Institute Project Code: AGENIASRISIL201701000096

Study of long memory and periodicities in climate variables in different Meteorological Subdivisions of India

Dr. Ranjit Kumar Paul (Principal Investigator)

Dr. L M Bhar (Co-Principal Investigator)

Dr. A K Paul (Co-Principal Investigator)

सांख्यिकीय आनुवांशिकी प्रभाग
Division of Statistical Genetics

भा.कृ.अनु.प.- भारतीय कृषि सांख्यिकी अनुसंधान संस्थान
लाइब्रेरी एवेन्यू, पूसा, नई दिल्ली – 110 012
ICAR-Indian Agricultural Statistics Research Institute
Library Avenue, PUSA, New Delhi – 110 012
www.iasri.res.in

(2020)

PREFACE

Climatic change has been much debated in the scientific world in recent decades. The detection and estimation of trends in the presence of noise, long memory, periodicities, or discontinuous patterns is important in climate research studies. The problem of predicting seasonal monsoon rainfall, and indeed of assessing the degree of predictability in the monsoon, continues to be of great fundamental and practical importance. Detection of significant periodicities in the available rainfall data can be of great value in prediction and has attracted much attention for nearly a century. For modelling time-series in presence of long memory, the autoregressive fractionally integrated moving-average (ARFIMA) model is used. A time-series may perhaps show long memory, not because it is really I(d) but sometimes because of the neglected structural breaks in the series. Ignoring the presence of breaks can lead to seriously biased estimates and forecasts. Most of the periodicities tests have been developed to deal with periodic component processes with additive white noise. In recent years, the wavelet transform has been developed as an alternative to the Fourier transform. The discrete wavelet transform (DWT) has been extensively used for the analysis of FD processes, since it matches the structure of such processes.

A common issue that often arises when working with real time-series which might exhibit long memory is the possibility of structural change. In literature of studying long memory in presence of structural break, the joint test of fractional integration and structural break in Indian condition has hardly been used. The existing test for testing periodicities in presence of long memory in climate variable needs to be used in the Indian climate. It is well recognized that the MK test is not robust against autocorrelation and cross correlation, and also depends on the sample size as well as magnitude of the trend to be identified. As such, there remains a need for using new types of methods in order to detect and test for trends. Forecasting of climate variables using wavelets is of prime importance now a days.

The long memory behaviour of monthly maximum temperature of India for the period 1901 to 2007 is investigated. Significant increasing trend is found in the maximum temperature series in India. The rate of increase in maximum temperature accelerated after 1960s as compared to the earlier period. Here, an attempt is also made to detect the structural break for seasonally adjusted monthly maximum temperature series. It is found

that there is a significant break in maximum temperature during July, 1963. Two Stage Forecasting (TSF) approach to deal with the coexistence of long memory and structural change in temperature pattern is discussed thoroughly. The forecast performance of the fitted model is assessed on the basis of RMAPE (Relative Mean Absolute Prediction Error), SSE (Sum of Squared Errors) and MSE (Mean Squared Errors) for different forecast horizons.

A procedure for estimating the fractional differencing parameter in semiparametric contexts, namely exact local Whittle (ELW) estimator proposed by Shimotsu and Phillips (2005) is employed to analyze seasonal rainfall data 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. The sub-divisional rainfall data of India during the period 1871 to 2016 has been investigated by using wavelet analysis. On the decomposed series, ARIMA as well as Artificial Neural Network (ANN) model is applied and by means of inverse wavelet transform, the prediction of rainfall for different sub-divisions have been obtained. To this end, empirical comparison was carried out toward forecast performance of the approaches namely Wavelet-ANN, Wavelet-ARIMA and ARIMA. It is reported that Wavelet-ANN and Wavelet-ARIMA approach out performs the usual ARIMA model for forecasting of rainfall for the data under consideration.

The authors express their sincere thanks to the Director, IASRI for his support and for providing all necessary facilities to carry out the research work successfully. The cooperation received from Head, Division of Statistical Genetics, IASRI and other scientists in the Division are thankfully acknowledged. We also express our thanks to the internal referee whose suggestions helped in improving the contents and presentation of this report.

Project Investigators

CONTENTS

Chapter I

INTRODUCTION AND REVIEW OF LITERATURE

Introduction

Climatic change has been much debated in the scientific world in recent decades. The detection and estimation of trends in the presence of noise, long memory, periodicities, or discontinuous patterns is important in climate research studies. The problem of predicting seasonal monsoon rainfall, and indeed of assessing the degree of predictability in the monsoon, continues to be of great fundamental and practical importance. Detection of significant periodicities in the available rainfall data can be of great value in prediction and has attracted much attention for nearly a century. One persistent question that arises in the quest for identifying periodicities is the heterogeneity of the Indian rainfall at regional level. In spite of such heterogeneity, the analysis is still focused on all-India level.

Time series modeling and forecasting is a dynamic research area which has attracted the attentions of researcher community over last few decades for analyzing climatic data. Although most economic time series are non-stationary and do require differencing of some kind it is not necessarily true that taking integer order differences and then using an ARMA model will be the best remedy. In Box-Jenkins ARIMA methodology, it is assumed that if the series is non-stationary, the integer order differenced series will be well behaved, as long as there are no seasonal components. In particular, it is hoped that differenced series will have rapidly decaying autocorrelations, so that it can be well described by a stationary invertible ARMA model. But, this is not always the case. However, some series evidently do not possess a further unit root, while they show signs of dependence and possess long memory. The long memory models provide us with a way to define such a fractional difference, and would provide a useful alternative to an ARMA model. For modeling time-series in presence of long memory, the autoregressive fractionally integrated moving-average (ARFIMA) model (Granger and Joyeux, 1980) is used.

A time-series may perhaps show long memory, not because it is really I(d) but sometimes because of the neglected structural breaks in the series (Granger and Hyung, 2004) Literally, structural change can be described as fundamental shift in the structure of the series under consideration. Ignoring the presence of breaks can lead to seriously biased estimates and forecasts. Many time series in different fields can be modeled as an underlying process with some periodic variations, contaminated by additive random noise. Fourier showed that almost any periodic function can be represented by a series of sine and cosine functions (sinusoids) or a mixtures of sine and cosine functions (Priestley 1981). But for many time series, it is more complicated in that several periodic terms may be present with unknown cycles. The so-called hidden periodicities can be determined from peaks in the periodogram, although not all peaks correspond to a genuine periodic component in the process.

Most of the periodicities tests have been developed to deal with periodic component processes with additive white noise. Many processes, however, exhibit some sort of correlation structure which makes the additive white noise assumption unrealistic. A variety of such dependence at long ranges, can be modeled by a class of processes called fractional difference processes (FD). In recent years, the wavelet transform has been developed as an alternative to the Fourier transform. The discrete wavelet transform (DWT) has been extensively used for the analysis of FD processes, since it matches the structure of such processes. The key property of the DWT is that it transforms a time series into coefficients that reflect changes at various scales and at particular times. For FD processes, the DWT wavelet coefficients for a given scale are approximately uncorrelated; see Percival and Walden (2000).

Frequency-domain analysis tools are very appealing to study the variables that exhibit a cyclical behaviour and/or are affected by seasonal effects. Spectral analysis and Fourier transforms can be used to quantify the importance of the various frequency components of the variable under investigation. Many methods have been used for trend detection and testing in literature, none have emerged as standard. Perhaps the most common test for trends is the rank-based nonparametric Mann–Kendall (MK) method.

It is, however, well recognized that the MK test is not robust against autocorrelation and cross correlation, and also depends on the sample size as well as magnitude of the trend to be identified. Moreover, the test of presence of trend in climate variable needs to be studied properly by using recent development in this field.

A common issue that often arises when working with real time-series which might exhibit long memory is the possibility of structural change. It would be interesting to see the effect of

structural break in long memory time series model with reference to forecast performance. In literature of studying long memory in presence of structural break, the joint test of fractional integration and structural break in Indian condition has hardly been used. There is a need to use the test for more precise interpretation. Sometimes long memory is coupled with volatility and the corresponding model may be ARFIMA-GARCH model. For this model, optimal out-of-sample forecast formulae does not exist. The existing test for testing periodicities in presence of long memory in climate variable needs to be used in the Indian climate. It is well recognized that the MK test is not robust against autocorrelation and cross correlation, and also depends on the sample size as well as magnitude of the trend to be identified. As such, there remains a need for using new types of methods in order to detect and test for trends. Forecasting of climate variables using wavelets is of prime importance now a days. But it is very difficult to apply in the real data because of non-availability of readymade module in software. There is a need to develop a package (R) for the above propose.

Paul (2014) and Paul et al. (2015a,b) have applied ARFIMA model for forecasting of agricultural commodity prices. The authors have also compared different estimation techniques for estimating the long memory parameter by means of MCMC and concluded that wavelet based estimation outperforms the other techniques

Agricultural performance of a country, generally, depends to a large extent on the quantum and distribution of rainfall. Therefore accurate modelling is vital in planning and policy making. Azad et al. (2008) developed a Wavelet-based significance test for periodicities in Indian monsoon rainfall. Sunilkumar and Prajneshu (2008) carried out modelling and forecasting of marine fish production of India using Wavelet thresholding with autocorrelated errors.

Paul et al. (2011) applied Wavelet methodology for detection of trend in Indian monsoon rainfall and found that there is a significant declining trend. Ghosh et al. (2013) applied Periodic ARCH model for modelling rainfall data. Ghosh et al. (2010) investigated wavelet frequency domain approach for statistical modelling of rainfall time-series data

Recently wavelet based hybrid time series models are widely used in forecasting time series data. (Anjoy and Paul, 2017). The authors have applied the technique for forecasting agricultural commodity prices. Paul et al. (2014) studied the structural changes in mean temperature over agro-climatic zones in India. Paul et al. (2015) studied Temperature trend in

different agro-climatic zones in India. Paul and Birthal (2015). Investigated rainfall trend in different agro-climatic zones over India using the nonparametric wavelet technique. Granger and Joyeux (1980), Granger (1980, 1981) and Hosking (1981) developed an autoregressive fractionally integrated (ARFIMA) model and provided a parametric tool for long memory analysis. Booth et al. (1982) and Helms et al. (1984) stated that the long memory process is the explanation of irregular cyclical patterns of certain financial series. Helms et al. (1984), Peters (1989, 1991, 1994), applied rescaled range (R/S) analysis to detect the existence of long memory in the futures prices of the soybean. Daubechies (1988), Hernandez and Weiss (1996) have found rapid decay in the wavelet coefficient's covariance. A good description of Long-Memory Processes can be found in Beran (1994).

Estimation of long-memory parameter

(1) Frequency domain approach

GPH estimator (Geweke and Porter-Hudak, 1983)

Gaussian semiparametric estimator (Robinson, 1995, 2003)

(2) Wavelet (time-frequency) domain approach (Abry and Veitch, 1996 and Jensen, 1999)

Lo (1991) modified the above R/S method and produced a new statistic that is robust. Robinson (1995), Hurvich et al. (1998) and Tanaka (1999) have analyzed the GPH estimate in detail. Diebold and Inoue (2001) studied first the joint analysis of long memory and structural break and proved how structural change can be misinterpreted as long memory. Wang et al. (2013) suggested to fit an autoregressive model directly to the long memory series choosing the order, k, via a selection criteria, i.e. AIC or Cp. Papailias and Dias (2015) introduced a two-stage approach for forecasting of long memory time-series focusing on spurious long memory cases

Using time domain, Fryzlewicz et al. (2003) developed Wavelet process model for forecasting nonstationary time-series. For the frequency approach, Renaud et al. (2003) developed methodology for prediction of time-series data based on multiscale decomposition. Aminghafari and Poggi (2007, 2012) used Wavelets and kernel smoothing approach for forecasting nonstationary time-series. Recently hybrid time series models are widely used in forecasting time series data. (Aminghafari and Poggi, 2012). But it is very difficult to implement the hybrid model based on wavelet in the real data due to non-availability of readymade software. So there is a need to develop an R package

Mann–Kendall tests are widely used in environmental science because they are simple, robust and can cope with missing values and values below a detection limit. Since the first proposals of the test by Mann (1945) and Kendall (1975), the test was extended in order to include seasonality (Hirsch et al., 1982), multiple monitoring sites (Lettenmaier, 1988), and covariates representing natural fluctuations (Libiseller and Grimvall, 2002). In addition to the studies mentioned earlier, the nonparametric Mann–Kendall test has been used in a variety of climate and streamflow studies in Canada (e.g. Gan, 1995, 1998; Gobena and Gan, 2006). Kallache et al. (2005) used a Discrete Wavelet Transform (DWT) to assess trends in flood data. Partal and Kucuk (2006) used a DWT to assess trends in precipitation data. Almasri et al. (2008) used a DWT to assess trends in temperature. Almasri (2010) introduced two tests for testing the trend in long memory processes, based on the wavelet and band periodogram. They have compared with the OLS-based test. The results have shown that the OLS tests perform better when the FD parameter is smaller, and the wavelet based tests and the band periodogram based tests are much better as d approximate 0.5. Periodicity tests have a long history, including, for example, the contributions of Schuster (1898), Fisher (1929), and Siegel (1980). Almasri (2011) proposed a New Approach for Testing Periodicity. Adamowski et al. (2009) Developed of a new method of wavelet aided trend detection and estimation.

Keeping the above points in mind, this study was undertaken with the following objectives:

1. To investigate the trend and variability in climate variable in presence of long memory using wavelets.
2. To investigate the joint effect of long memory and structural break on forecasting
3. To develop formulae for out-of-sample forecast for ARFIMA-GARCH model
4. To identify periodicities in the climate variables
5. To develop R-package for forecasting using hybrid models based on wavelets

Chapter II

LONG MEMORY IN MAXIMUM AND MINIMUM TEMPERATURE SERIES IN INDIA

1. Introduction

Large number of research papers have been published on long memory and fractionally integrated processes since the initial publication of the work of Granger (1980); Granger and Joyeux (1980) and Hosking (1981) which parameterized the processes of Hurst (1951) on the time series with hyperbolically decaying autocorrelations. The long memory or long term dependence property describes the high-order correlation structure of a time series. If a series exhibits long memory, there is persistent temporal dependence even between distant observations. Such series are characterized by distinct but non-periodic cyclical patterns. The presence of long memory dynamics causes nonlinear dependence in the first moment of the distribution and hence a potentially predictable component in the series dynamics. Fractionally integrated processes can give rise to long memory (Beran, 1994).

A popular class of models for time series with long memory behaviour is the autoregressive fractionally integrated moving average (ARFIMA) model by Granger and Joyeux (1980). This kind of models extended classical ARIMA models by assuming the differencing parameter *d* as a fractional value. It is well known that ARFIMA models are linear time series model. Fractional integration is part of the larger classification of time series, commonly referred to as 'long memory' models. The recent empirical evidence suggests that temperature series may be well described in terms of fractionally integrated processes (Gil-Alana, 2004). Fractionally integrated I(d) processes have attracted growing attention among empirical researchers. In fact this is because I(d) processes provide an extension to the classical dichotomy of I(0) and I(1) time series and equip us with more general alternatives in long range dependence (Shimotsu, 2010). Empirical research continues to find evidence that I(d) processes can provide a suitable description of certain long range characteristics.

Understanding the nature and scale of possible climate changes in India is of importance to the policy makers and farmers as it gives them a chance to be prepared for better mitigation and adaptation measures. For that purpose time series analysis of weather data can be a very valuable tool to investigate its variability pattern and, maybe, even to predict short- and long-term changes in the time series. Various researchers have carried out studies on temperatures. Woodcock (1984) described some experimental MOS forecasts of daily maximum and minimum temperature for seven Australian cities. Raj (1998) evolved a scheme for predicting minimum temperature at Pune by analogue and regression methods. Mohan *et al.* (1989) developed a method for forecasting maximum temperature over Ozar situated in Maharastra using maximum and dew point temperature of the previous day. In the present investigation monthly minimum temperatures in India, for the period 1901–2007 were examined by means of fractional integration techniques. Dhimri *et al*. (2005) have carried out forecast of minimum temperature at Manali, India. Paul *et al*. (2014) have investigated the trend in mean surface temperature in different agro-climatic zones in India. Paul *et al*. (2015) have also investigated the structural break in mean surface temperature in different agro-climatic zones in India and reported that there is significant structural break during 1970's. However, in none of the above studies, long memory nature in maximum and minimum surface temperature in India has been investigated. In the present investigation, an attempt has been made to apply long memory model for forecasting maximum and minimum surface temperature in India with more accuracy. To this end, nonparametric wavelet technique has also been applied to study the pattern of maximum and minimum surface temperature in India over the last century or so both globally as well as locally. Some applications of this technique in modelling climate variables may be found in Paul *et al*. (2013), Paul *et al.* (2011), Paul and Birthal (2015). The paper is organized as follows: section 2 describes the data set used in the present investigation; section 3 describes the long memory definition, ARFIMA model, testing stationarity and testing of presence of long memory followed by section 4 which deals with results and discussion.

2. DATA

For the present investigation, all India monthly maximum and minimum temperature data during the period January, 1901 to December, 2007 is used. The data is collected form Indian Institute of Tropical Meteorology, Government of India. The data for the period January 1901 to December, 2006 have been used for model building and the remaining data have been used for model validation purpose.

3. METHODOLOGY

Long Memory Process

Long memory in time-series can be defined as autocorrelation at long lags (Robinson, 1995). According to Jin and Frechette (2004), memory means that observations are not independent (each observation is affected by the events that preceded it). The autocorrelation function (acf) of a time-series y_t is defined as

$$\rho_k = cov(y_t, y_{t-1})/var(y_t) \quad (1)$$

for integer lag k. A covariance stationary time-series process is expected to have autocorrelations such that $\lim_{k\to\infty} \rho_k = 0$. Most of the well-known class of stationary and invertible time-series processes have autocorrelations that decay at the relatively faster exponential rate, so that $\rho_k \approx |m|^k$, where |m|<1 and this property is true, for example, for the well-known stationary and invertible ARMA(p,q) process. For long memory processes, the autocorrelations decay at an hyperbolic rate which is consistent with $\rho_k \approx Ck^{2d-1}$, as k increases without limit, where C is a constant and d is the long memory parameter.

