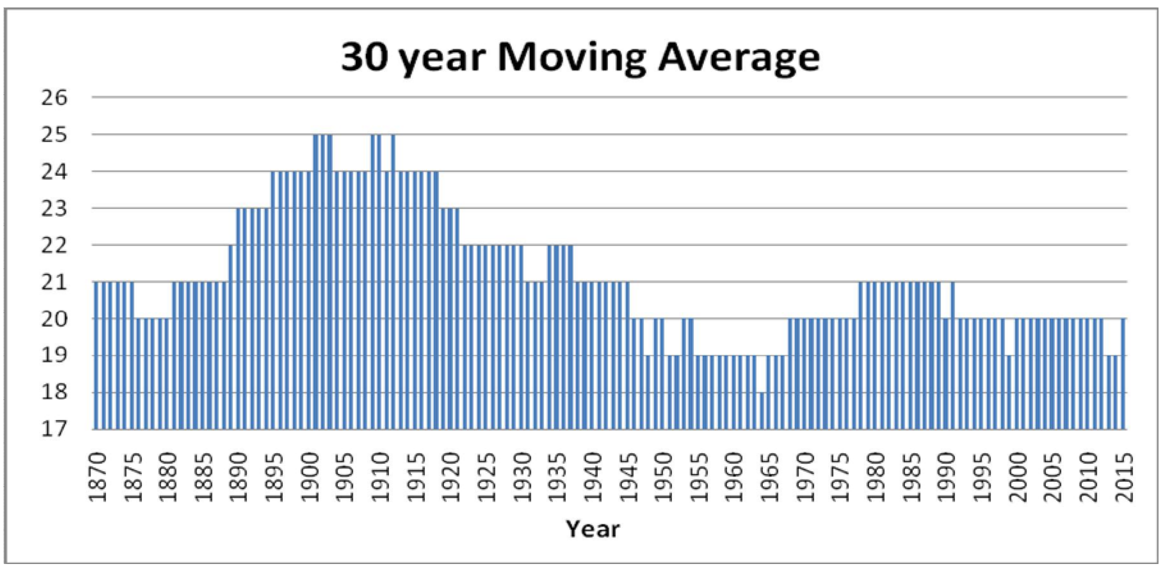

Figure2. The 30-year moving average of the date of onset of the southwest monsoon in India.

The two groups of data, with 51 observations each, were analyzed using the Kolmogorov–Smirnov one-sample test for normality. The test rejected the null hypothesis with Z = 1.403 and p= 0.039 for the data from 1870–1920, and Z = 1.539 and p= 0.018 for the data from1965–2015. The two groups were analyzedto determine whether there were any significant changes in the distribution using the Kolmogorov–Smirnov two-sample test. While Levene's test for homogeneity of variances showed no significant variation in data, the Komogorov–Smirnov two-sampletestshowed a change inthe median, with Z = 1.386 and p=0.043.

To determinethe distribution of onset dates from 1870 to 2015 using nonparametric estimation, CV(d) wascalculated for d=1,2,3,… using the corresponding c_d and the frequency of onset data.Thed–CV(d) curve identified an optimum d of 83,andacorresponding $c_d$of5.163 (Fig. 3). Using these values,the distribution of the data using polynomial estimation and plotted in Figure 4 is:$Q(x) = \frac{1}{146}\sum_{i=1}^{146} 5.163\left[1 - \left(\frac{Xi - x}{38}\right)^2\right]^{83}$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

