Special Proceeding of the 21st Annual Conference of SSCA held at SV Agricultural College (ANGRAU), Tirupati, during January 29-31, 2019; pp 83-92

Long Memory and Structural Break in Seasonal Rainfall in India

Ranjit Kumar Paul, Dipankar Mitra, A.K. Paul and L.M. Bhar
ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

Received: 24 June 2019; Revised: 29 June 2019; Accepted: 01 July 2019

Abstract

Agricultural performance of a country depends to a great extent on the quantity and distribution of rainfall. Modeling and forecasting of monthly seasonal rainfall across the country is of great concern among the researches. In many situations the datasets may exhibit long memory pattern and break in their structure. Sometimes a stationary short memory process that encounters occasional structural breaks in mean can show a slower rate of decay in the autocorrelation function and misinterpreted as long memory process. This phenomenon is called as spurious long memory. In this paper, we have employed a procedure for estimating the fractional differencing parameter in semiparametric contexts, namely exact local Whittle (ELW) estimator proposed by Shimotsu and Phillips (2005) to analyse seasonal rainfall data sets across different zones of India. The results indicate that some of the series exhibit long memory. Furthermore, an empirical fluctuation process using the ordinary least square (OLS)-based Chow test for the break date is applied. Break dates are detected in North-East and Central-North data sets. Moreover, Qu test (Qu, 2011) has been considered for testing for true versus spurious long memory and no test is found to be significant.

Key words: Long memory; Structural breaks; Spurious long memory.

1 Introduction

Indian agriculture is rain-dependent, with almost two-thirds of the net cropped area being rain-fed. Modeling and forecasting of monthly seasonal rainfall for different metrological zones over different seasons is important for climatic assessment and planning. Any modeling effort on the rainfall data will have to be based on an understanding of the time dependency among the past observations. Over last few decades, Box-Jenkins's Autoregressive integrated moving average methodology (ARIMA) (Box *et al*., 2007) has been efficiently used to get forecasts of the time series having dependency at fewer lags *i.e.* short memory process. But in many situations, the observations separated by distant time lags may be dependent and this phenomenon is commonly known as long memory or long range dependency. Under this situation ARIMA class of short memory model fail to get accurate and reliable forecasts. Moreover, some long memory time series models like Autoregressive fractionally integrated

Corresponding Author: Ranjit Kumar Paul
E-mail: ranjitstat@gmail.com

moving average methodology (ARFIMA)(Granger and Joyeux, 1980) need to be called for capturing the long memory. Hence, detection of presence of long memory in the dataset prior to fitting any forecasting models is important for proper understanding of the underlying pattern. Paul *et al*. (2013) have investigated the modelling of Indian monsoon rainfall data and concluded that wavelet methodology has greater accuracy than that of ARIMA model. Paul *et al*. (2015) investigated the trend in mean temperatures in different agro-climatic zones in India using both parametric and nonparametric methods. Paul and Birthal (2016) have applied advanced statistical approach for describing variability in rainfall in different agro-climatic zones of India. Paul (2017) found the significant presence of long memory in maximum and minimum temperatures in India.

In econometrics and statistics, a structural break is an unexpected change over time in the parameters of regression models, which can lead to huge forecasting errors and unreliability of the model in general. Rainfall data may be subjected to the structural break due to numerous reasons like El-Nino, etc. Ignorance of the break points can lead to serious bias and error in the estimates of the model parameters. Studies by Cheung (1993) and Diebold and Inoue (2001)have shown that there is a bias in favour of finding long memory processes when structural breaks are not accounted for in a time series. Observed long memory behaviour can be due to neglected structural breaks. Before fitting any model the detection of break points of long memory process is essential to capture the long memory pattern over different horizons. Paul *et al*. (2014) investigated structural break in mean temperature in different agro-climatic zones of India. In the last decade, a lot of interest has been paid to the issue of confusing long memory and occasional structural breaks in mean (Diebold and Inoue, 2001 and Granger and Hyung, 2004). Indeed, there is evidence that a stationary short memory process that encounters occasional structural breaks in mean can show a slower rate of decay in the autocorrelation function and other properties of fractionally integrated (I(d)) processes. Therefore, a time series with structural breaks can generate a strong persistence in the autocorrelation function, which is an observed behaviour of a long memory process. On the contrary, long memory processes may cause breaks to be detected spuriously. In this paper, we have used the term ‘true long memory’ to refer to fractionally integrated series. Sometimes real time-series data may exhibit long memory pattern due to possible presence of structural change. This phenomenon is commonly called as spurious long memory. The literature on the tests to distinguish between true long memory and various spurious long memory models has been steadily growing. For example, Berkes *et al*. (2006) and Shao and Zhang (2010) proposed a testing procedure to discriminate a stationary long memory time series from a short-range dependent time series with change points in the mean; Qu (2011) proposes a test in the frequency domain based on the profiled local Whittle likelihood function.