ARFIMA Model

Fractional integration is the primary conceptual framework for describing long memory in financial time-series. Fractional integration is a generalization of integer integration, under which time-series are usually presumed to be integrated of order zero or one. For example, an autoregressive moving-average process integrated of order d [denoted by ARFIMA(p, d, q)] can be represented as

$$(1 - L)^{\mathrm{d}}\varphi(L)y_t = \theta(L)u_t \quad (2)$$

where u_t is an independently and identically distributed (i.i.d.) random variable with zero mean and constant variance, L denotes the lag operator; and $\varphi(L)$ and $\theta(L)$ denote finite polynomials in the lag operator with roots outside the unit circle. For d = 0, the process is stationary, and the effect of a shock to $u(t)$ on $y(t + j)$ decays geometrically as j increases. For d = 1, the process is said to have a unit root, and the effect of a shock to $u(t)$ on $y(t + j)$ persists into the infinite future. In contrast, fractional integration defines the function $(1 - L)^{-d}$ for noninteger values of the fractional differencing parameter d. It turns out that for -0.5< d< 0.5 the process $y(t)$ is stationary and invertible. A detail description of ARFIMA model can be found in Robinson (2003).

Estimation procedures

We deal with some well known estimation methods of the long memory parameter d. The first one is the semiparametric method based on an approximated regression equation obtained from the logarithm of the spectral density function of a model. This method is proposed by Geweke and Porter-Hudak (1983). The second is the Gaussian semiparametric method developed by Robinson (1995).

Testing of Long Memory

H=1-*d*, a Hurst exponent produced by the rescaled range analysis, or R/S, analysis and applied to economic price analysis by Booth *et al*. (1982) and Helms *et al*. (1984). For a given time-series, the Hurst exponent measures the long-term non-periodic dependence, and indicates the average duration the dependence may last.

The time period spanned by the time series of length *T* is divided into *m* contiguous sub-periods of length *n* such that *m*n = T*. In each sub-period X_{ij}, the elements have two subscripts. The first subscript (i = 1,…,n) denotes the number of elements in each sub-period and the second one (j = 1,…,m) denotes the sub-period index. For each sub-period j the R/S statistic is calculated as follows:

$$\left(R/S\right)_j = \left(S_j\right)^{-1} \max_{1\le k\le n} \sum_{i=1}^{k} \left(x_{ij} - \bar{x}_j\right) - \min_{1\le k\le n} \sum_{i=1}^{k} \left(x_{ij} - \bar{x}_j\right) \tag{3}$$

where S_j is the standard deviation for each sub-period. In (3), the k deviations from the sub-period mean have zero mean; therefore the last value of the cumulative deviations for each sub-period will always be zero. Because of this, the maximum value of the cumulative deviations will always be greater or equal to zero, while the minimum value will always be less or equal to zero. Rescaling the range is crucial since it allows diverse phenomena and time periods to be compared, which means that *R/S* analysis can describe time series with no characteristic scale. The $(R/S)_n$ is computed by the average of the $(R/S)_j$ values for all the m contiguous sub-periods with length n as

$$\left(R/S\right)_n = m^{-1}\left(R/S\right)_j \tag{4}$$

Eq. (4) computes the *R/S* value which corresponds to a certain time interval of length n. This is repeated by increasing n to the next integer value, until $n = T/2$, since at least two sub-periods are needed, to avoid bias.

As n increases**,** the following holds:

$$\log\left((R/S)_n\right) = \log\alpha + H\log n \tag{5}$$

When 0.5<H<1, the long memory structure exists. If H ≥ 1, the process has infinite variance and is nonstationary. If 0<H<0.5, anti-persistence structure exists. If H=0.5, the process is white noise.

Geweke and Porter-Hudak (*GPH*) *estimate*

The GPH estimation procedure is a two-step procedure, which begins with the estimation of d. This method is based on least squares regression in the spectral domain, exploits the sample form of the pole of the spectral density at the origin: $f_x(\lambda) \sim \lambda^{-2d}$, $\lambda \to 0$. To illustrate this method, we can write the spectral density function of a stationary model X_t , t = 1, . . . ,T as

$$f_x(\lambda) = \left[4\sin^2\left(\frac{\lambda}{2}\right)\right]^{-d} f_\varepsilon(\lambda)$$

where $f_\varepsilon(\lambda)$ is the spectral density of ε_t , assumed to be a finite and continuous function on the interval $[-\pi, \pi]$. Taking the logarithm of the spectral density function $f_x(\lambda)$ the log-spectral density can be expressed as

$$\log\left(f_x(\lambda)\right) = \log\left(f_\varepsilon(0)\right) - d\log\left[4\sin^2\left(\frac{\lambda}{2}\right)\right] + \log\frac{f_\varepsilon(\lambda)}{f_\varepsilon(0)}$$

Let $I_x(\lambda_j)$ be the periodogram evaluated at the Fourier frequencies $\lambda_j = 2\pi_j / T$; j = 1, 2, . . . , m; T is the number of observations and m is the number of considered Fourier frequencies, that is the number of periodogram ordinates which will be used in the regression

$$\log\left(I_x(\lambda_j)\right) = \log\left(f_\varepsilon(0)\right) - d\log\left[4\sin^2\left(\frac{\lambda}{2}\right)\right] + \log\frac{f_\varepsilon(\lambda)}{f_\varepsilon(0)} + \log\frac{I_x(\lambda_j)}{f_x(\lambda_j)}$$

where $\log\left(f_\varepsilon(0)\right)$ is a constant, $\log\left[4\sin^2(\lambda/2)\right]$is the exogenous variable and $\log\left[I_x(\lambda_j)/\, f_x(\lambda_j)\right]$ is a disturbance error. The GPH estimate requires two major assumptions related to asymptotic behaviour of the equation

H_1: for low frequencies, we suppose that $\log\left[f_\varepsilon(\lambda)/\, f_\varepsilon(0)\right]$ is negligible.

H_2: the random variables $\log\left[I_x(\lambda_j)/\, f_x(\lambda_j)\right]$; j = 1, 2, . . ., m are asymptotically iid.

Under the hypotheses H_1 and H_2, we can write the linear regression

$$\log\left(I_x\left(\lambda_j\right)\right)=\alpha-d\log\left[4\sin^2\left(\frac{\lambda_j}{2}\right)\right]+e_j$$

where $e_j \sim iid\left(-c,\pi^2/6\right)$. Let $y_j=-\log\left[4\sin^2\left(\lambda_j/2\right)\right]$ the GPH estimator is the OLS estimate of the regression log $I_x\left(\lambda_j\right)$ on the constant α and y_j. The estimate of d is

$$\hat{d}_{GPH}=\frac{\sum_{j=1}^{m}\left(y_j-\bar{y}\right)\log\left(I_x\left(\lambda_j\right)\right)}{\sum_{j=1}^{m}\left(y_j-\bar{y}\right)^2}, \text{ where } \bar{y}=\sum_{j=1}^{m}y_j/m$$

The parameter m is selected so that $m = T^u$, with u = 0.5; 0.6; 0.7. Robinson (1995), Hurvich *et al*. (1998) and Tanaka (1999) have analyzed the GPH estimate in detail. Under the assumption of normality for X_t, it has been proved that the estimate is consistent and asymptotically normal. An alternative semiparametric estimator has been proposed by Robinson (1995).

Wavelets

Wavelets are fundamental building block functions, analogous to the trigonometric sine and cosine functions. As with a sine or cosine wave, a wavelet function oscillates about zero. This oscillating property makes the function a *wave*. However, the oscillations for a wavelet damp down to zero, hence the name *wavelet*. If $\psi(.)$ is a real-valued function defined over the real axis $(-\infty,\infty)$ and satisfies two basic properties: (i) Integral of $\psi(.)$ is zero, i.e. $\int_{-\infty}^{\infty}\psi(u)\,du=0$ (ii) Square of $\psi(.)$ integrates to unity, i.e. $\int_{-\infty}^{\infty}\psi^2(u)\,du=1$, then the function $\psi(.)$ is called a wave. A good description of wavelets can be found in Daubechies (1992), Ogden (1997) and Percival and Walden (2000).

Maximal Overlap Discrete Wavelet Transforms (MODWT)

The Maximal overlap discrete wavelet transforms (MODWT) is a linear filtering operation that transforms a series into coefficients related to variations over a set of scales. It is similar to DWT, in that, both are linear filtering operations producing a set of time-dependent

wavelet and scaling coefficients. Both have basis vectors associated with a location *t* and a unit less scale $\tau_j = 2^j - 1$ for each decomposition level *j* =1,..,J_0. Both are suitable for the analysis of variance (ANOVA) and for multiresolution analysis (MRA). However, MODWT differs from DWT in the sense that it is a highly redundant, nonorthogonal transform (Percival and Walden, 2000). It retains downsampled values at each level of the decomposition that would otherwise be discarded by DWT. The MODWT is well defined for all sample sizes *N*, whereas for a complete decomposition of *J* levels, DWT requires *N* to be a multiple of 2^J.

MODWT Coefficients

For a redundant transform, like MODWT, an *N* sample input time-series will have an *N* sample resolution scale for each resolution level. Therefore, features of wavelet coefficients in a multiresolution analysis (MRA) will be lined up with original time-series in a meaningful way. For a time-series $\boldsymbol{X}$ with arbitrary sample size *N*, the j^{th} level MODWT wavelet ($\widetilde{\boldsymbol{W}}_j$) and scaling ($\widetilde{\boldsymbol{V}}_j$) coefficients are defined as:

$$\widetilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \widetilde{h}_{j,l} X_{t-l \bmod N}, \qquad \widetilde{V}_{j,t} \equiv \sum_{l=0}^{L_j-1} \widetilde{g}_{j,l} X_{t-l \bmod N}, \tag{6}$$

where $\widetilde{h}_{j,l} \equiv h_{j,l} / 2^{j/2}$ are j^{th} level MODWT wavelet filters, and $\widetilde{g}_{j,l} \equiv g_{j,l} / 2^{j/2}$ are j^{th} level MODWT scaling filters, L_j is width of j^{th} level equivalent wavelet and scaling filters. For a time-series $\boldsymbol{X}$ with *N* samples, MODWT yields an additive decomposition or MRA given by

$$\boldsymbol{X} = \sum_{j=1}^{J_0} \widetilde{\boldsymbol{D}}_j + \widetilde{\boldsymbol{S}}_{J_0}, \tag{7}$$

where

$$\widetilde{D}_{j,t} = \sum_{l=0}^{N-1} \widetilde{u}_{j,l}\ \widetilde{W}_{j,t+l \bmod N}, \qquad \widetilde{S}_{j,t} = \sum_{l=0}^{N-1} \widetilde{v}_{j,l}\ \widetilde{V}_{j,t+l \bmod N}, \tag{8}$$

$\widetilde{u}_{j,l}$ and $\widetilde{v}_{j,l}$ being the filters obtained by periodizing $\widetilde{h}_{j,l}$ and $\widetilde{g}_{j,l}$. According to eq. (7), at a scale *j*, a set of coefficients {$\boldsymbol{D_j}$} each with the same number of samples (*N*) as in the original signal ($\boldsymbol{X}$) is obtained. These are called wavelet "details" and capture local fluctuations over whole period of a time-series at each scale. Set of values $\boldsymbol{S}_{J_0}$ provide a "smooth" or overall "trend" of the original signal and adding $\boldsymbol{D}_j$ to $\boldsymbol{S}_{J_0}$, for $j = 1, 2, \ldots, J_0$, gives an increasingly more accurate approximation for it. This additive form of reconstruction allows prediction of

each wavelet subseries ($\boldsymbol{D}_j, \boldsymbol{S}_{J_0}$) separately and adding individual predictions an aggregate forecast is generated.

Choosing Number of Levels

A time-series can be completely or partially decomposed into a number of levels. For complete decomposition of a series of length $N = 2^J$ using DWT, maximum number of levels in the decomposition is J. In practice, a partial decomposition of level $J_0 \leq J$ suffices for many applications. A J_0 level DWT decomposition requires that N be an integral multiple of 2^{J_0}. The MODWT can accommodate any sample size N and, in theory, any J_0. In practice, largest level is commonly selected such that $J_0 \leq log_2(N)$ in order to preclude decomposition at scales longer than total length of the time-series. In particular, for alignment of wavelet coefficients with the original series, condition $L_{J_0} < N$ (i.e. width of equivalent filter at J_0 th level is less than sample size) should be satisfied to prevent multiple wrappings of the time-series at level J_0. Selection of J_0 determines the number of octave bands and thus number of scales of resolution in the decomposition.

4. RESULTS AND DISCUSSION

A perusal of the Figure 1 and 2 indicate that both the series are stationary. In order to test for stationarity, two tests namely Augmented Dickey-Fuller unit root test (Said and Dickey, 1984) and Philips-Peron unit root test (Philips and Peron, 1988) are conducted. The descriptive statistics for monthly maximum and minimum temperature have been computed and are reported in table 1. A perusal of table1 reveals, the variability in minimum temperature is more than maximum temperature. The same can also be observed from the histogram plotted in Figure 3 and 4. The results of the stationarity tests are given in Table 2. A perusal of Table 2 reveals that both the test statistics reject the null hypothesis of presence of unit root indicating that series are stationary.

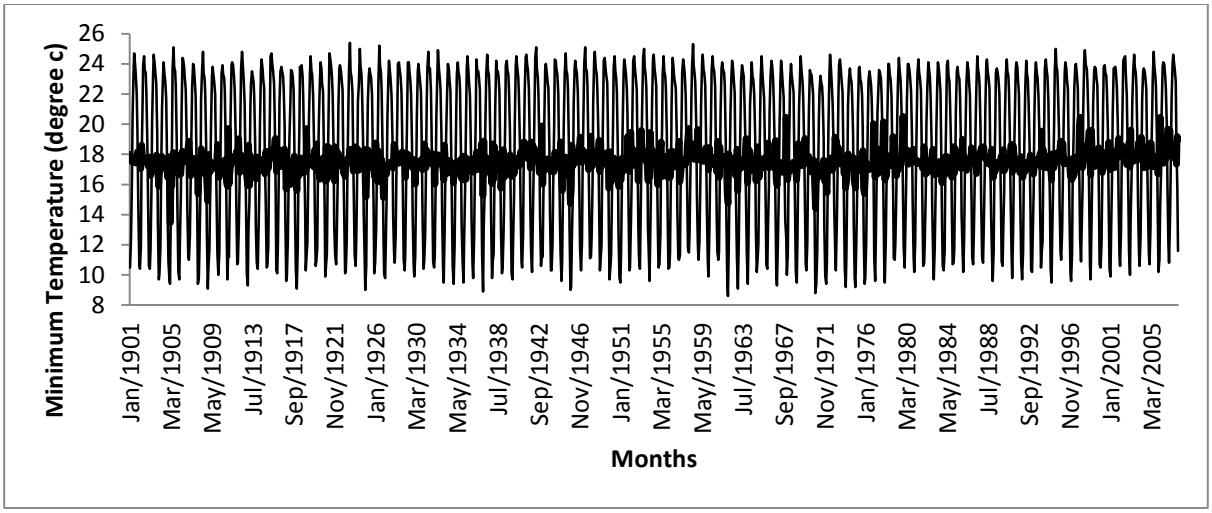


Fig. 1. All India minimum temperature in degree centigrade (dark line is seasonally adjusted series and light line is average monthly minimum temperature)

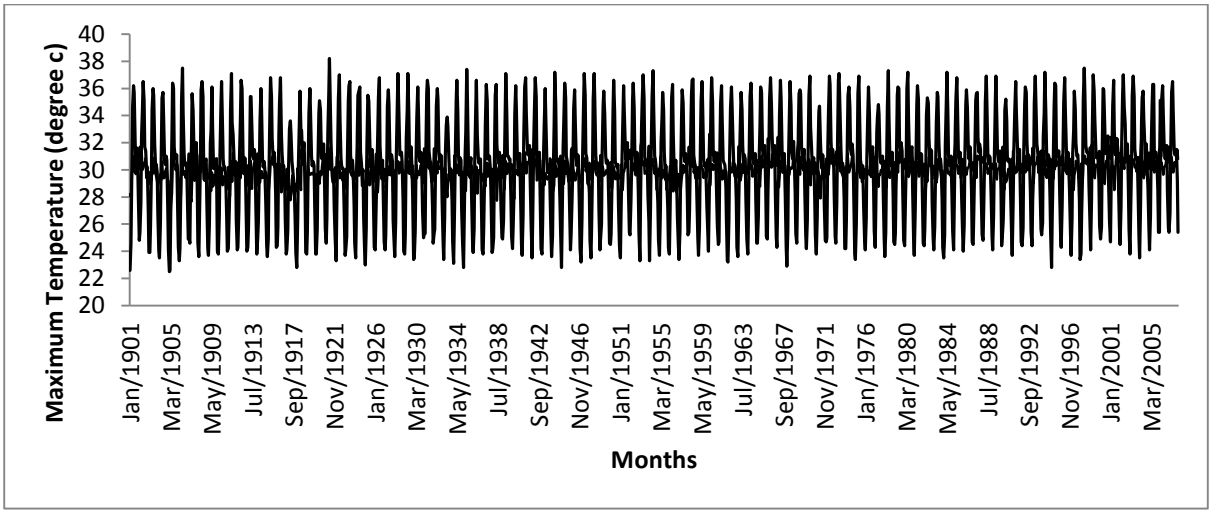


Fig. 2. All India maximum temperature in degree centigrade (dark line is seasonally adjusted series and light line is average monthly minimum temperature)