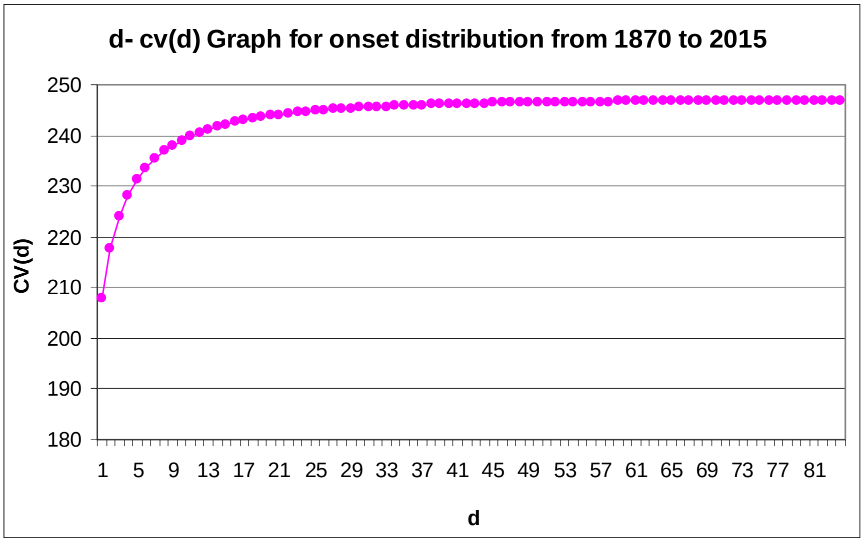

Figure3. d–CV(d) curve for date of onset of southwest monsoon in India from 1870 to 2015.

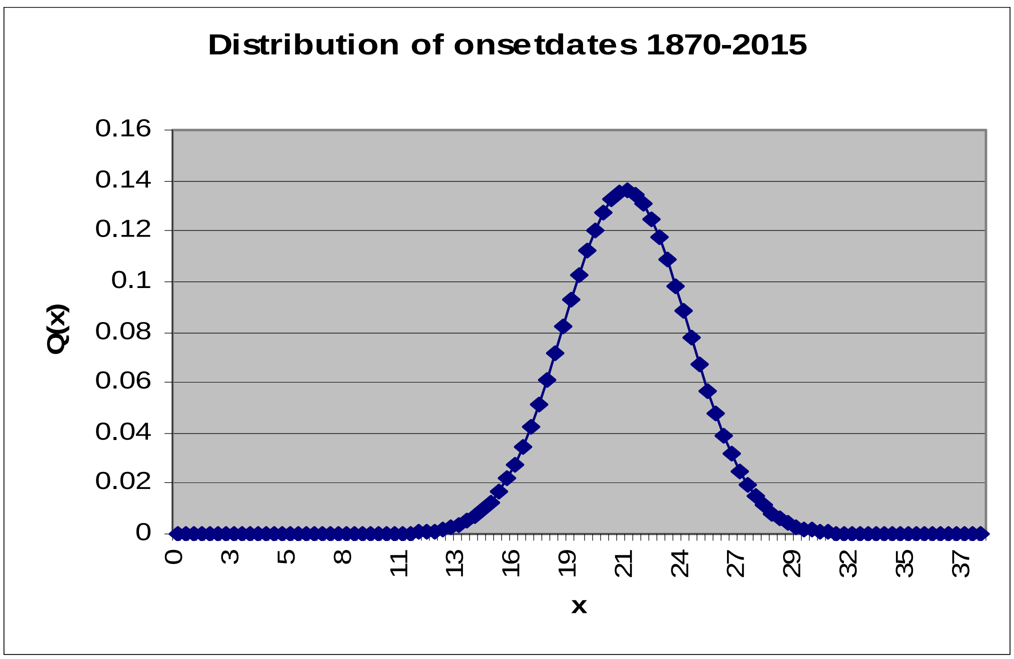

Figure4. Distribution of onset date of the southwest monsoon in India using $Q(x) = \frac{1}{146}\sum_{i=1}^{146} 5.163\left[1 - \left(\frac{Xi - x}{38}\right)^2\right]^{83}$

To determine the distribution of the data from 1870–1920 using nonparametric estimation, CV(d) was calculated for d=1,2,3,… using the corresponding c_d and the frequency of onset data.Thed–CV(d) curve identified an optimum d of 57 and a corresponding $c_d$of4.287 (Fig. 5). Using these values, the distribution of the data using polynomial estimation and plotted in Figure 5 is:

$$Q(x) = \frac{1}{51}\sum_{i=1}^{51} 4.287\left[1 - \left(\frac{Xi - x}{38}\right)^2\right]^{57}$$