2 Methodology

2.1 Long Memory

Long memory models are statistical models that describe strong correlation or dependence across time series data. This kind of phenomenon is often referred to as “long memory” or “long-range dependence.” It refers to persisting correlation between distant observations in a time series. For scalar time series observed at equal intervals of time that are covariance stationary, so that the mean, variance, and auto-covariances (between observations separated by a lag j) do not vary over time, it typically implies that the auto-covariances decay so slowly, as j increases, as not to be absolutely summable. However, it can also refer to certain non-stationary time series, including ones with an autoregressive unit root, that exhibit even stronger correlation at long lags. Evidence of long memory has often been found in economic and financial time series, where the noted extension to possible non-stationarity can cover many macroeconomic time series, as well as in such fields as astronomy, agriculture, geophysics, and chemistry.

Most of the research works in time-series analysis assume that the observations separated by long time span are independent of each other or nearly so. But in many empirical series it is seen that the distant observations are dependent, though the correlation is small but not negligible. The statistical dependency of any time-series data is generally measured by plotting the ACF of the dataset. Let x_t $(t = 1,2, \ldots, n)$ be an equally spaced, real valued and covariance stationary time-series process so that the mean$\mu = E(x_t)$and lag$-k$ autocovariances (or variance when $k = 0$)

$$\gamma(k) = Cov(x_t, x_{t+k})$$

do not depend on t.

Further, consider that the autocorrelation function of the time-series with a time lag of k is given as

$$\rho_k = cov(x_t, x_{t-1})/var(x_t).$$

The series x_t $(t = 0,1,2, \ldots.)$ is said to have short memory if the autocorrelation coefficient at lag k approaches to zero as k tends to infinity, *i.e.*$\lim_{k\to\infty} \rho_k = 0$.

The autocorrelation functions of most of stationary and invertible (ARMA) time-series process decay very rapidly at an exponential rate, so that $\rho_k \approx |m|^k$, where $|m| < 1$.

For long memory processes, decaying of autocorrelations functions occur at much slower rate (hyperbolic rate) which is consistent with $\rho_k \approx Ck^{2d-1}$, as$k$ increases indefinitely, where C is a constant and d is the long memory parameter. The autocorrelation function of a long memory process exhibits persistency structure which is neither consistent with an *I(1)* process nor an*I(0)* process.

There are different approaches for estimating long memory parameter$d$; these are R/S statistic (Hurst, 1951), ACF plot, Maximum likelihood method of estimation (MLE) (Beran, 1995), GPH (Geweke and Porter-Hudak, 1983), exact local Whittle estimator (ELW) of Shimotsu and Phillips (2005). In this study we have used ELW estimator to estimate long memory parameter as it is consistent and asymptotically normal and provides estimate of stationary as well as non-stationary processes. An interesting application of long memory in climate data can be found in Paul and Anjoy (2018).

2.2 Structural Break

A structural break occurs when there will be a sudden change in a time series or a relationship between two time series. Literally, structural change can be described as fundamental shift in the structure of the series under consideration which may be due to economic growth, policy decisions, revolution, etc. This change could involve a change in mean or a change in the other parameters of the process that produce the series. Being able to detect the structural changes of the time series can give insights into the problem which are under study. Structural break tests help to determine when and whether there is a significant change in the dataset. If the presence of breaks is completely ignored, forecasts become poor and inaccurate. Therefore, detection of structural break is prime importance prior to the analysis of time-series data. For detection of structural break Cumulative sum (CUSUM), Chow, Andrew's LR tests are commonly used.

2.3 Spurious Long Memory

Sometimes real time-series data may exhibit long memory pattern due to possible presence of structural change. This phenomenon is commonly called as spurious long memory. There are mainly two situations - one is the structure of the time-series process might be mistaken as long memory due to presence of structural break and next one is the co-existence of long memory and structural break in the given data set. Two important features of a long-memory process are that its spectral density at the origin is unbounded and that its autocorrelation function decays at a hyperbolic rate at long lags. But these features also can be present for a short-memory process affected by a regime change or a smooth trend, leading to so-called "spurious" long memory. This has been widely documented (Perron and Qu, 2010). Recent contributions for tests against true long memory (in the sense of fractional integration) include Ohanissian et al. (2008), Perron and Qu (2010), Qu (2011). Qu (2011) proposes a test in the frequency domain based on the profiled local Whittle likelihood function. The test turns out to be powerful when the series is solely generated by random level shifts, non-monotonic trends or Markov regime switching.