Table 1. Descriptive Statistics of maximum and minimum temperature

Descriptive Statistics	Maximum Temperature	Minimum Temperature
Mean	30.24	18.33
Median	30.7	19.7
Maximum	38.2	25.4
Minimum	22.5	8.6
Std. Deviation	3.81	5.15
CV	12.60	28.10

Skewness	-0.12	-0.39
Kurtosis	2.13	1.6

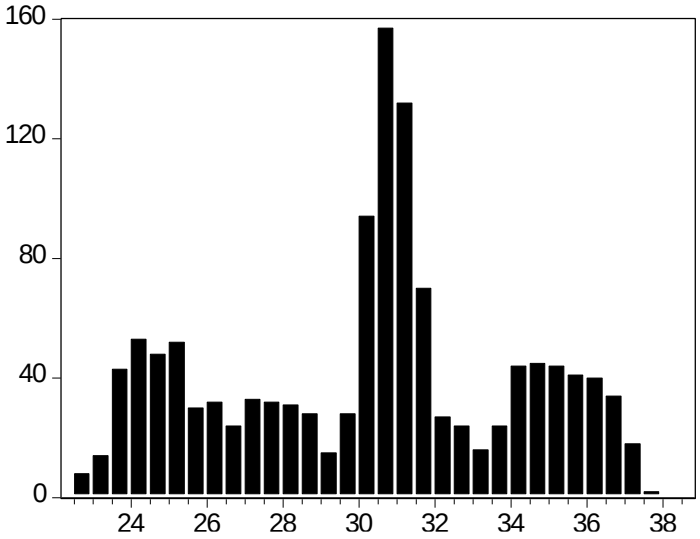


Fig. 3. Histogram of maximum temperature

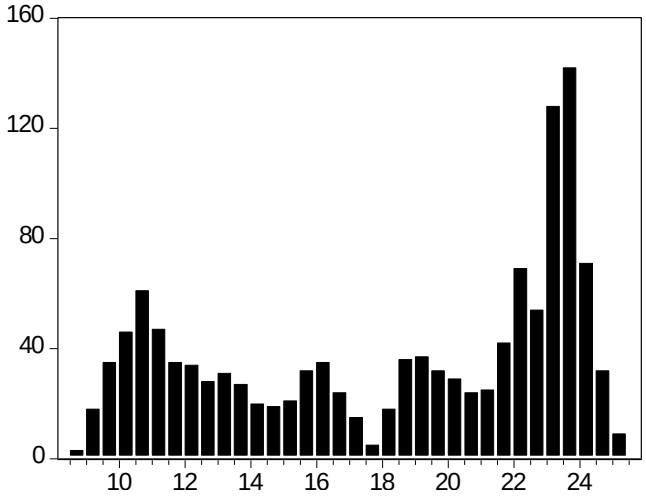


Fig. 4. Histogram of minimum temperature

Table 2: Testing stationarity of seasonally adjusted temperature series

Test	Test Statistic		1% Critical Value	5% Critical Value
	Max Temp	**Min temp**		

Augmented Dickey Fuller (ADF)	-40.821	-12.174	-3.438	-2.864
Philips and Peron (PP)	-12.190	-27.966	-3.438	-2.864

Structure of Autocorrelations

For a linear time series model, typically an autoregressive integrated moving average (ARIMA(*p*,*d*,*q*)) process, the patterns of autocorrelations and partial autocorrelations could indicate the plausible structure of the model. At the same time, this kind of information is also important for modelling nonlinear dynamics. The long lasting autocorrelations of the data suggest that the processes are nonlinear with time-varying variances. The basic property of a long memory process is that the dependence between the two distant observations is still visible. For the series of daily wholesale price, autocorrelations were estimated up to 100 lags, i.e., *j*=1,...,100. The autocorrelation functions of these series are plotted in figure 5-6. A perusal of figures indicate that, these do not decay exponentially over time span, rather, there is hyperbolic decay of the autocorrelations functions towards zero and they show no clear periodic patterns. There is no evidence that the magnitude of autocorrelations become small as the time lag, *j*, becomes larger. No seasonal and other periodic cycles were observed.

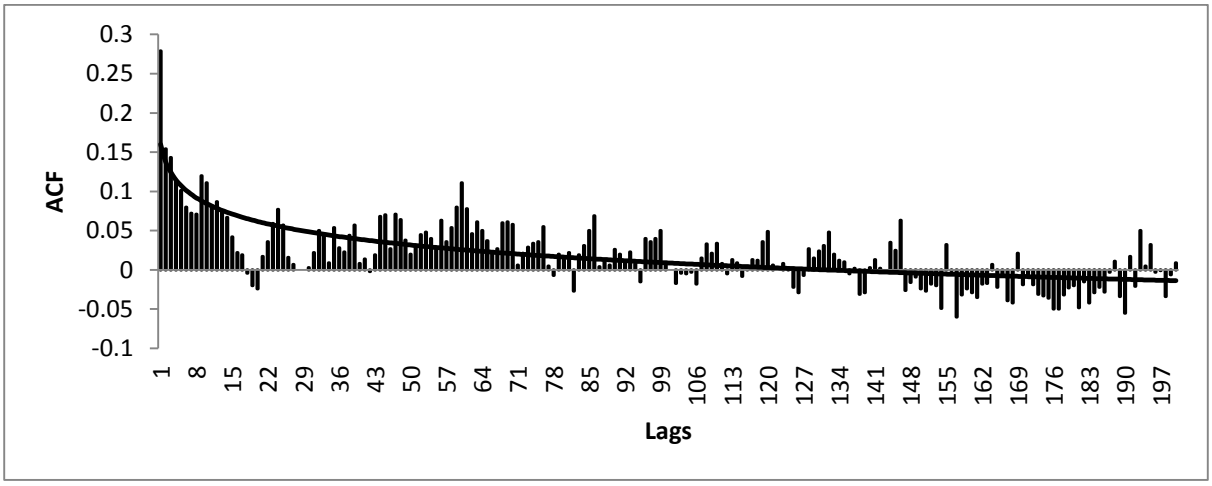


Fig. 5. Autocorrelation function (ACF) of seasonally adjusted Min temp series up to 200 lags

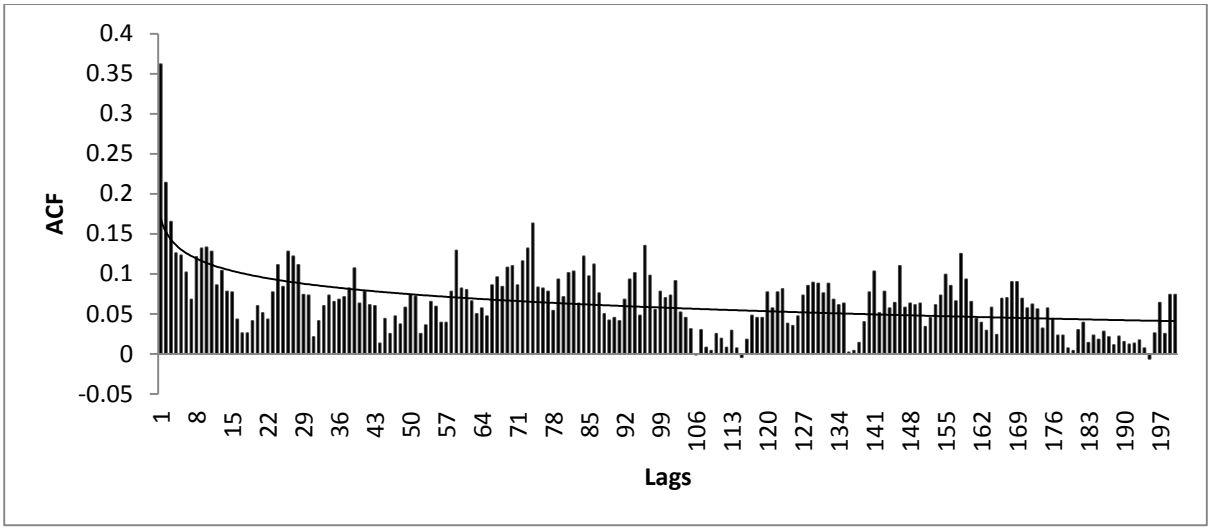


Fig. 6. Autocorrelation function (ACF) of seasonally adjusted Max temp series up to 200 lags

Accordingly, ARFIMA model was fitted to the above seasonally adjusted dataset. The best ARFIMA model has been selected on the basis of minimum Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC) values. It is found the both the maximum as well as minimum temperature series follow ARFIMA (0,d,0) process. The value of long memory parameter for maximum and minimum temperature are found to be 0.262 and 0.217 respectively. Both values are also significant at 1% level as reported in table 3.

Table 3. Parameter estimate of ARFIMA(0,d,0)

	Maximum Temperature			**Minimum Temperature**		
Parameters	**Estimate**	**Standard Error**	**P Value**	**Estimate**	**Standard Error**	**P Value**
const	30.025	0.137	<0.001	17.543	0.093	<0.001
d	0.262	0.022	<0.001	0.217	0.021	<0.001
Log likelihood	406.956			452.398		
Akaike Information Criteria (AIC)	-807.912			-898.797		
Bayesian Information Criteria (BIC)	3038.09			2947.200		

Diagnostic Checking and Validation

The model verification is concerned with checking the residuals of the model to see if they contained any systematic pattern which still could be removed to improve the chosen ARFIMA. For this purpose, autocorrelations of the residuals were computed and it was found that none of these autocorrelations was significantly different from zero at any reasonable level. This proved that the selected ARFIMA model was an appropriate model for forecasting the data under study.

One-step ahead forecasts of temperature series using naïve approach for the period January, 2007 to December, 2007 in respect of above fitted model are computed. For measuring the accuracy in fitted time series model, Root mean square prediction error (RMSPE), Mean absolute error (MAE) and Relative mean absolute prediction error (RMAPE) are computed by using the formulae given below and are reported in Table 4.

$$\text{MAE} = 1/12\sum_{i=1}^{12}\left|y_{t+i} - \hat{y}_{t+i}\right|$$

$$\text{RMSPE} = \left[1/12\sum_{i=1}^{12}\left\{\left(y_{t+i} - \hat{y}_{t+i}\right)^2\right\}\right]^{1/2}$$

$$\text{RMAPE} = 1/12\sum_{i=1}^{12}\left\{\left|y_{t+i} - \hat{y}_{t+i}\right| / y_{t+i}\right\}\times 100$$

Table 4: Validation of Models

Temperature	**MAE**	**RMSPE**	**RMAPE (%)**
Maximum	0.46	0.54	1.51
Minimum	0.44	0.56	2.72

A perusal of above table indicates that for both the temperature series data, RMAPE is less than 5% indicating the performance of the model is satisfactory. The fitted vs observed maximum as well as minimum temperature series are plotted in figure 7 and 8 respectively. Both the figures justify the accuracy of the model fitting.

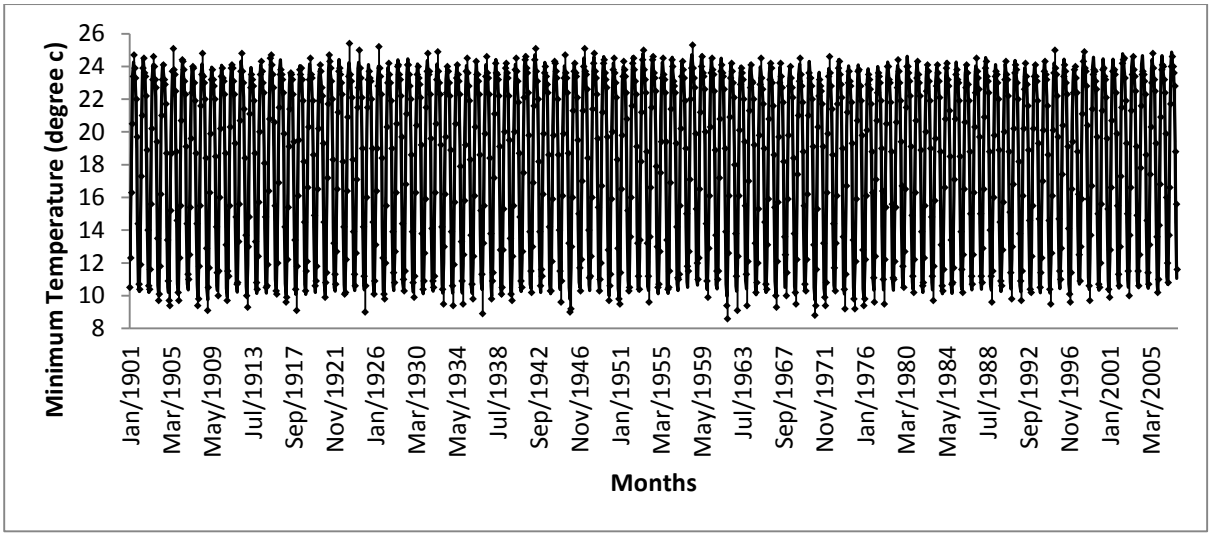


Fig. 7. Observed (marker) vs predicted (line) of minimum temperature

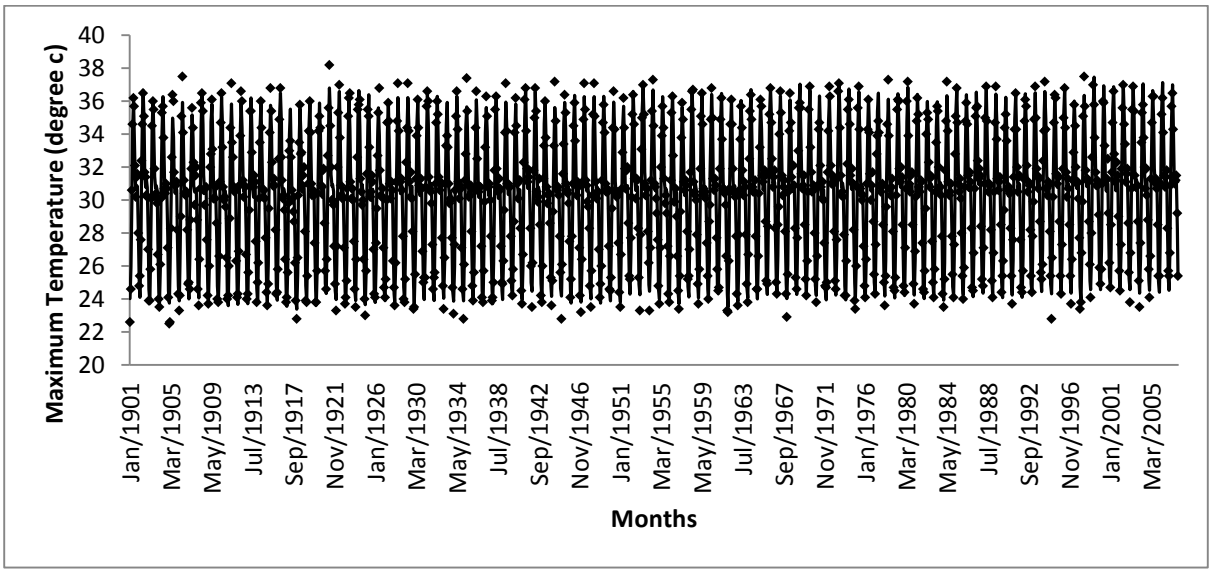


Fig. 8. Observed (marker) vs predicted (line) of maximum temperature

Modelling of Temperature Series by Wavelet Approach

For computation of MODWT of temperature series by Wavelet approach, methodology discussed in methodology section is followed. Here, we take J_0 as 7. Haar wavelet is used for analysing the data on a scale by scale basis to reveal its localized nature as exhibited by MRA

coefficients at level 7 in Fig. 9 and 10. A perusal indicates that localized variation in the data is detected at lower scale, whereas global variation is detected at higher scale. The wavelet coefficients are related to differences (of various order) of (weighted) average values of portions of X_t concentrated in time. Coefficients at the top (below) provide "high frequency" ("low frequency") information. Wavelet coefficients do not remain constant over time and reflects changes in the data at various time-epochs. Locations of abrupt jumps can be spotted by looking for vertical (between levels) clustering of relatively large coefficients.