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

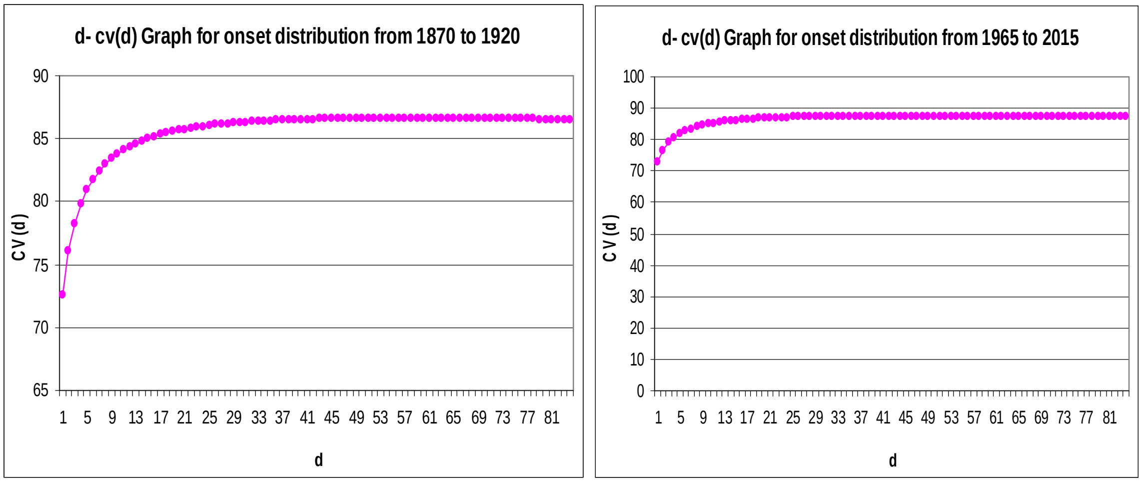

Figure5. The d–CV(d) curve for onset date of the southwest monsoon in India.

Similarly, to determine the distribution of the data from 1965–2015 using nonparametric estimation, CV(d) was calculated for d=1,2,3,… using the corresponding c_d and the frequency of onset data.Thed–CV(d) curve identified an optimum d of48andacorresponding$c_d$of 3.939 (Fig. 5). Using these values, the distribution of the data using polynomial estimation plotted in Figure 6 is: $Q(x) = \frac{1}{51}\sum_{i=1}^{51} 3.939\left[1-\left(\frac{Xi-x}{38}\right)^2\right]^{48}$