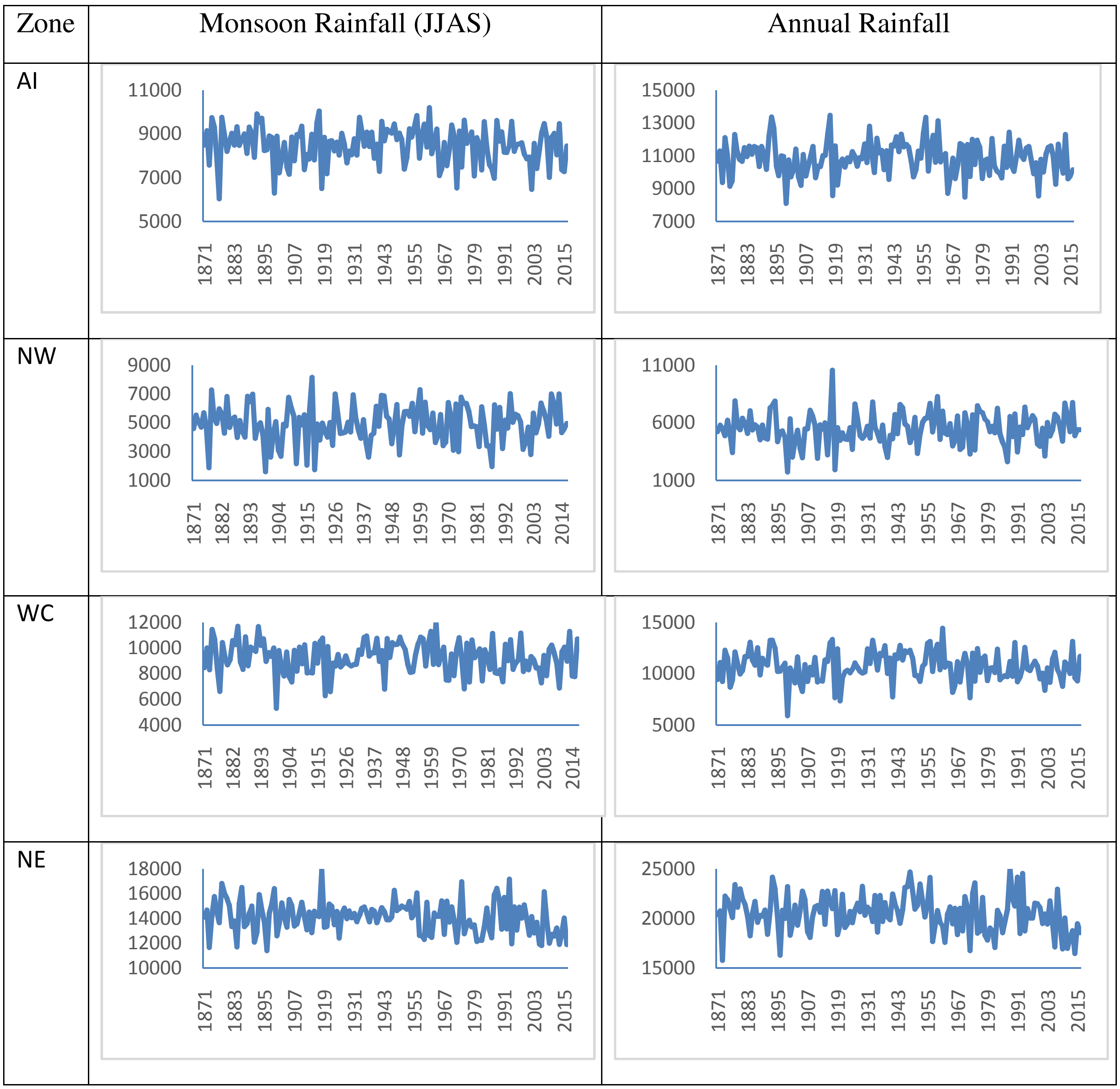
3 Illustration

3.1 Dataset

In this study monthly rainfall data corresponding to five zones of India *viz*., North-West (NW), West-Central (WC), North-East (NE), Central-North (CN), Peninsular (P) as well as all India (AI) are collected from Indian Institute of Tropical Meteorology (www.tropmet.res.in), Pune, India for analysis. The data set comprises of monthly rainfall over 146 years (from 1871 to 2016), measured in mm. The monthly data is accumulated to obtain seasonal data corresponding to four seasons *viz*., January and February (JF); March, April and May (MAM); June, July, August and September (JJAS); October, November and December (OND). Finally, the annual rainfall series (ANN) is obtained by summing over 12 months. The pattern of monsoon rainfall *i.e*. total rainfall of the season JJAS along with annual rainfall in different zones as well as at All India level is depicted in Figure 1.