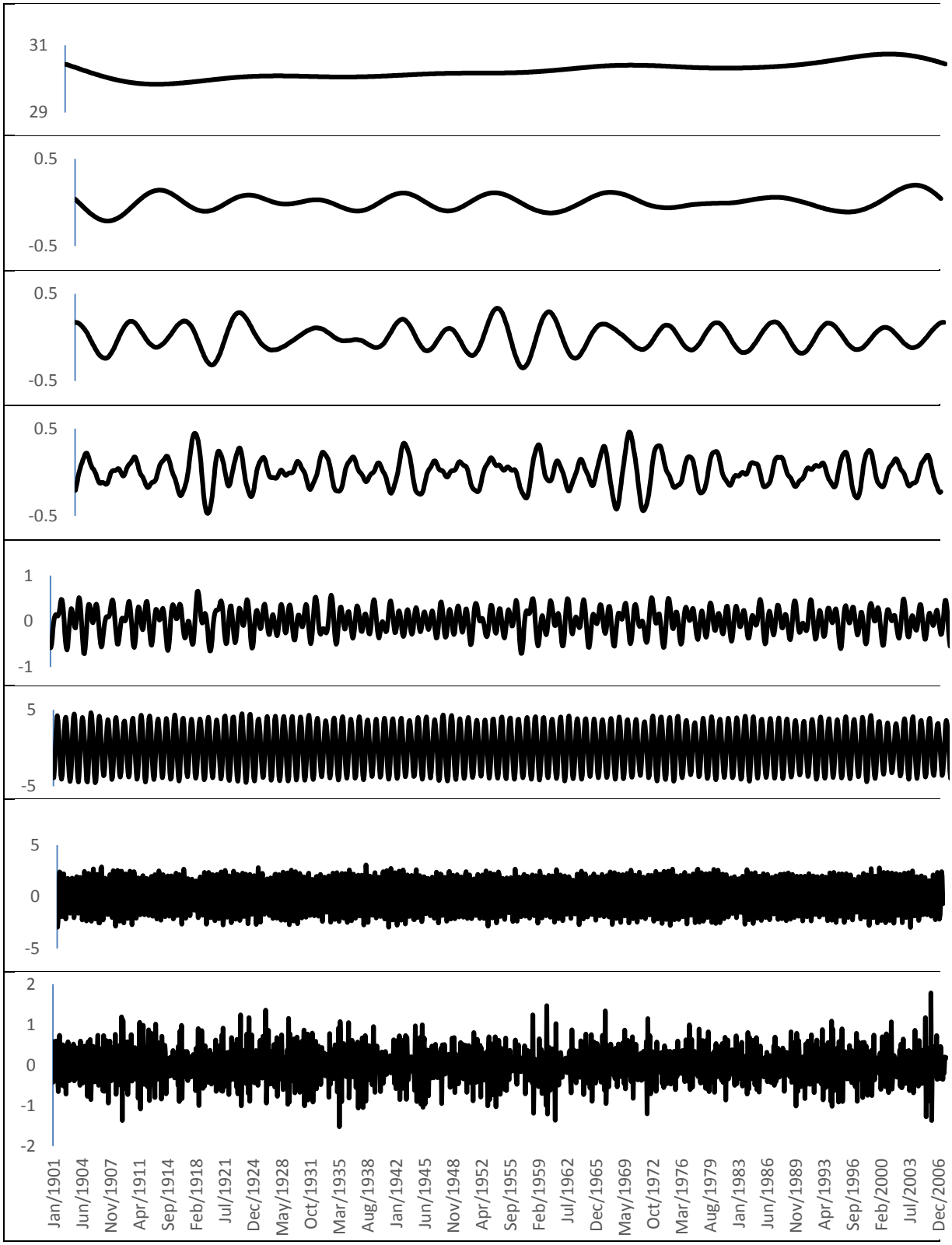


Fig. 9. MRA of Maximum temperature at level 7 D1, D2, D3, D4, D5, D6, D7 and S7 (From bottom to top)

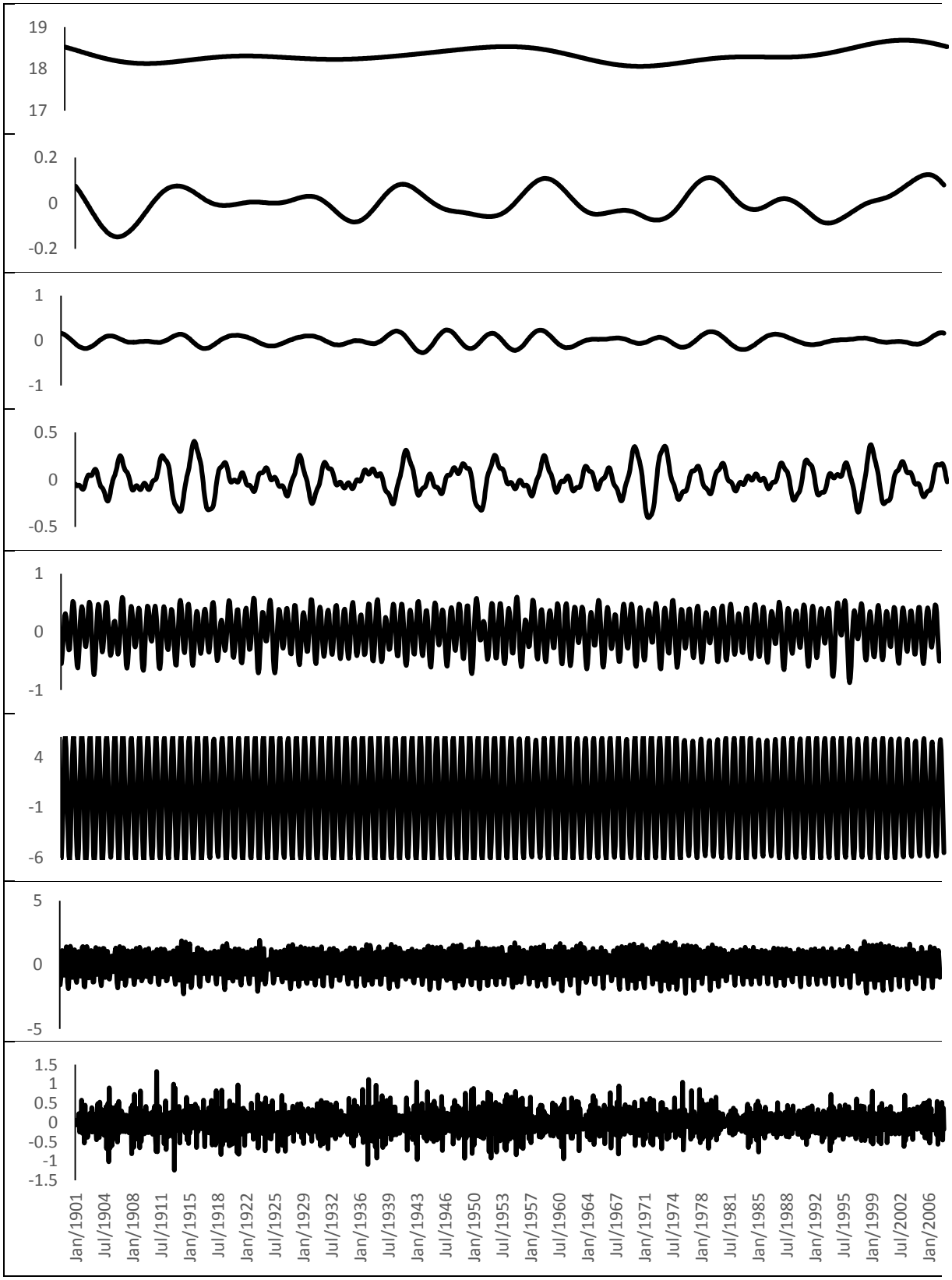


Fig. 10. MRA of Minimum temperature at level 7 D1, D2, D3, D4, D5, D6, D7 and S7 (From bottom to top)

Conclusion

Long memory time series have been analysed by using ARFIMA models. Model parameter *d* reflects the long memory in the maximum and minimum temperature series. It is found that in the both the series long memory parameter is significant. The study has revealed that the ARFIMA model could be used successfully for modelling the temperature series. The predictive ability of ARFIMA model was investigated in terms of relative mean absolute percentage error. The model has demonstrated a good performance in terms of explained variability and predicting power. Multiresolution analysis (MRA) was carried out to explore the local as well as global variations in both the temperature series over the years. The variability in minimum temperature is found to be more than maximum temperature. The study reveals that there are pockets of change in the temperature pattern (both in maximum as well as in minimum temperature) which may be clearly visible by vertical clustering of coefficients in MRA.

Chapter III

FRACTIONALLY INTEGRATED MAXIMUM TEMPERATURE SERIES IN INDIA IN PRESENCE OF STRUCTURAL BREAK

1. Introduction

Being an integral part of the natural environment, Climate continues to play a leading role in determining natural and human life. Climatic behaviours constitute a complex domain of learning, which is atypical and uncontrollable. Studying the past climate pattern or through simulation studies we can only understand its behavioural development, whereas sound statistical and numerical methodology paves an indispensable way. According to the various climate change impact assessment reviews, agriculture, vegetation, water resources and tourism are the sectors worst affected directly due to climatic vagaries and temperature being one of the most important climatic parameters. It has a direct impact on the evaporation, snow melting, frost and an indirect impact on the atmospheric stability and precipitation conditions (Ustaoglu *et al*., 2008). Therefore, there is a need to forecast temperature accurately in order to prevent unexpected hazards caused by temperature variation, such as frost and drought which may cause financial and human losses.

Climatic conditions have a crucial effect on the variability and production of crops while the economy of the region is based on agriculture. Climate change has become a global concern because of its potential threats to agricultural development. Agriculture is more sensitive to climate change (Mendelsohn *et al*., 2006). The consequences of climate change are predicted to be more severe for the developing countries like India on account of their heavy dependence on agriculture (Birthal *et al*., 2014). Available literatures suggest a considerable reduction in agricultural productivity due to climate change in most developing countries (DE Salvo *et al*., 2013; Birthal *et al*., 2014). In India, the literature on climate impacts on agriculture is limited but has been growing fast. Aggarwal (2009), using production function approach estimated that a 1.0°C rise in mean temperature would reduce yields of wheat, soybean, mustard, groundnut and potato by 3-7 per cent. The forecasting of surface

temperature has long been an area of interest of the scientific community in general, and meteorologists in particular. Since the late 1930s, different statistical methodologies have been attempted to forecast the surface temperature on hourly (Spreen, 1956), daily (Mantis and Dickey, 1945; Gilbert, 1953), monthly (Kangieser, 1959), and seasonal (Van Loon and Jenne, 1975; Kumar *et al*., 1997) time scales. Paul *et al*. (2015) have investigated the trend in mean temperature in different agro-climatic zones in India. Mills (2016) has analyzed the temperature series of Kefalonia with Seasonal autoregressive-integrated moving average (SARIMA) model.

It is a well-known fact that temperatures are time-dependent. Mathematically, there exist different ways of modelling that dependence and a key issue here is to determine if the series is stationary (around a linear time trend) or not (Gil-Alana, 2008). Climate memory has been a well-known statistical concept since last several decades. Affected by the slowly responding subsystems (such as the ocean), the climate variability usually exhibits long-term memory, which means the present climate states may have long-term influences on the states in far future (Malamud *et al*., 1999). Normally, this kind of characteristic is considered as fractal properties or scaling behaviors in climate (Yuan *et al*. 2014). Rohini *et al*., (2016) investigated the variability and increasing trends of heat waves over India. They found that the anomalous persistent high with anti-cyclonic flow, supplemented with clear skies and depleted soil moisture are primarily responsible for the occurrence of heat waves over India. Variability of heat waves is also influenced by both the tropical Indian Ocean and central Pacific SST anomalies. Nowadays, an important new environmetric application for long range dependence (LRD) is to climate research. Here Autoregressive fractionally integrated moving average (ARFIMA) model plays an important role in understanding long-term climate variability and in trend estimation, but remains less well known in the researcher's communities compared to, for example, Autoregressive-integrated moving average (ARIMA) models of the Box-Jenkins type, of which AR(1) is still the most frequently applied model. A number of studies suggest that atmospheric temperature have long-range memory. Although there is increasing evidence of the presence of LRD in such time series, a precise physical explanation of the phenomenon has been elusive. Studies of instrumental data indicate that sea surface temperature is more persistent than air temperature over land (Pelletier, 1997, Eichner *et al*., 2003, Monetti *et al*., 2003, Lennartz and Bunde, 2009), so ocean dynamics seem to be an important component. Spatial averaging also influences the persistence, as temperature averages over larger regions are more persistent than local data (Lennartz and Bunde, 2009). Some studies also indicate that persistence is largely close to the equator, and

is reduced with location closer to the poles (e.g., Pattantyús-Ábrahám *et al*., 2004, Huybers and Curry, 2006, Vyushin and Kushner, 2009).

India is situated north of the equator between 8°4' to 37°6' north latitude and 68°7' to 97°25' east longitude. In the south, India projects into and is bounded by the Indian Ocean in particular, by the Arabian Sea on the southwest, the Lakshadweep Sea to the south, and the Bay of Bengal on the southeast. Present paper discusses the modeling aspect of Maximum temperature series of India which possess a long memory pattern. For modeling such long-range persistence pattern in climatic behaviours, traditional Autoregressive-integrated moving average (ARIMA) or SARIMA, termed as short-memory models are not suitable to capture the long term behaviour of the series. Moreover, possible structural shift in the temperature series has also been identified. Now, to deal the coexistence of long memory and/or structural break use of ARFIMA is mostly observed (Gil-Alana, 2005). Wang *et al*. (2013) investigated that for series with possible break in the mean, ARFIMA model may not lead optimal forecasting result (), and they proposed infinite AR truncation method to deal with the situation. Papailias and Dias (2015) reported that the AR truncation method may also lead to loss of information and accordingly they have proposed two stage forecasting (TSF) approach for this situation. Hence, TSF approach has been addressed to deal the modeling and forecasting of maximum temperature series possessing the property of long memory with structural break which is claimed to be more robust to tackle the problem of the coexistence of long memory and structural shift. Besides this, Wavelet analysis tool has been envisaged to capture the variation in maximum temperature series in both time and frequency scale.

2. Long Memory: Concepts and Forecasting

2.1 Basic Concept

Long memory processes generalize linear ARIMA models by allowing for non-integer differencing parameter and thereby providing a more flexible framework for analyzing time series data. This flexibility enables fractional processes to model stronger data dependence than that is allowed in stationary ARMA models. The autocorrelation functions of ARMA or conventional short memory processes decay faster (exponentially) than the fractionally integrated processes which decay hyperbolically and are therefore known as long memory processes. Long memory in time series can be defined as autocorrelation at long lags. The autocorrelation function (acf) of a time-series x_t is defined as, $\rho_k = \text{cov}(x_t, x_{t-k}) / \text{var}(x_t)$ for

integer lag *k*. A covariance stationary time-series process is expected to have autocorrelations such that $\lim_{k\to\infty}\rho_k = 0$. For short memory process $\rho_k = |m|^k$, where $|m| < 1$ and this property is true, for example, for the well-known stationary and invertible ARMA structure. For long memory processes, the autocorrelations decay at a hyperbolic rate which is consistent with $\rho_k \approx Ck^{2d-1}$, as *k* increases without limit, where *C* is a constant and *d* is the long memory parameter. For more details on long memory one can refer to Beran (1995).

2.2 Existing Framework

Long memory process for x_t can be modeled as a fractionally integrated *I(d)* process

$$(1-L)^d x_t = \varepsilon_t,\ t=1, 2,\ldots, T \tag{1}$$

where *L* denotes the lag operator, *d* is the fractional difference parameter and ε_t is white noise with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_t \varepsilon_s) = 0 \forall t \neq s$. The process x_t is stationary and invertible provided *-1/2<d<1/2,* $d \neq 0$ and exhibits non-stationary long memory if *1/2≤d<1*.

A flexible parametric process called the ARFIMA (*p, d, q*) model incorporates both long term and short term memory,

$$\phi(L)(1-L)^d x_t = \theta(L)\varepsilon_t,\ t=1, 2,\ldots, T \tag{2}$$

where $\phi(L) = 1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p$ and $\theta(L) = 1+\theta_1 L+\theta_2 L^2+\ldots+\theta_q L^q$ are the autoregressive and moving average polynomials respectively, with roots lie outside the unit circle and ε_t is white noise.

Now, let us represent x_t as infinite autoregressive process

$$x_t = \sum_{i=1}^{\infty} \beta_i x_{t-i} + \varepsilon_t,\ t=1, 2,\ldots, T \tag{3}$$

where $\beta_i = \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(d)}$, with $\Gamma(.)$ being the gamma function.

For AR-truncation method, theoretical *h*-step head forecast is obtained from canonical representation of the model using a truncated version of the autoregressive weights as

$$\hat{x}_{T+h} = \sum_{i=1}^{P} \hat{\beta}_i x_{T+h-i} \tag{4}$$

Akaike Information Criteria (AIC) or other different criterions can be used to determine the lag order *P*.

Wang *et al.* (2013) have recommended the use of AR approximation method as an alternative to the ARFIMA set up for modeling long memory process with spuriousness. This methodology suggests to apply an autoregressive model directly to the long memory series with order *k* chosen by AIC or C_p (Malllows, 1973) statistic.

$$\hat{x}_{T+h} = \sum_{i=1}^{k} \hat{\beta}_i^c x_{T+h-i} \tag{5}$$

where, $\hat{\beta}_i^c$ denote the autoregressive parameter corresponding to i^{th} lag.

For long memory series with a break in the mean at point *b,* the process is represented as follows

$$y_t = \begin{cases} \mu + x_t, 1 \le t \le b \\ \mu + \Delta + x_t, b+1 \le t \le T \end{cases} \tag{6}$$

There are various popular approaches for estimating long memory parameter such as Rescaled Range Analysis (Hurst, 1951), GPH (Geweke and Porter-Hudak, 1983), Maximum Likelihood Method of Estimation (MLE) (Sowell, 1992), Wavelet method (Jensen, 1999) etc. Paul *et al*. (2015) have demonstrated the performance of different estimation techniques of long memory parameter by means of Markov Chain Monte Carlo (MCMC) and reported that wavelet method performs better than the other methods.

2.3 A two-stage forecasting Algorithm

Let us specify the ARFIMA model as

$$x_t = (1-L)^{-d} \left(\phi(L)\right)^{-1} \theta(L)\varepsilon_t \tag{7}$$

The short run effects are obtained by setting $d = 0$ and the behavior of the fractionally differenced process can be described from $(1-L)^d x_t$.