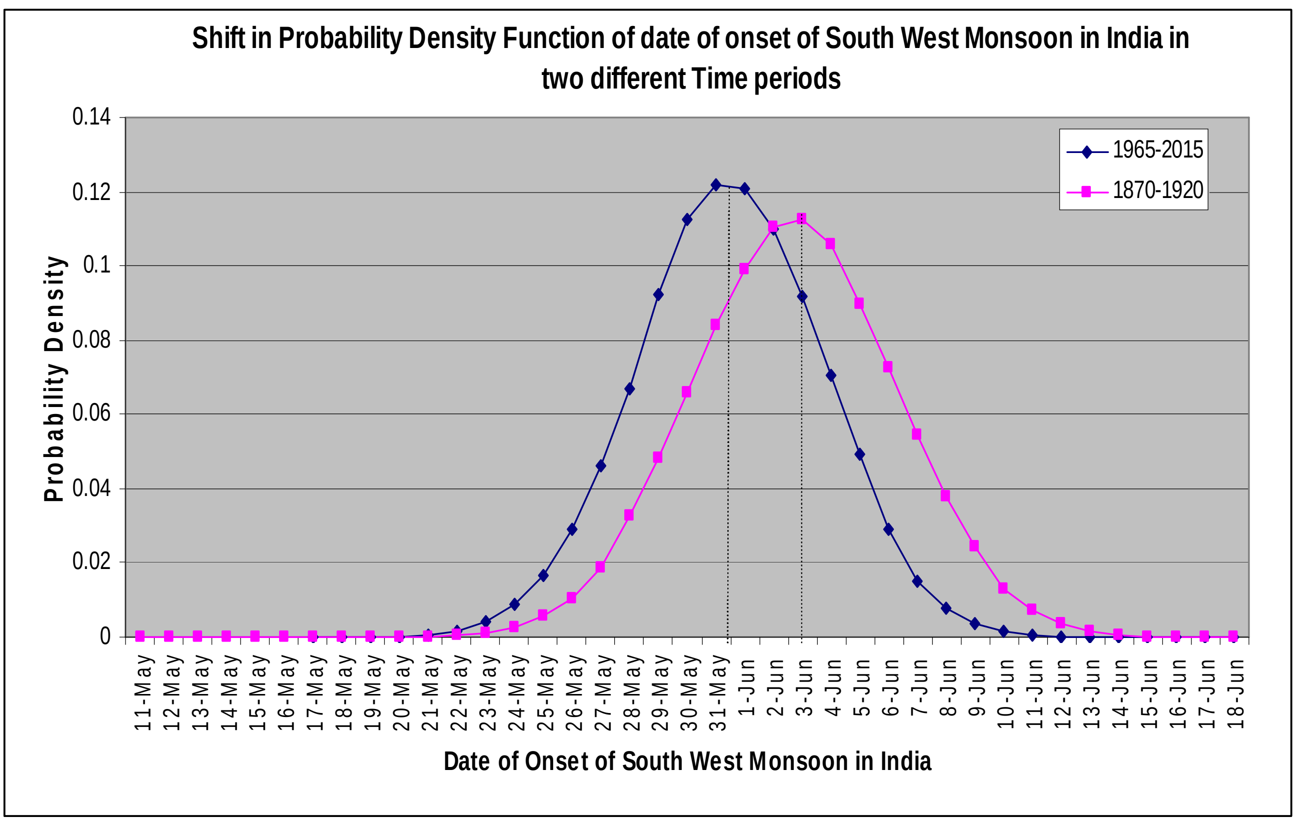

Figure6. The shift in the distribution of onset date of the southwest monsoon in India.

Superimposing the two graphs, a clear shift in the onset date of the southwest monsoon in India to earlier than 'normal' is evident Fig. 6).

IV. CONCLUSIONS

The Kolmogorov–Smirnov two-sample test showed a shift in the median value of the onset date of the southwest monsoon in India to earlier than 'normal'. Comparing the variance using Lavene's test for homogeneity, there was no significant change in variance. Since the Kolmogorov–Smirnov test showed that the frequency was not normally distributed, the optimum probability density function for fitting the data was determined using a polynomial estimator, which showed a shift in the distribution of onset of the southwest monsoon in India**.** The optimum choice of smoothing parameter d, using the maximum CV(d) score, is unaffected by small changes in the data as the sample size increases. Hence, the approach is recognized as an excellent data-driven method for selecting the smoothing parameter of the proposed estimator. The idea of maximization of CV(d) for finding the value of d was used in the present study to fit the distributions.

V. ACKNOWLEDGMENTS

The authors thank the Department of Agricultural Meteorology, Kerala Agricultural University for providing the data for the study and the Principal, Farook College, Calicut for making arrangements to undertake the research.

REFERENCES

[1] Ananthakrishnan, R., Soman, M.K., 1988. The onset of the southwest monsoon over Kerala:1901–1980.J.Climatol.8, 283–296.
[2] Deveroy,L.,Gyorfi,L.,1985. Nonparametric Density Estimation.Wiley, New York.
[3] Hamza,K.K., 2009. Some contributions to renewal density estimation. Thesis, Department of Statistics, University of Calicut.
[4] Prakash Rao, 1983. *Nonparametric*Functional Estimation. Academic Press, Orlando.
[5] Rudin, W.,1976.Principles of Mathematical Analysis. 3rd Edition, McGraw-Hill, New York.
[6] Silverman,B.W., 1981. Using kernel density estimate to investigate multi modality, J. R. Stat. Soc.43, 833–836.
[7] .Silverman,B.W., 1986.Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.
[8] Subbaramayya, I., Bhanukumar, O.S.R.U., 1978. The onset and the northern limit of the southwest monsoon over India.Metereol. Mag.107, 37–44.
[9] Subbaramayya, I., Babu, S.V., Rao, S.S., 1984. Onset of the summer monsoon over India and its variability.Metereol. Mag.113, 127–135.
[10] Unnikrishnan, T., Ajitha, T.K., 2009. A univariate statistical approach to predict the onset of South West Monsoon in India, in: Proceedings of the National Workshop on Climate and Development, 29 June 2009.Vellayani, Trivandrum, pp. 8–16.