The summary statistics of the seasonal rainfall as well as annual rainfall for all the zones of India along with All India rainfall is reported in Table 1. We considered mean, median, maximum value, minimum value, standard deviation, coefficient of variation, Skewness, Kurtosis and Jarque-Bera statistic. A perusal of Table 1 indicates that the monsoon rainfall as well as annual rainfall is highest NE zone and lowest in NW zone. In terms of CV the variability in monsoon rainfall and annual rainfall is highest in NW zone. J-B statistic implied that in all the zones, annual rainfall is normally distributed.

Figure 1. **The pattern of monsoon rainfall and annual rainfall in different zones of India**



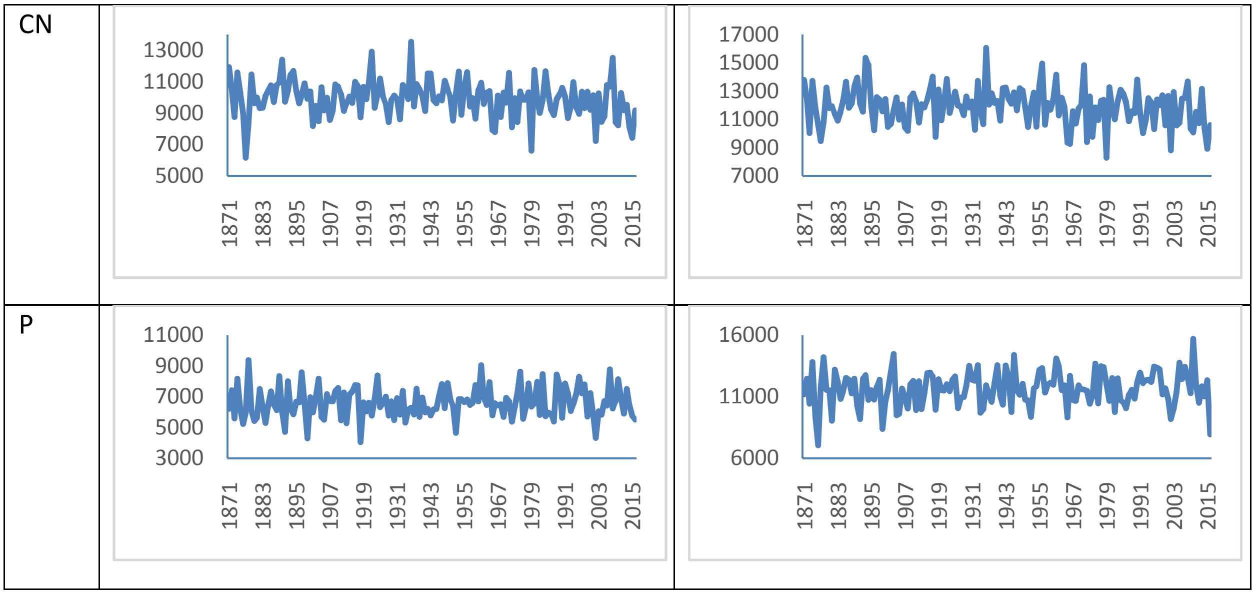


Table 1. Summary statistics of seasonal and annual rainfall in different zones of India