TSF approach is suggested by Papailias and Dias (2015) to deal with spurious long memory. The procedure can be envisaged in the underlying steps:

1. Estimate the long memory parameter via any consistent method of estimation to obtain fractional difference operator $\hat{d}$ such that $(\hat{d} - d) = O_p(T^{-\delta})$, where $\delta > 0$. Then apply the differencing operation to have the weakly dependent process η_t,

$$\eta_t = (1-L)^{\hat{d}} x_t, \; t = 1, 2, \ldots, T \tag{8}$$

2. Fit an AR(*P*) to η_t and compute the one step ahead forecast

$$\hat{\eta}_{t+1} = \sum_{i=1}^{P} \hat{\alpha}_i \eta_{t+1-i},\ t=1, 2,\ldots,T \tag{9}$$

where $\hat{\alpha}_i$ is the autoregressive term associated with *i* th lag. The lag *P* can be selected in two different ways either via AIC or using higher order with respect to sample size, $\hat{P} = \left\lfloor \left(\ln T\right)^2 \right\rfloor$; $\lfloor . \rfloor$ denoting the integer part.

3. Expand the weakly dependent series η_t including the one step ahead forecast, such that $\tilde{\eta}_t = (\eta_1,...,\eta_t,\hat{\eta}_{t+1})'$. Apply the fractional cumulation operator using the estimated $\hat{d}$ and obtain the one step ahead forecast $\tilde{x}_{t+1}$ for original long memory series.

$$\tilde{x}_t = (1-L)^{-\hat{d}} \tilde{\eta}_t, t = 1,2,...,T,T+1 \tag{10}$$

4. Iterating over the previous steps recursively provide the *h* step ahead forecast $\tilde{x}_{t+h}$ for the long memory process.

The main advantages of using TSF approach in modeling long memory series is two-fold, first, weakly dependent series obtained through fractional differencing operation, is less persistent as compared to the original long memory series and hence the idea of implementing AR based forecast to the weakly dependent process η_t enable TSF method to provide better forecast than the existing methodologies. Secondly, employing AR-truncation approach or relying on the post-break data to provide forecasting results may lead to loss of information which is avoidable through TSF technique.

3. Results and Discussion

Seasonally adjusted all India monthly maximum temperatures (in degree centigrade) for the period January, 1901 to December, 2007 has been used. Out of total 1284 observations, 1272 observations have been used for estimation of models and the remaining 12 observations for validation purpose. The data has been obtained from Indian Institute of Tropical Meteorology (IITM), Pune, India (http://www.tropmet.res.in). For computation of monthly maximum temperature in India, IITM considers the data of 388 well-spread stations of India Meteorological Department (IMD). They first interpolate the station wise monthly temperature anomaly time series onto a 0.5° x 0.5° grid. Then, the climatological normals of temperature at all stations are interpolated onto the same grid, resulting in high-resolution grid point temperature climatology for the country. The gridded monthly anomaly values are

then added to the gridded climatology, finally producing a long-term gridded data set of actual temperatures for India for the period 1901-2007. India's monthly temperature series are computed by simple averages of the constituent grid point data of the respective regions. For more details regarding the procedure see Kothawale and Rupakumar (2005).

3.1. Descriptive Statistics

Table 1 reports the descriptive-statistics of the maximum temperature series. It is clear from table 1 that the mean maximum temperature is 30.24^0C. The variability as expressed by coefficient of variation (CV) is around 13%. And the maximum temperature series is negatively skewed and leptokurtic. The time series plot of seasonally adjusted temperature data is exhibited in figure 1. A perusal of the plot indicates that the data series is stationary. In order to test for stationarity, Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests have been performed and the result is reported in table 2. The stationarity test results reveal that the series is weakly stationary.

Table 1. Descriptive statistics of temperature series (degree centigrade)

Statistics	Temperature series	Statistics	Temperature series
Mean (^{0}C)	30.24	Standard Deviation (^{0}C)	3.81
Median (^{0}C)	30.7	CV (%)	12.60
Maximum (^{0}C)	38.2	Skewness	-0.12
Minimum (^{0}C)	22.5	Kurtosis	2.13

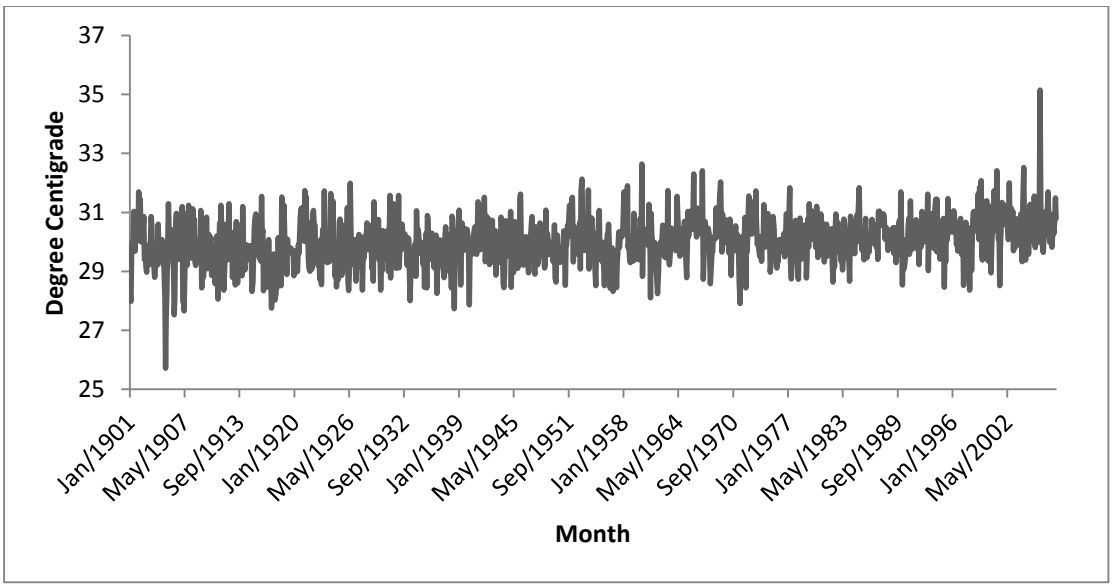


Figure 1. Seasonally adjusted monthly maximum temperature series

Table 2. Testing for stationarity

Temperature series	ADF Test			PP Test		
	Test statistic	P-value	Result	Test statistic	P-value	Result
Single Mean	11.52	<0.001	Stationary	-1344.04	<0.001	Stationary
With Trend	28.55	<0.001	Stationary	-1006.50	<0.001	Stationary

3.2. Change Point Detection

The usual assumption of constant mean and variance over time for a time series data is often violated in practice. Sometimes, the series shows an interesting behaviour of stationarity for some time, then suddenly the variability of the error term changes, it stays constant again for some time at this new value, until another change occurs. Detection of change points in a series is very important to correctly interpret the behaviour of the series. For single structural change one can use the At most one Change point (AMOC) algorithm (Killick *et al*., 2014). Paul *et al.* (2014) applied different techniques for identification of structural break in mean temperature of different agro-climatic zones in India. Werner *et al*. (2015) modeled global and hemisphere temperature anomalies by piecewise linear regression and found break points in the temperature evolution. Paul (2017) has investigated the presence of long memory in maximum and minimum temperatures in India. In the present study, an attempt has been made to detect the structural break in mean for seasonally adjusted monthly maximum temperature series. By using AMOC algorithm, it is found that there is a significant break in maximum temperature during July, 1963 which may be due to drought year as found in Nagarajan (2009). According to study by Paul *et al*. (2014), anthropogenic activities are the main drivers behind climate change. The changes in land use, intensive cropping, deforestation, and industrial development have led to a rise in carbon-di-oxide emissions and a structural break in temperature series. It is to be noted here that overall trend in the maximum temperature series is positive and statistically significant. The trend coefficient is found out to be 0.00064 i.e. there is an average increase of 0.00064^0C temperature per month in the maximum temperature in India. When the trend estimation was computed in two phases i.e. during 1901 to 1962 and during 1963 to 2007, the respective trend component found out to be 0.00039 and 0.00083. Both the trend are statistically significant. It is clear that after 1962 the rate of increase in maximum temperature is higher in comparison to the

years up to 1962. Figure 2 displays the shift in mean temperature in the time series plot for the entire study period.

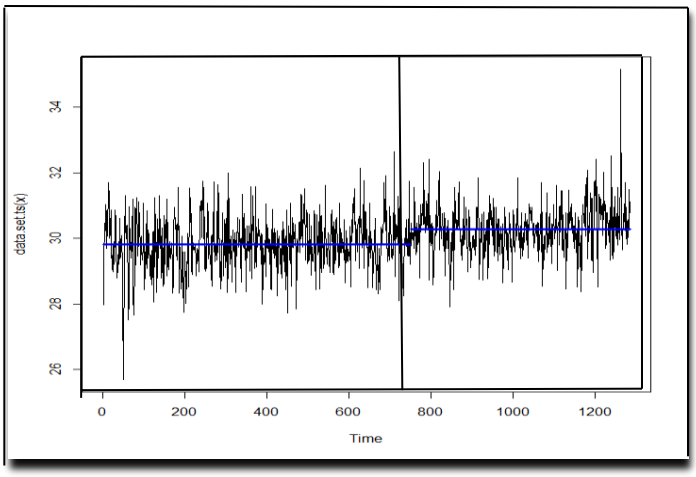


Figure 2. Change point in mean for the maximum temperature series

3.3. Quantitative Analysis of Long Memory Process

Estimating the long memory parameter (*d*) is the milestone of modeling long memory property. The hyperbolically slow decay of autocorrelation function (ACF) and partial autocorrelation function (PACF) initially gives an indication of possible presence of long memory in a time series. But to be sure about the presence of long memory and estimating the value of long memory parameter one should rely on statistical techniques as described in section 2.2, Figure 3 depicts the ACF and PACF of the temperature series estimated up to 500 lag respectively. As clearly indicated, correlogram is following a hyperbolic decay; the correlation decreases very slowly, a typical shape for time series having the long memory property. The estimated long memory parameter has been found to be 0.266 which is within the range of persistence and highly significant at 1% significance level. The estimated long memory parameter value is less than 0.5 ensures the stationary long memory phenomena of the series.

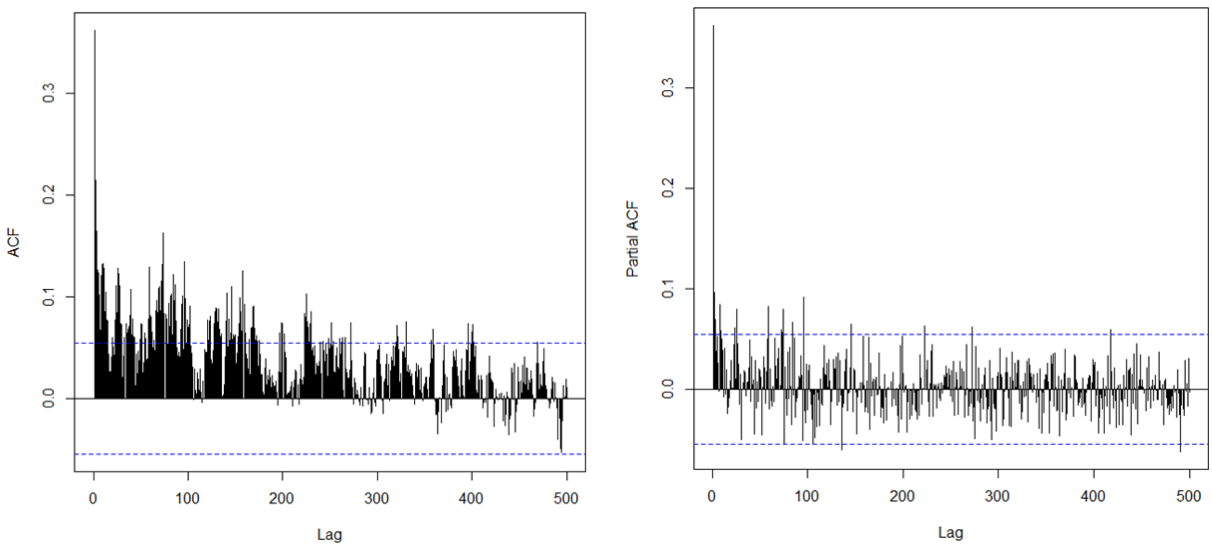
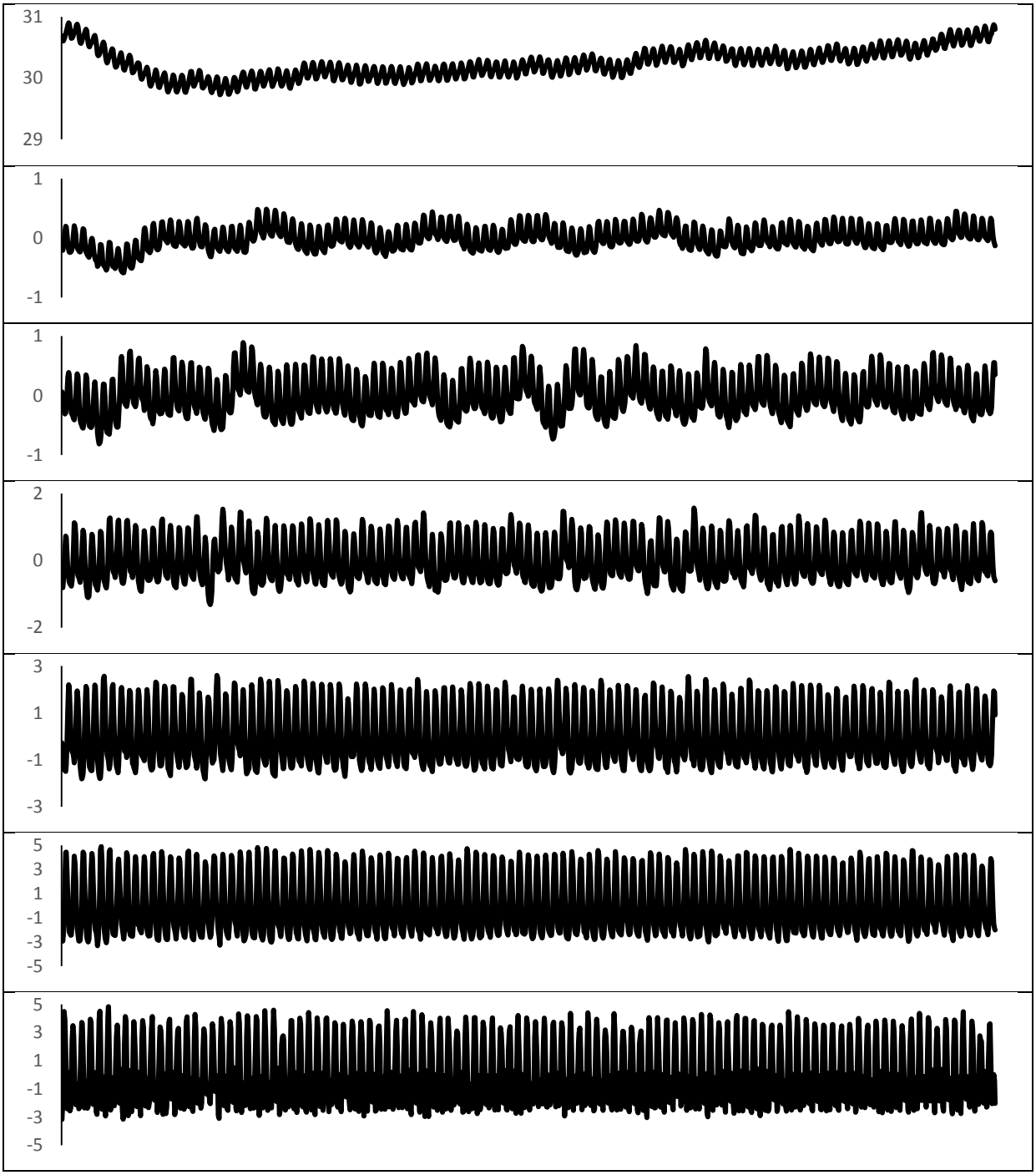


Figure 3. Correlogram of temperature series

3.4. Examining variation in Temperature Series by Wavelet Approach

For computation of Maximal overlap discrete wavelet transform (MODWT) of temperature series by Wavelet approach, methodology by Percival and Walden (2000) is followed. The MODWT can accommodate any sample size *N*. In practice, largest level (J_0) is commonly selected such that $J_0 \leq log_2(N)$ in order to preclude decomposition at scales longer than total length of the time-series. Here, we take J_0 as 7. Haar wavelet is used for analyzing the data on a scale by scale basis to reveal its localized nature as exhibited by MODWT coefficients at level 7 in Figure 4. A perusal indicates that localized variation in the data is detected at lower scale, whereas global variation is detected at higher scale. The wavelet coefficients are related to differences (of various order) of (weighted) average values of portions of X_t concentrated in time. Coefficients at the top (below) provide "high frequency" ("low frequency") information. Wavelet coefficients do not remain constant over time and reflects changes in the data at various time-epochs. Locations of abrupt jumps can be spotted by looking for vertical (between levels) clustering of relatively large coefficients. The trend in maximum temperature is detected as top most plot of figure 4. As in the wavelet, the series is smoothen, therefore the break in mean may not be prominent in this case. A

closer look at the top most plot in figure 4 indicates that initially there was decreasing trend in maximum temperature series till 1914 then started increasing and after 1962 the rate of increase in temperature becomes higher.



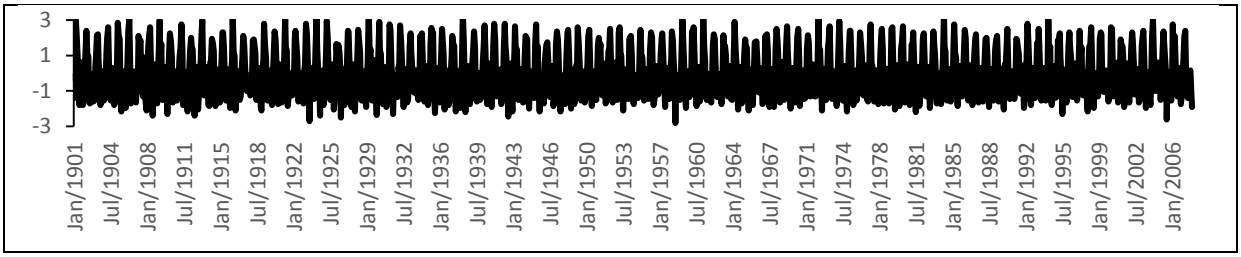


Figure 4. MODWT of Maximum temperature at level 7 (W_1, W_2, W_3, W_4, W_5, W_6, W_7 and V_7 From bottom to top)

3.5. Modeling of the temperature series

For the seasonally adjusted stationary temperature series data, presence of long range dependence has been established including the presence of a break in the mean of the series. We compute multi-step ahead forecasts using the following methodologies:

(i) **AR(AIC)-Trunc:** Truncate the infinite autoregression representation up to a specific lag chosen by the AIC (Wang *et al*., 2013), as in equation (4).

(ii) **AR(AIC):** Fit an AR(*k*) model to the original long memory process with order *k* selected by AIC (Wang *et al*., 2013) using equation (5).

(iii) **TSF:** Using TSF algorithm with order chosen by AIC for the underlying weakly dependent process (Papailias and Dias, 2015).

Table 3 represents the parameter estimates of AR(AIC)-Trunc and AR(AIC) model. For the AR-Trunc model, the optimum lag was selected as 2 and for AR(AIC) model lag 3 was chosen as indicated from minimum AIC criterion. Both this models have been compared with the TSF approach empirically for the aforementioned temperature series. In case of TSF algorithm parameter estimates varies in each subsequent steps including AR parameters and fractional difference operator.

Table 3. Parameter estimates of the AR-Trunc and AR model

Model	Parameters	Estimate	Probability
AR(AIC)-Trunc	***d***	0.26	<0.0001
	AR 1	0.70	<0.0001
	AR 2	0.10	0.0003
AR(AIC)	***d***	0.26	<0.0001
	AR 1	0.32	<0.0001
	AR 2	0.08	0.01

	AR 3	0.07	0.018

3.6. Evaluation of Forecast Performance

The forecasting performance of TSF procedure has been computed for an out of-sample cross-validation period of 12 observations (i.e., 12 months). Predictive abilities of different models have been compared using SSE, MSE and RMAPE for different forecast horizons. The corresponding results of forecast comparison have been reported in table 4.

Table 4. Comparing predictive abilities of various models

Forecast Horizon	Models	AR(AIC)-Trunc	AR(AIC)	TSF
h=4	**SSE**	2.37	2.27	2.13
	MSE	0.59	0.58	0.53
	RMAPE (%)	1.99	1.98	1.81
h=8	**SSE**	2.71	2.65	2.54
	MSE	0.34	0.34	0.32
	RMAPE (%)	1.37	1.40	1.38
h=10	**SSE**	3.23	3.28	2.75
	MSE	0.32	0.33	0.27
	RMAPE (%)	1.38	1.45	1.26
h=12	**SSE**	5.85	6.08	4.49
	MSE	0.49	0.51	0.37
	RMAPE (%)	1.73	1.81	1.51

Comparing the validation results of all the three models, it is observed that TSF produces the best results over the other methods. For AR(AIC)-Trunc method, there is possibility of loss of information due to truncation at specific lag. Moreover, for this procedure break may be mistaken as long memory, so sometimes may render moderately good result and provide minor differences with the result of AR(AIC). Overall, for the aforesaid maximum temperature series data, TSF methodology provides smallest forecast errors in most of the

forecast horizons. Employing the fractional filtering, the presence of break is smoothed, which is the key strategy of TSF that makes the out-performance of the TSF approach over the other methodologies.

4. Conclusions

The surface temperatures over a given region vary seasonally and annually depending upon latitude, altitude and location with respect to geographical features such as a water body (river, lake or sea), mountains, etc. Probably one of the most widely quoted aspects of climatic change is the significant increase in global mean air temperature during the past century. Since the hydrologic cycle is a thermally driven system, rise in global temperature is likely to accelerate this cycle. Demand for the identification of the temperature trend/shift and their projection has been growing substantially. Present paper addresses the modeling issue of maximum temperature series in India which possesses characteristic of long-term memory and TSF has proven to be a robust approach of capturing such long memory even in the presence of detected structural break. Time series modeling of temperature series assume prime importance both in the local and global levels. One of the major significance is the worldwide interest in the issues of global warming and climate change. Wavelet transformation was applied to decompose the temperature series into time–frequency domain in order to study the local as well as global variation over different scale and time epochs. It is found that there is significant increase in maximum temperature over the years in India. The shift in maximum temperature in India occurred during mid of 1963 as detected by the statistical test. Temperature data series exhibiting long-range dependence property combined with structural break required to be modeled with a concrete and valid technique which can overcome the issue of loss of information, biased estimates and inaccurate forecast. In this regard, TSF approach has been found to serve the better results. For long memory processes with a change in mean level, the present study on maximum temperature data has established the outperformance of TSF methodology in terms of SSE, MSE and RMAPE criterions over both AR-Truncation and AR(AIC) approach. Implementation of TSF has been found to be more robust with regard to multi-step ahead forecast for spurious long memory. Further, the TSF approach can be employed for forecasting of temperature series in different agro-climatic zone and existence of structural shifts in different zones of India can be investigated.

Chapter IV

Wavelets Based Estimation Of Trend In Sub-Divisional Rainfall In India

1. Introduction

A common issue in time series analysis is the decomposition of several time series components viz. low frequency (trend) component and high frequency (noise and periodicity) component. Most of the time series of aggregated variables show a steadily increasing or decreasing pattern, known as trend. Over last few decades Box Jenkin's Autoregressive integrated moving average (ARIMA) methodology (Box *et al*., 2007), a parametric approach of time series analysis has been used for forecasting time series data. But there are certain circumstances where it is not possible to postulate appropriate parametric relationship for the underlying phenomena; in this case nonparametric approach is called for.

A plausible statistical model for such series can consist of non-stochastic or trend component and stochastic component: $Y(t) = T(t) + X(t)$, where $Y(t)$ is the value of an observable time series at time *t*, $T(t)$ is the trend component and $X(t)$ is called the noise process with mean zero. Trend assessment is the problem of determining whether or not $T(t)$ is actually present in the time series, i.e. test the null hypothesis H_0: $T(t)= 0$. In many agricultural data, like daily commodity price data, daily rainfall and temperature data it is seen that the distant observations are dependent that means the dataset have characteristic feature of long memory or long range dependency. A time-series process is called as long memory or Fractional differenced (FD) process if the autocorrelation function decays very slowly towards zero unlike the exponential decay in usual ARIMA model. A major problem in practice is the discrimination of a FD(d) (d denotes the long memory parameter) process with trend from a FD(d) process without trend, since FD(d) processes for $0 < d < ½$ usually looks like having some trend in the series. This makes it hard to distinguish the existence of trend in the time series.

Here our interest lies on the testing of low-frequency part or called as trend, in wavelet domain (Antoniadis 1997). Let us consider the assumptions that the trend T(t) is well approximated at least locally by a low order polynomial (such as linear) and that the stochastic component X(t) is a FD process. Under these assumptions, the discrete wavelet transform (DWT) (Percival and Walden 2000) based upon the Haar family of wavelet filters can be used to transform the series Y(t). The ability of the DWT to cleanly separate Y(t) into these components allows us to test for its significance.

Almasri *et al*. (2010) have proposed a test statistic by using wavelet decompositions to test the significance of trend in a time series data. The most difficult problem of testing for linear trend is the presence of dependence among the residuals because of which, tests for trend based on the classical ordinary least squares (OLS) regression are inappropriate. In many situations, the error autocovariance function exhibits a slow decay reflecting the possible presence of long memory process. The wavelet analysis, however, has been extensively used for such purposes, since it suitably matches the structure of these processes. Paul *et al*. (2011) studied wavelet methodology for estimation of trend in Indian monsoon rainfall time-series data and reported significant trend. Indian agriculture is mainly rain-dependent; approximately two-third net cropped area is under rain-fed cultivation. Therefore, regular and uniform rainfall pattern is crucial for betterment of agriculture; extreme situations can affect the agriculture devastatingly. But problem is that the regular rainfall pattern may not be observed all the time. Several studies have been done on the pattern of regional rainfall in India (Kumar *et al*. 2004; Gowsami *et al*. 2006). Guhathakurta and Rajeevan (2006) have found that there is a decline in rainfall in the months of June, July and September and an increase in the month of August in few sub-divisions in India. Goswami *et al*. (2006) have studied the rainfall behaviour over central India of the period 1951-2000 and found that a significant increase in heavy rainfall events and a decrease in moderate rainfall events. Some studies have also carried out to predict the rainfall and estimate the trend using some non-parametric methods like wavelet (Kallache *et al*., 2005). 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 wavelet approach for describing variability in rainfall in different agro-climatic zones of India. Paul (2017) found the significance presence of long memory in

maximum and minimum temperatures in India. In this study we use wavelet based test statistic and non-parametric Mann-Kendall test statistic to test the existence of trend in the time series data. The paper is organized as follows: section 2 describes about long memory process; section 3 deals with the basics of wavelets; section 4 describes methods of detection of trend; section 5 deals with empirical illustration followed by conclusions in section 6.

2. Long memory process

Most of the research works in time-series analysis assume that the observation separated by long time span are independent of each other or nearly so. But in many practical situations it is seen that many empirical economic series show that the distant observations are dependent, though the correlation is small but not negligible. Let X_t; $(t = 0,1,2, \dots)$ be a stationary time-series process and the autocorrelation function of the time-series with a time lag of k is ρ_k. 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. A study of long memory time series in climate can be found in Paul and Anjoy (2018).

3. Wavelet

Wavelets (Daubechies 1992; Ogden 1997 and Vidakovic 1999) are fundamental building block functions, analogous to the trigonometric sine and cosine functions. As with a sine or cosine wave, a wavelet function oscillates about zero. This oscillating property makes the function a *wave*. However, the oscillations for a wavelet damp down to zero, hence the name *wavel*et. If $\psi(.)$ be a real valued function defined over the real axis $(-\infty, \infty)$ and satisfying two basic properties: the integral of $\psi(.)$ is zero and the square of $\psi(.)$ integrates to unity. The detail discussion of wavelets can be found in (Percival and Walden 2000).

3.1. Discrete Wavelet Transform (DWT)

DWT of a time-series observation is used to capture high and low frequency components. The DWT re-expresses a time-series in terms of coefficients that are associated with a particular time and a particular dyadic scale 2^{j-1} (Nason *et al*. 1999). Let, $\mathbf{X} = (X_0, X_1, \dots, X_{N-1})'$ be a column vector containing *N* observations of a real-valued time series, where we assume that *N* is an integer

multiple of 2^M, where M is a positive integer. The DWT of level J is an orthonormal transform of $\mathbf{X}$ defined by

$$\mathbf{d} = \left(\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_j, \ldots, \mathbf{d}_J, \mathbf{s}_J\right)' = \mathbf{WX} \tag{1}$$

where $\mathbf{W}$ is an orthonormal $N \times N$ real-valued matrix, i.e. $\mathbf{W}^{-1} = \mathbf{W}'$ so $\mathbf{W}'\mathbf{W} = \mathbf{W}\mathbf{W}' = \mathbf{I}^N$, and called the wavelet matrix. $\mathbf{D}j = \{d_{j,k}\}$, $j = 1, 2, \ldots, J$, are $N/\lambda j \times 1$ real-valued vectors of wavelet coefficients at level j associated with scale λ_j and location k, where $\lambda_j = 2^j$. The real-valued vector $\mathbf{S}_J = \{s_{J,k}\}$ is made up of $N/2^J$ scaling coefficients. Thus, the first $N - N/2^J$ elements of $\mathbf{D}$ are wavelet coefficients and the last $N/2^J$ elements are scaling coefficients, where $J \leq M$. In DWT several filters can be used, "Haar filter" is one of them.

4. Trend

Trend is defined as a long term change in the underlying mean level per unit time (Jain and Kumar, 2012). There are different models for describing various forms of trend, both linear and nonlinear, and both stochastic and deterministic. The model that represents the time series by using a j^{th}-order polynomial function is given as

$$T_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_j t^j \tag{2}$$

when j = 0, there is no long-run increase or decline in the time series over time, and j = 1, implies that there is a straight-line long-run growth (if $a_1 > 0$) or decline (if $a_1 < 0$) over time.

4.1. Mann-Kendall Test

The non-parametric Mann-Kendall test is commonly employed to detect monotonic linear trend (Jayawardane *et al*., 2005) in time series data. The null hypothesis, H_0, is that the data come from a population with independent realizations and are identically distributed. The alternative hypothesis, H_1, is that the data follow a monotonic trend. The Mann-Kendall test statistic is calculated according to

$$S = \sum_{k=1}^{n-1} \sum_{j-k+1}^{n} \operatorname{sgn}\left(X_j - X_k\right); \qquad \text{with,} \qquad sgn(x) = \begin{cases} 1 \; if \; x > 0 \\ 0 \; if \; x = 0 \\ -1 \; if \; x < 0 \end{cases}$$

where, sgn is the signum function. If S is greater than zero, trend is said to be increasing otherwise if S is less than zero trend is decreasing. More details of the test can be found in Paul *et al*. (2014).

4.2. Wavelet-based estimation of the trend in presence of long memory

Yajima (1988) considered a polynomial regression consisting of a polynomial trend with p = 1 and a stationary process with long memory. Based on that we consider the following model:

$$X_t = T_t + Z_t = a + \beta t + Z_t \quad for \;\; t=0,\ldots,N-1 \tag{3}$$

where the process Z_t is a residual term which is a long-memory process defined by

$$(1-B)^{\delta} Z_t = \varepsilon_t \quad , \tag{4}$$

where $0 \le \delta < 1/2$ is the long-memory parameter, $\{\varepsilon_t\}$ is a Gaussian white noise process with mean zero and $\sigma_\varepsilon^2 > 0$.

From (1), we can write $$\mathbf{d} = \mathbf{d}_W + \mathbf{d}_s \tag{5}$$

where $\mathbf{d}_w$ is an N×1 vector containing the wavelet coefficients and zeros at all other positions, and $\mathbf{d}_s$ is an N×1 vector containing the scaling coefficients and zeros at all other positions. Since $\mathbf{X} = \mathbf{W}'\mathbf{d}$, we can write,

$$\mathbf{X} = \mathbf{W}'\mathbf{d} = \mathbf{W}'\mathbf{d}_S + \mathbf{W}'\mathbf{d}_w = \hat{\mathbf{T}} + \hat{\mathbf{Z}}, \tag{6}$$

where $\hat{\mathbf{T}}$ is an estimator of the polynomial trend $\mathbf{T}$ at level J, while $\hat{\mathbf{Z}}$ is a tapered 'version' of $\mathbf{X}$. An important issue is how to choose the wavelet filter. The Haar wavelet, which is a piecewise constant function, preserves the discontinuities, and therefore it is most suitable to identify a structural break in the data.

4.3. A Wavelet-Based Test for Testing the Trend

A test statistics to test the null hypothesis H_0: $T_t = 0$ against the alternative H_1: $T_t \neq 0$ (Almasri *et al.*, 2010) is defined as follows:

$$G = \frac{\hat{\sigma}_{Sj}^2}{\sum_{j=1}^{J} \hat{v}_X^2(\lambda_j)} \tag{7}$$

$$\hat{\sigma}_X^2 \equiv \frac{1}{N}\sum_{t=0}^{N-1}\left(X_t - \bar{X}\right)^2 = \frac{1}{N}\|\mathbf{D}\|^2 - \bar{X}^2 = \frac{1}{N}\sum_{j=1}^{J}\|\mathbf{D}_j\| + \frac{1}{N}\|\mathbf{S}_j\|^2 - \bar{X}^2 \;\; = \sum_{j=1}^{J}\hat{v}_X^2(\lambda_j) + \hat{\sigma}_{Sj}^2$$

$\hat{v}_X^2(\lambda_j)$ is the estimated variance of the wavelet coefficients at scale λ_j, and $\hat{\sigma}_{Sj}^2$ is the estimated variance of the trend. The test statistic $(N - N/2^J)/(N/2^J - 1)\boldsymbol{G}$ will follow an F-distribution with

$(N/2^J - 1)$ and $(N - N/2^J)$ degrees of freedom (under the normality assumption of the scaling coefficients). The null hypothesis is rejected if the calculated value is greater than the tabulated F-value.

5. Empirical Illustration

5.1. Dataset

In this study monthly rainfall data corresponding to different zones of India is collected from Indian Institute of Tropical Meteorology (www.tropmet.res.in), Pune, India for detection as well as estimation of trend in monsoon rainfall data of different sub-divisions of India. The data set comprises of monthly rainfall over 128 years (from 1887 to 2014) measured in mm. There are total 30 sub-divisions in India, out of which 10 zones are considered in this present study. These 10 zones are selected in such a manner that it can represent the whole India. Here we are considering only four monsoon months (namely June, July, August, and September). The names of the selected sub-divisions along with their short names are listed below:

1. Assam & Meghalaya sub-divisions (ASMEG)
2. Gangetic West Bengal sub-divisions (GNWBL)
3. West Uttar Pradesh Plains sub-divisions (WUPPL)
4. Punjab sub-divisions (PUNJB)
5. East Rajasthan sub-divisions (ERJST)
6. East Madhya Pradesh sub-divisions (EMPRA)
7. Gujarat Sub-divisions (GUJRT)
8. Madhya Maharashtra sub-divisions (MADMH)
9. Coastal Andhra Pradesh sub-divisions (COAPR)
10. Tamilnadu and Pondichery sub-divisions (TLNAD)

5.2 Descriptive statistics:

The descriptive statistics of the rainfall data is reported in table 1. Here we consider mean, maximum value, minimum value, standard deviation (SD), coefficient of variation (CV), skewness and kurtosis to represent the data set. The values of these statistics are calculated for each of the monsoon months corresponding to each of the sub-divisions under consideration. A perusal of

table 1 indicates that the average monthly monsoon rainfall is higher for Assam & Meghalaya (ASMEG) and Gangetic West Bengal (GNWBL) sub-divisions than the other sub-divisions. The table also indicates that the maximum rainfall is higher for Assam & Meghalaya (ASMEG), Gangetic West Bengal (GNWBL), Gujarat (GUJRT) and Madhya Maharashtra (MADMH) sub-divisionss for each of the monsoon months. In terms of CV the variability in the monthly monsoon rainfall data is higher for West Uttar Pradesh Plains (WUPPL), Punjab (PUNJB) and Gujarat (GUJRT) sub-divisions. Almost all the series under consideration are positively skewed (except July month of WUPPL and MADMH) and all the series are leptokurtic.