	AI					NW				
	JF	MAM	JJAS	OND	ANN	JF	MAM	JJAS	OND	ANN
Mean	232	943	8481	1201	10859	140	203	4924	207	5476
Med.	216	919	8585.5	1204	10879.5	108.5	165.5	4944.5	129	5469.5
Max.	611	1665	10202	2099	13470	421	977	8168	1382	10572
Min.	30	552	6040	501	8109	7	5	1620	0	1755
SD	116.25	205.47	834.52	345.61	1013.68	101.16	158.24	1291.89	225.34	1368.44
Skew.	0.67	0.70	-0.51	0.36	-0.02	0.91	2.26	-0.27	2.28	0.10
Kurt.	3.21	3.66	2.91	2.72	3.10	2.98	9.88	2.90	9.43	3.72
CV (%)	50.10	21.77	9.84	28.77	9.33	71.96	77.66	26.23	108.74	24.99
J-B	11.08	14.48	6.42	3.58	0.08	20.12	412.81	1.82	377.37	3.43
Prob.	0.00	0.00	0.04	0.17	0.96	0.00	0.00	0.40	0.00	0.18
	WC					NE				
Mean	189	437	9276	843	10746	422	4251	14072	1764	20511
Med.	151	384.5	9377.5	814	10711	402	4206	14117.5	1707.5	20692
Max.	665	1297	12116	2147	14433	1182	6565	17929	3738	25044
Min.	0	106	5323	79	5933	9	2266	11399	263	15764
SD	142.11	223.19	1237.22	428.48	1412.31	228.25	820.13	1275.02	722.26	1850.81
Skew.	0.97	1.21	-0.32	0.53	-0.19	0.69	0.20	0.20	0.39	-0.02
Kurt.	3.39	4.64	2.91	3.00	3.26	3.52	2.69	2.86	2.80	2.81
CV (%)	74.94	51.05	13.34	50.79	13.14	54.04	19.29	9.06	40.93	9.02
J-B	24.03	52.04	2.58	6.96	1.34	13.39	1.56	1.12	3.94	0.23
Prob.	0.00	0.00	0.28	0.03	0.51	0.00	0.46	0.57	0.14	0.89

	CN					P				
Mean	340	729	9921	915	11906	199	1386	6609	3432	11628
Med.	324.5	681.5	9983.5	837.5	11990	124.5	1295.5	6533	3378.5	11672
Max.	1039	1836	13536	2539	16055	993	2808	9378	5855	15677
Min.	8	176	6144	76	8275	1	524	4044	929	7052
SD	210.29	308.58	1154.02	515.14	1335.69	190.61	456.99	984.50	929.71	1373.27
Skew.	0.76	0.65	-0.13	0.74	0.07	1.42	0.96	0.26	0.04	-0.23
Kurt.	3.51	3.35	3.96	3.31	3.39	4.81	3.71	3.05	2.90	3.50
CV (%)	61.77	42.32	11.63	56.25	11.22	95.42	32.97	14.89	27.08	11.81
J-B	15.49	10.97	5.98	13.74	1.03	69.13	25.50	1.64	0.10	2.82
Prob.	0.00	0.00	0.05	0.00	0.60	0.00	0.00	0.44	0.95	0.24

3.2 Test for Long Memory

ELW estimator is applied to estimate the long memory parameter for all the series and the results are provided in Table 2. The results show that the parameter is significant for JF and OND series of all India data, JF and OND of North-west, OND and ANN of West-Central, ANN of North-west and JF of Peninsular. It establishes the presence of long range dependency in aforesaid rainfall data.

Table 2.Exact local Whittle estimator of long memory

Region	**JF**	**MAM**	**JJAS**	**OND**	**ANN**
All India	0.158*	-0.092	0.035	0.150*	0.116
North-West	0.084*	0.080	0.013	0.231*	0.027
West-Central	0.161	-0.025	0.111	0.204*	0.194*
North-East	0.074	-0.024	0.117	0.106	0.176*
Central-North	0.097	0.037	0.110	0.018	0.080
Peninsular	0.231*	0.086	-0.129	0.038	0.008

*significant at 5% level

3.3 Test for the Presence of Structural Break

The OLS based Chow test has been applied to the dataset to see the presence of structural break and the results are reported in Table 3. The results depicts that the test is significant for JJAS and ANN rainfall data of North-East and Central-North of India indicating presence of break in the respective datasets. According to this test there is a break point in the data set which at 1957 *i.e.* at 87th observation for North-East datasets and in the year of 1965 i.e. at 95th observation for Central-North datasets.