Table 1: Descriptive Statistics

Zone	Month	Mean	SD	Maximum	Minimum	CV	Skewness	Kurtosis
ASMEG	June	4405	1060.08	7760	2333	24.065	0.380	3.217
	July	4158	1029.17	7812	2398	24.751	0.844	3.625
	Aug	3508	823.553	6672	1952	23.476	0.664	4.104
	Sept	2867	785.427	5325	1308	27.395	0.641	3.291
GNWBL	June	2502	980.389	5835	699	39.184	0.731	3.477
	July	3316	900.475	6736	1590	27.156	0.647	3.837
	Aug	3236	865.187	5632	1721	26.736	0.713	2.894
	Sept	2522	939.822	5967	924	37.258	1.308	4.813
WUPPL	June	890	646.674	3007	39	72.660	1.116	3.889
	July	2589	876.867	4320	201	33.869	-0.263	2.585
	Aug	2568	937.033	4948	650	36.489	0.149	2.663
	Sept	1497	905.677	3732	41	60.499	0.451	2.355
PUNJB	June	547	486.415	3360	13	88.875	2.597	13.572
	July	1774	800.455	5390	221	45.122	1.078	6.016
	Aug	1669	788.405	3847	237	47.238	0.637	3.134
	Sept	953	968.369	4412	0	101.55	1.749	5.671
GUJRT	June	1279	941.213	5682	8	73.590	1.322	6.302
	July	3433	1567.91	8625	126	45.672	0.59	3.27
	Aug	2441	1311.06	6385	101	53.710	0.533	2.738
	Sept	1435	1184.49	5800	25	82.560	0.979	3.532
EMPRA	June	1559	857.622	3863	262	55.025	0.667	2.790
	July	3842	1082.27	6716	869	28.169	0.113	2.615
	Aug	3715	1041.44	7424	1779	28.033	0.607	3.776
	Sept	1977	999.278	5627	177	50.545	0.856	4.122
MADMH	June	1261	515.891	2642	233	40.905	0.457	2.995

	July	1815	614.799	3201	409	33.873	-0.025	2.507
	Aug	1315	515.319	3198	261	39.191	0.296	3.289
	Sept	1403	621.782	2977	212	44.312	0.208	2.396
COAPR	June	893	433.041	2528	252	48.509	1.222	4.714
	July	1334	459.005	3155	477	34.408	0.764	4.169
	Aug	1343	468.634	2920	285	34.895	0.486	3.317
	Sept	1556	623.493	3910	587	40.070	0.934	4.119
TLNAD	June	464	222.054	1838	141	47.836	2.289	13.644
	July	622	317.790	1873	162	51.116	1.500	5.357
	Aug	903	402.817	2491	282	44.599	0.885	3.977
	Sept	1100	412.480	2173	267	37.498	0.311	2.572
ERJST	June	743	507.121	2275	30	68.299	0.847	3.063
	July	2257	837.666	4287	213	37.114	0.154	2.814
	Aug	2223	978.826	4652	58	44.032	0.233	2.532
	Sept	1026	747.785	3442	25	72.855	0.939	3.723

*SD: Standard Deviation and CV: Coefficient of Variation

5.3. Test for long memory

The presence of long memory in a time series data can be confirmed either by investigating the autocorrelation function (ACF) plot of the data set or by using some parametric and nonparametric tests like GPH (Geweke and Porter-Hudak, 1983) and Sperio test (Reisen, 1994). In this study we have applied Sperio test to test the presence of long memory for each of the monsoon month rainfall data of the selected zones and the test results are provided in table 2. It is found that the test is significant at 5% level for some of the series which establishes the presence of long range dependency in the respective monthly rainfall data like July(ASMEG), August(WUPPL) etc.

Table 2: Test for long memory

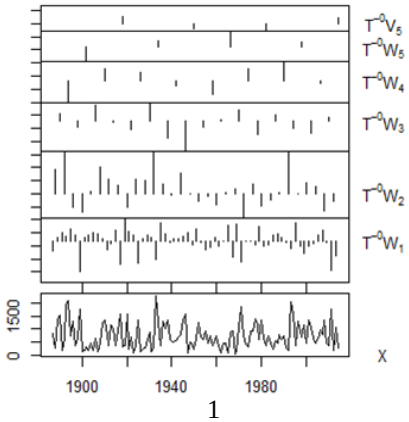
Zone	Month	d	Zone	Month	d
ASMEG	June	0.233	EMPRA	June	-0.189
	July	0.266*		July	0.429*
	Aug	0.101		Aug	0.239*
	Sept	0.211		Sept	-0.339*
GNWBL	June	0.142	MADMH	June	0.132
	July	-0.111		July	0.206
	Aug	-0.153		Aug	0.000

	Sept	0.172		Sept	-0.107
WUPPL	June	0.089	COAPR	June	-0.622*
	July	0.000		July	0.129
	Aug	0.449*		Aug	-0.039
	Sept	-0.170		Sept	-0.242*
PUNJB	June	0.112	TLNAD	June	-0.181
	July	0.421*		July	0.170
	Aug	0.010		Aug	-0.089
	Sept	0.039		Sept	0.111
GUJRT	June	-0.160	ERJST	June	-0.079
	July	-0.149		July	0.087
	Aug	-0.439*		Aug	-0.174
	Sept	0.133		Sept	-0.031

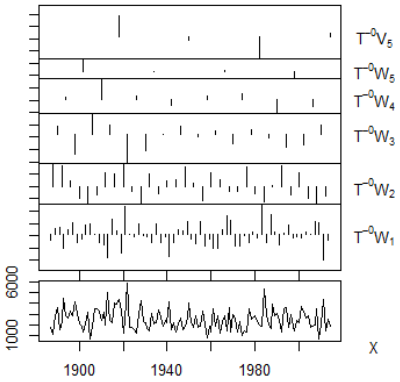
*denotes significance at 5% level.

5.4. Variation in rainfall using DWT

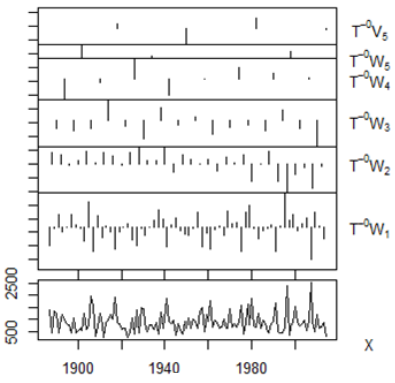
DWT plots of rainfall in different zones for the months of June, July, August and September using Haar wavelet filter at level 5 are reported in fig.1. The number of the plots in fig. 1 are given for the different zones accordingly to the order as per section 5.1. Wavelet coefficients are plotted as bars, up or down. The sizes of the bars are relative to magnitudes of coefficients. The number of wavelet coefficients at the lowest resolution level (level = 1) is exactly half the number of original data points and the number of coefficients decreases by half at each level (Nason and Sachs, 1999). The plot of actual data is shown at the bottom of each plot of fig 1 which is very rough and it is full of noise component. DWT attempts to extract the actual signal from the noisy time series data at each level. The graph of W_2 is much smoother than the W_1. Similarly smoothness increases as we are going to top of the graphs with upper coefficients. In the graph V_5 scaling coefficient shows the smooth plot.


$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
0 1500
X
1900 1940 1980

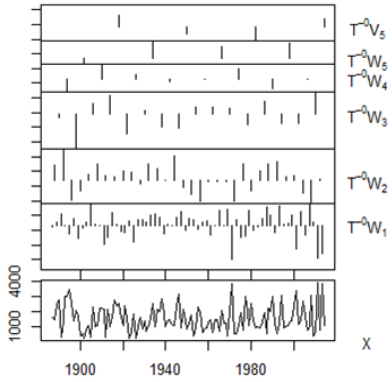
1


$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
1000 6000
X
1900 1940 1980

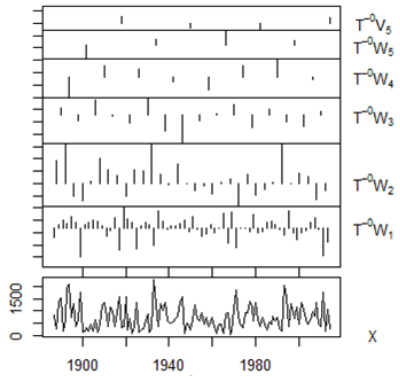
2


$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
500 2500
X
1900 1940 1980

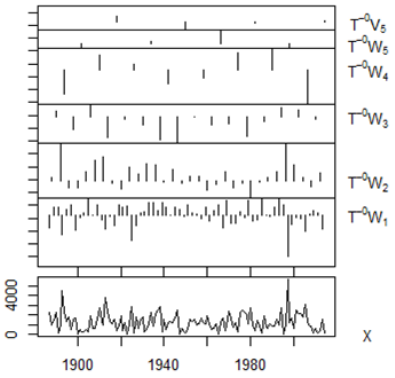
3


$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
1000 4000
X
1900 1940 1980