Table 3: Chow F-test for structural break

Region	**JF**	**MAM**	**JJAS**	**OND**	**ANN**
All India	6.387	1.982	6.387	3.788	3.747
North-West	4.765	5.235	1.850	3.890	2.034
West-Central	4.211	3.075	3.895	2.535	3.335
North-East	6.836	3.459	11.400*	4.892	10.812*
Central-North	4.315	3.787	8.844*	1.666	9.163*
Peninsular	5.397	6.349	1.896	1.522	3.373

*significant at 5% level

3.4 Test for Spurious Long Memory

Qu test for differentiation of true long memory from the spurious long memory has been conducted to all of the series and the results are listed in Table 4. Since the test statistics values are less than the critical value at 5% level i.e. 1.155 leading to acceptance of null hypothesis that the series is a true long memory process. In order words, the test results indicate that no series is wrongly detected as long memory process.

Table 4: Qu test for true long memory against spurious long memory

Region	JF	MAM	JJAS	OND	ANN
All India	0.731	0.553	0.794	0.646	0.612
North-West	0.475	0.315	0.653	0.766	0.652
West-Central	0.280	0.818	1.109	0.665	0.664
North-East	0.399	0.698	0.832	0.523	0.997
Central-North	0.625	0.643	0.565	0.564	0.800
Peninsular	0.733	1.048	0.399	0.845	0.793

4 Conclusions

It may happen that long memory and structural changes are easily confused and the time series is mistakenly detected as long memory process. However, most researchers choose to ignore the problem of structural break in testing for long memory. It is a known fact that short memory with structural break may exhibit the properties of long memory. To avoid the confusion test has to be performed to differentiate true long memory from spurious long memory. The main contribution of the paper is to detect if the DGP of monthly seasonal rainfall series of some zones across India is generated by a true long memory process. In this paper, we have employed exact local Whittle (ELW) estimator to estimate the long memory parameter. The results indicate that some of the series exhibit long memory pattern. Next, an empirical fluctuation process using the ordinary least square (OLS)-based Chow test is applied to detect the break date. Break dates are detected in two series of North-East and Central-North data sets in the year 1957 and 1965, respectively. Moreover, Qu test (Qu, 2011) has been considered for testing for true versus spurious long memory and no test is found to be significant. It means that the series are truly long memory processes.

References

Beran, J. (1995). Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models. *Journal of the Royal Statistical Society*, Series B (Methodological), **57**(**4**), 659-672.

Berkes, I., Horváth, L., Kokoszka, P. and Shao, Q.M. (2006). On discriminating between long-range dependence and changes in mean. *The Annals of Statistics*, **34**(**3**), 1140-1165.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2007). *Time-Series Analysis: Forecasting and Control*. Pearson Education, India.

Cheung, Y.W. (1993). Tests for fractional integration: A Monte Carlo investigation. *Journal of Time Series Analysis*, **14**(**4**), 331-345.

Diebold, F.X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, **105**(**1**), 131-159.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long-memory time series models. *Journal of Time Series Analysis,* **4**, 221–238.

Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis,* **4**, 221–238.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S and P 500 absolute stock returns. *Journal of Empirical Finance*, **11**(**3**), 399-421.

Hurst, H.E. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Agricultural Engineers*, **116**, 770–799.

Ohanissian, A., Russell, J. R. and Tsay, R.S. (2008). True or spurious long memory? A new test. *Journal of Business and Economic Statistics*, **26**(**2**), 161-175.

Paul, R.K. (2017). Modelling long memory in maximum and minimum temperature series in India. *Mausam,* **68**(**2**), 317-326.

Paul, R.K. and Anjoy, P. (2018). Modeling fractionally integrated maximum temperature series in India in presence of structural break. *Theoretical and Applied Climatology*, **134**, (**1**&2), 241-249.

Paul, R.K., Birthal, P.S. and Khokhar, A. (2014). Structural breaks in mean temperature over agro-climatic zones in India. *The Scientific World Journal*; dx.doi.org/10.1155/2014/434325.

Paul, R.K., Birthal, P.S., Paul, A.K. and Gurung, B. (2015). Temperature trend in different agro-climatic zones in India. *Mausam*, **66**(**4**), 841-846.

Paul, R.K. and Birthal, P.S. (2016). Investigating rainfall trend over India using the wavelet technique. *Journal of Water and Climate Change,* **7**(**2**), 353-364.

Paul, R.K., Prajneshu, and Ghosh, H. (2013). Wavelet frequency domain approach for modelling and forecasting of Indian monsoon rainfall time-series data. *Journal of the Indian Society of Agricultural Statistics*, **67** (**3**), 319-327.

Perron, P. and Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics*, **28**(**2**), 275-290.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business and Economic Statistics*, **29**(**3**), 423-438.

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, **105**(**491**), 1228-1240.

Shimotsu, K. and Phillips, P.C. (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, **33**(**4**), 1890-1933.