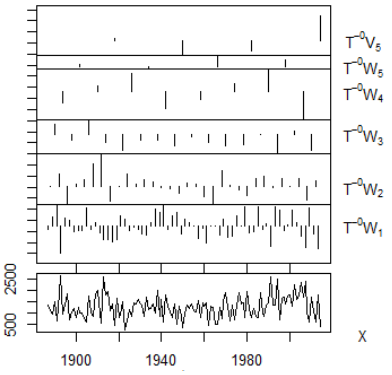
4


$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
0 1500
X
1900 1940 1980

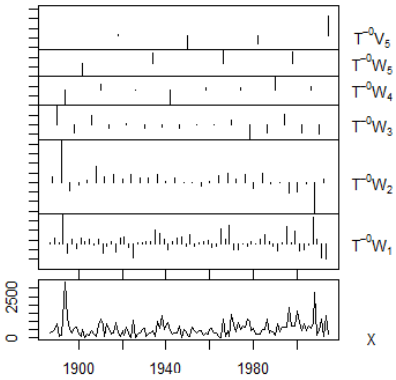
5


$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
0 4000
X
1900 1940 1980

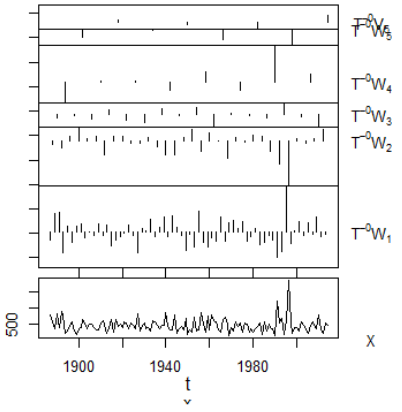
6

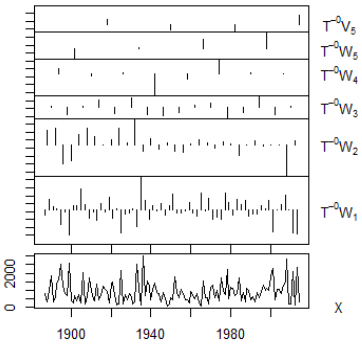

$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
500 2500
X
1900 1940 1980

7


$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
0 2500
X
1900 1940 1980

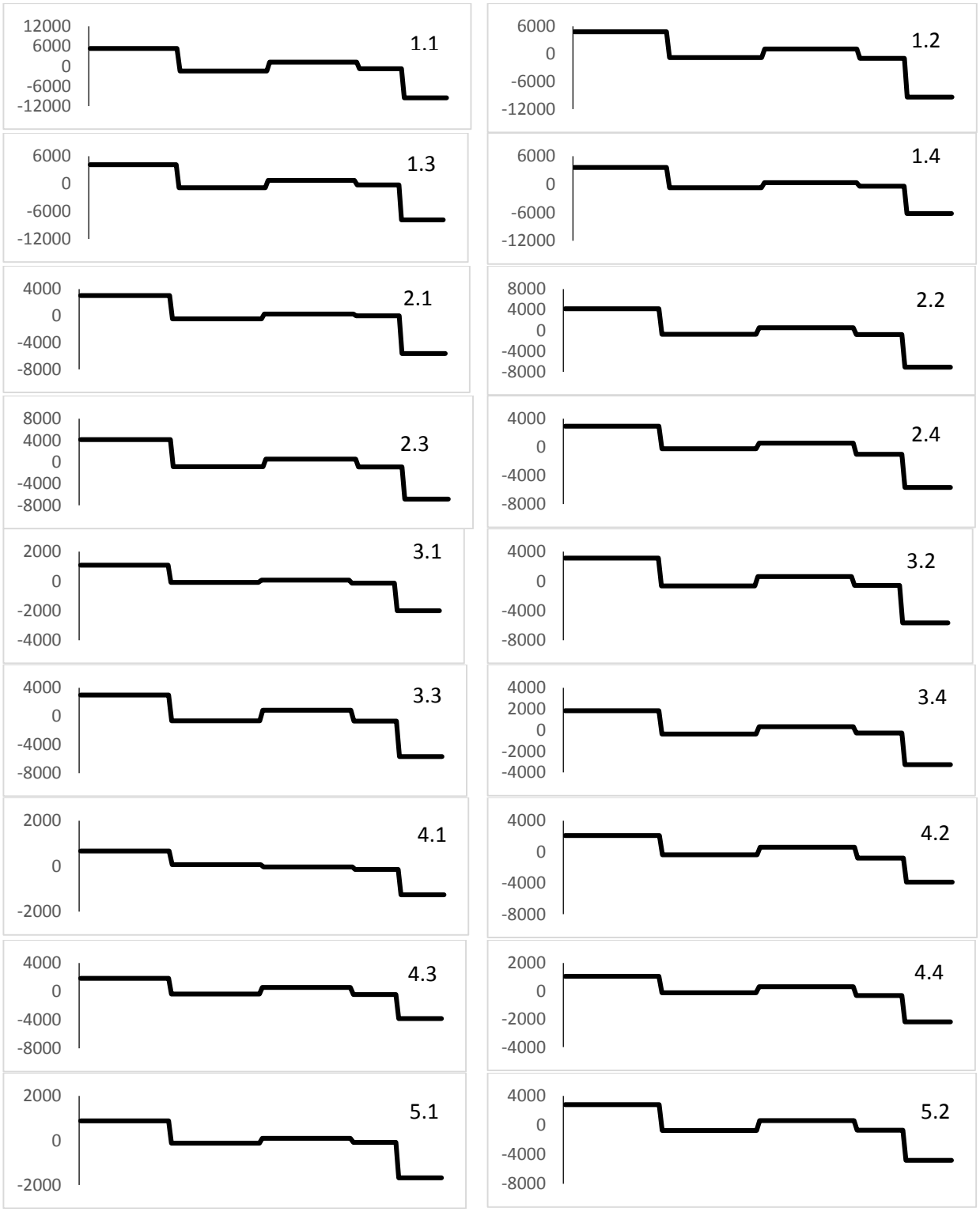
8

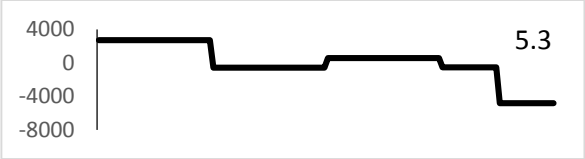

$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
500
X
1900 1940 1980
t

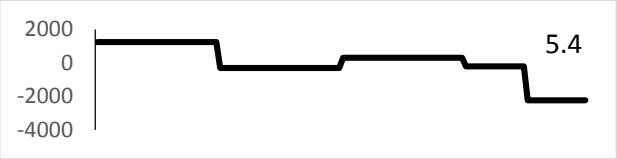

$T^{-0}V_5$
$T^{-0}W_5$
$T^{-0}W_4$
$T^{-0}W_3$
$T^{-0}W_2$
$T^{-0}W_1$
0 2000
X
1900 1940 1980

10

Fig. 1. DWT plot of different zones using Haar wavelet filter

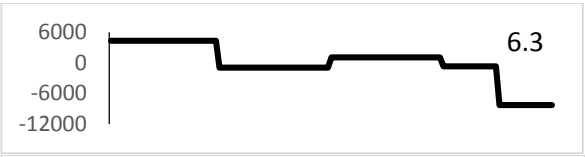



4000
0
-4000
-8000
5.3


2000
0
-2000
-4000
5.4

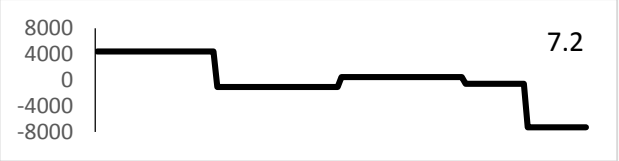

4000
0
-4000
6.1


6000
0
-6000
-12000
6.2

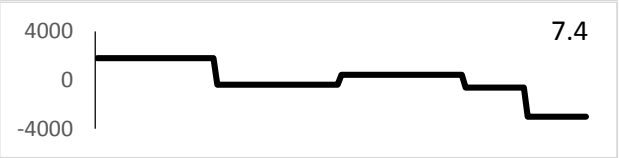

6000
0
-6000
-12000
6.3


4000
0
-4000
-8000
6.4


2000
0
-2000
-4000
7.1


8000
4000
0
-4000
-8000
7.2

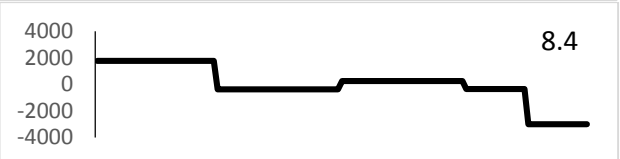

4000
0
-4000
-8000
7.3

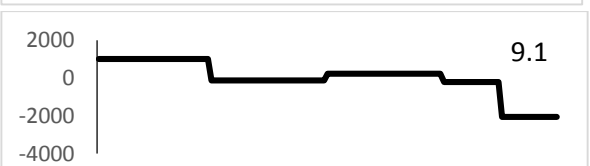

4000
0
-4000
7.4


2000
0
-2000
-4000
8.1


4000
0
-4000
-8000
8.2


2000
0
-2000
-4000
8.3


4000
2000
0
-2000
-4000
8.4


2000
0
-2000
-4000
9.1

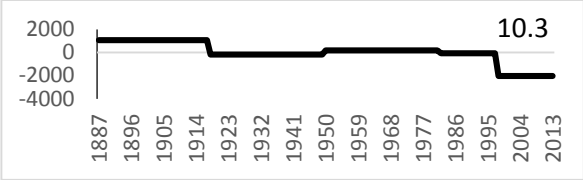

2000
0
-2000
-4000
9.2


2000
0
-2000
-4000
9.3


4000
2000
0
-2000
-4000
9.4


2000
0
-2000
10.1


2000
0
-2000
10.2



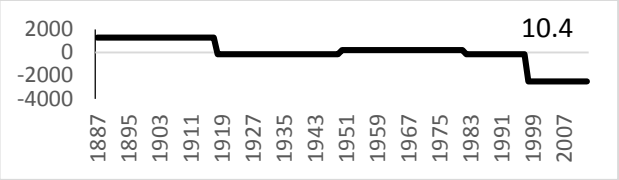


Fig. 2.Trend plots of different zones

5.5. Detection of trend using Wavelet

Trends of different sub-divisional zones are estimated using "Haar" DWT and plotted in the Fig-2. Each plot in figure 2 is represented by two subscripts. The first subscript refers to zone as described in section 5.1 and the second subscript refers to the monsoon months i.e. 1 for June; 2 for July; 3 for August and 4 for September. Therefore, the first plot i.e. 1.1 represents the trend in rainfall in ASMEG zone for the month June. The rainfall trend in the month of June for ASMEG zone clearly indicates that trend is negative. From the year 1887 to 1915 trend remains constant then it falls during 1916. Again it remains stagnant during the year 1918 to 1949. After that a slight increase in the rainfall has been detected till the year 1951. A rapid decreasing trend has been observed from 1997 to 2014. For other three months of this zone, the rainfall pattern remains same with a difference in magnitude. The trend in other zones can be interpreted in similar pattern. The overall pattern of trend in rainfall remains almost same in all the zones except the variation in different time epochs.

5.6. Test for detection of trend

In this section we test the existence of trend using Mann-Kendall test. The values of test statistics along with corresponding p-values are given in table 3. This table also includes the values of G statistic along with their corresponding F-values for testing the significance of existing trend as described in section 4.3. It has been seen that the Mann-Kendal test could not capture the significant trend in many of the series whereas the Wavelet based test successfully captured the trend in almost all the months of different zones. This is because, wavelet based test has more power than that of the Mann-Kendall test in capturing the trend particularly when there is long memory present in the series.

Table 3: Test for detection of trend

Zone	Months	Method				
		Haar Filter		Mann-Kendall		
		G	F	S	tau	p-value
ASMEG	June	0.171	7.068*	-1121	-0.138*	0.021
	July	0.046	1.901	-229	-0.028	0.630
	Aug	0.043	1.777	-1316	-0.162*	0.006
	Sept	0.034	1.405	-520	-0.064	0.285
GNWBL	June	0.494	20.419*	-505	-0.062	0.299
	July	0.232	9.589*	8	0.001	0.989
	Aug	0.474	19.592*	569	0.070	0.242
	Sept	0.615	25.420*	1133	0.139*	0.019
WUPPL	June	0.127	5.249*	338	0.041	0.487
	July	0.08	3.307*	-235	-0.028	0.629
	Aug	0.345	14.260*	-407	-0.050	0.403
	Sept	0.018	0.744	-167	-0.020	0.732
PUNJB	June	0.442	18.269*	1168	0.144*	0.016
	July	0.51	21.080*	270	0.033	0.579
	Aug	0.277	11.449*	-28	-0.003	0.950
	Sept	0.102	4.216*	612	0.075	0.208
GUJRT	June	0.037	1.529	-111	-0.013	0.820
	July	0.272	11.243*	-190	-0.023	0.690
	Aug	0.1	4.133*	585	0.072	0.220
	Sept	0.134	5.539*	372	0.045	0.440
EMPRA	June	0.199	8.225*	-305	-0.030	0.530
	July	1.15	47.533*	-879	-0.108	0.070
	Aug	0.272	11.243*	-612	-0.075	0.208
	Sept	0.025	1.033	-517	-0.060	0.287
MADMH	June	0.475	19.633*	722	0.088	0.137
	July	0.17	7.027*	-568	-0.069	0.242
	Aug	0.69	28.520*	1145	0.141*	0.010
	Sept	0.112	4.629*	48	0.005	0.922
COAPR	June	0.195	8.060*	-32	-0.003	0.949
	July	0.259	10.705*	446	0.054	0.359
	Aug	0.441	18.228*	858	0.106	0.077
	Sept	0.021	0.868	-151	-0.018	0.757
TLNAD	June	0.054	2.232	-468	-0.057	0.336
	July	0.26	10.747*	403	0.049	0.407
	Aug	0.103	4.257*	29	0.003	0.954
	Sept	0.152	6.283*	-141	-0.017	0.773

ERJST	June	0.056	2.315	322	0.039	0.508
	July	0.566	23.395*	-34	-0.004	0.940
	Aug	0.063	2.604	94	0.011	0.840
	Sept	0.075	3.100*	178	0.021	0.710

*denotes significance at 5% level.

6. Conclusion

In the present study, tow non parametric methods namely wavelet analysis and Mann-Kendall test have been used for estimation of trend in monthly rainfall in ten selected sub-divisions of India. Presence of long memory in rainfall data was tested using suitable statistical test and it was found that long memory was significantly present in rainfall for the month of June in COAPR; July of ASMEG, PUNJB and EMPRA; August of GUJRT and EMPRA; September of EMPRA and COAPR. In presence of long memory, Mann-Kendall test may not be appropriate to detect the trend. Therefore, wavelet based trend test was advocated and it was found that in most of the sub-divisions, the rainfall trend was significant at 5% level. Mostly, significant negative trend in rainfall has been found in the monsoon months which is not a positive signal for the water sector in the country. The variation in monthly rainfall for June-September for the studied zones were also investigated by wavelet using Haar filter. The local as well as global variation in rainfall is observed by DWT plot of the respective months in the respective locations. The variability in rainfall is evident in the recent decades in all the locations.

Chapter V

Wavelets Based Combination Approach for Modeling Sub-Divisional Rainfall In India

1. Introduction

Modeling climatic variables in general and rainfall in particular is intended to be useful for many stakeholders namely the farmers, policy makers and governments. Modeling of time series helps in extracting many inherent features present in that data set and based on the extracted features it helps in extrapolating the time series into future. In the last few decades, much effort has been devoted to the development and improvement of time series forecasting models and there are infinitely many stochastic processes which can be implemented to model and forecast for a particular series. The quantum and distribution of rainfall mostly determines the performance of agriculture of a country. Therefore, from proper planning and policy making point of view, accurate modeling of rainfall is vital. In literature, several attempts have been found regarding developments of models for describing climatic variables like rainfall and temperature. In Indian context, Parthasarathy *et al*. (1995) reported some statistical details and long-term changes of the All-India monsoon rainfall. Rajeevan *et al*. (2004) have reported an excellent review of different statistical models employed since 1988 along with various modifications made in these models from time to time, particularly in the identification of relevant explanatory variables for modeling and forecasting of climatic variables. Kakatkar *et al*. (2017) investigated Indian summer monsoon rainfall variability.

In statistical and stochastic models, Box-Jenkins ARIMA methodology has virtually dominated analysis of time series data for last several decades. This model has gained much popularity in modeling linear dynamics. However, this models is based on the assumptions of stationarity. Moreover, many a times, presence of high chaotic nature and complex nonlinearity of series under consideration distorts the actual model specification and thereby in the forecast performance. Therefore, preprocessing of input data become essential to minimize the noise level

and improve the performance of forecasting model. An example of such a powerful method is nonparametric wavelet analysis. Wavelet analysis decomposes the original time series into both high and low frequency components, in order to improve the ability of a forecasting model by capturing useful information at different resolution levels (Antoniadis, 1997; Percival and Walden, 2000; Vidakovic, 1999). Quite good number of literatures is available towards theoretical development of wavelet functions and their application in economics and finance. However, the application of wavelet to real data analyses is scarce particularly in climatic series. Fryzlewicz *et al*. (2003) developed Wavelet process model for forecasting nonstationary time-series. Renaud *et al*. (2003) developed methodology for prediction of time-series data based on multiscale decomposition. Wavelet thresholding approach was applied by Sunilkumar and Prajneshu (2004) for modeling and forecasting of monthly meteorological subdivisions rainfall in Eastern Uttar Pradesh, India. Almasri *et al*. (2008) proposed a test statistic by using wavelet decompositions for testing presence of significant trend in a time-series data. Aminghafari and Poggi (2007, 2012) used Wavelets and kernel smoothing approach for forecasting nonstationary time-series. Ghosh *et al*. (2010) have computed the size and power of test based on wavelet for testing presence of trend in time series. Kisi (2010, 2011) studied wavelet regression model and compared the forecasting efficiency with that of neural network model. Paul *et al*. (2011) applied Wavelet methodology for detection of trend in Indian monsoon rainfall and found that there is a significant declining trend. Venkata Ramana *et al*. (2013) attempted to find an alternative method for rainfall prediction by combining the wavelet technique with ANN. Paul and Birthal (2015) have used wavelet method for estimation of trend in rainfall in different agro-climatic zones of India. Paul *et al*. (2013) studied wavelet frequency domain approach for forecasting of Indian monsoon rainfall. Paul (2017) studied modeling of maximum and minimum temperature in India by using wavelet methods. Paul and Anjoy (2018) used wavelet and long memory model for modeling maximum temperature in India. Meena *et al*. (2019) studied homogeneity of monthly, seasonal, and annual rainfall over arid region of Rajasthan, India. Paul *et al*. (2019) have applied wavelet technique for estimation and testing of trend in ten selected sub-divisional rainfall in India. There is a need to study the predictive performance of different combination models based on wavelet decomposition. In the present investigation, an attempt has been made for modeling and forecasting of annual rainfall for 30 sub-divisions of India based on combination of

decomposition approach e.g. wavelet, the stochastic model e.g. ARIMA and machine learning technique like ANN and evaluate the efficiency of forecast in comparison to ARIMA model.

2. Dataset

In this study annual rainfall data corresponding to different zones of India is collected from Indian Institute of Tropical Meteorology (www.tropmet.res.in), Pune, India. The data set comprises of annual rainfall of 146 years (from 1871 to 2016) measured in mm. For all the 30 sub-divisions of India, rainfall modeling is carried out. The names of the selected sub-divisions along with their short names are listed in table 1.

Table 1. The sub-divisions considered for present study

Abbreviation	**Name of Sub-divisions**
ASMEG	Assam & Meghalaya Sub-division
NMAMT	Nagaland, Manipur, Mizoram and Tripura Sub-division
SHWBL	Sub Himalayan West Bengal Sub-division
GNWBL	Gangetic West Bengal Sub-division
ORISS	Orissa Sub-division
JHKND	Jharkhand Sub-division
BIHAR	Bihar Sub-division
EUPRA	East Uttar Pradesh Sub-division
WUPPL	West Uttar Pradesh Sub-division
HARYA	Haryana Sub-division
PUNJB	Punjab Sub-division
WRJST	West Rajasthan Sub-division
ERJST	East Rajasthan Sub-division
WMPRA	West Madhya Pradesh Sub-division
EMPRA	East Madhya Pradesh Sub-division
GUJRT	Gujarat Sub-division
SAUKU	Saurashtra & Kutch Sub-division
KNGOA	Konkan and Goa Sub-division

MADMH	Madhya Maharashtra Sub-division
MARAT	Marathwada Sub-division
VDABH	Vidarbha Sub-division
CHHAT	Chattisgarh Sub-division
COAPR	Coastal Andhra Pradesh Sub-division
TELNG	Telangana Sub-division
RLSMA	Rayalaseema Sub-division
TLNAD	Tamil Nadu Sub-division
COKNT	Coastal Karnataka Sub-division
NIKNT	North interior Karnataka Sub-division
SIKNT	South interior Karnataka Sub-division
KERLA	Kerala Sub-division

3. Methodology:

3.1.Autoregressive Integrated Moving Average (ARIMA) Model

ARIMA model is a generalization of ARMA models which incorporate a wide range of non-stationary time-series by suitable order of differencing. The simplest example of a non-stationary process which reduces to a stationary one after differencing is 'Random Walk'. A process $\{y_t\}$ is said to follow an Integrated ARMA model, denoted by ARIMA (p, d, q), if $\nabla^d y_t = (1-B)^d \varepsilon_t$ is ARMA (p, q). The model is written as:

$$\varphi(B)(1-B)^d y_t = \theta(B)\varepsilon_t \qquad (1)$$

where, $\varepsilon_t \sim WN(0,\sigma^2)$, WN indicates white Noise, $\varphi(B) = 1-\varphi_1 B-\varphi_2 B^2 -...-\varphi_p B^p$ and $\theta(B) = 1-\theta_1 B-\theta_2 B^2 -...-\theta_q B^q$. The integration parameter d is a non-negative integer.

There are four major stages of ARIMA model building. These are identification, estimation, validation and forecasting. A detailed discussion on various aspects of this approach is given in Box *et al*. (2007).

3.2. Wavelets

Wavelets are fundamental building block functions, analogous to the trigonometric sine and cosine functions. A wavelet function oscillates about zero. If $\psi(.)$ is a real-valued function defined over the real axis $(-\infty,\infty)$ and satisfies two basic properties: (i) Integral of $\psi(.)$ is zero, i.e. $\int_{-\infty}^{\infty}\psi(u)\,du=0$ and (ii) Square of $\psi(.)$ integrates to unity, i.e. $\int_{-\infty}^{\infty}\psi^2(u)\,du=1$, then the function $\psi(.)$ is called a wave. A good description of wavelets can be found in Daubechies (1992), Ogden (1997) and Percival and Walden (2000).

There are mainly two types of wavelet transforms: the continuous wavelet transform (CWT) designed to work with time-series defined over the entire real axis and the discrete wavelet transform (DWT) which deals with series defined essentially over discrete time points. In order to capture high and low frequency components in a time series, DWT is applied. Again, through computation of inverse DWT, this, in turn would enable modeling of time-series data.

3.3. Maximal Overlap Discrete Wavelet Transforms (MODWT)

The MODWT is an undecimated wavelet transform over dyadic scales. Similar to DWT, the Maximal Overlap Discrete Wavelet Transforms (MODWT) also produces a set of time-dependent wavelets (details) and scaling (smooth) coefficients based on linear filtering operation. Both the transformations use basis vectors associated with a location parameter *t* and a scale parameter $\tau_j = 2^j - 1$ for each level of decomposition i.e. j =1,..,J_0. However, MODWT differs from DWT in the sense that it is a highly redundant, nonorthogonal transform (Percival and Walden, 2000). The main drawback associated with the application of the DWT in time series analysis is that it suffers from lack of translation invariance. Which implies that, circularly shifting the time series will not necessarily shift its DWT coefficients in a similar manner. The downsampled values at each level of decomposition are discarded by DWT but retained in MODWT. The MODWT is applicable for all the time series with any integer length of the series, whereas for a complete decomposition of levels J, DWT requires the length of the series to be a multiple of 2^J. Redundancy of MODWT facilitates alignment of the decomposed wavelet and scaling coefficients at each level with original time-series, thus enabling a ready comparison between the series and its decomposition. ANOVA derived using MODWT is not influenced by

circular shifting of input time-series, whereas values derived using DWT depend on starting point of a series (Percival and Mofjeld, 1997; Percival and Walden, 2000).

3.4 Artificial neural network (ANN)

Artificial neural networks (ANNs) are nonlinear data driven self-adaptive approach and are powerful tools for modeling, especially when the underlying data relationship is unknown. A very important feature of these networks is their adaptive nature, where "learning by example" replaces "programming" in solving problems. Most commonly used ANN is multi-layer perceptron (MLP), a class of feed forward neural network. MLP consists of three types of nodes viz. input nodes, hidden nodes and output nodes. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised training scheme, which means, it learns from labeled training data. In reality, most of the time series data is not linearly separable by using linear perceptron therefore MLP is used which has multiple layers and non-linear activation function.

Mathematically, MLP network is a function consisting of compositions of weighted sums of the functions corresponding to the neurons. Let us consider following notations with p input and h hidden nodes: x_i(i=1,2…p) are network inputs; w_{ij},refer the synaptic weight connection between neuron i and j; w_j refer the synaptic weight connection between j^{th} neuron of hidden node and output node.a_0, β_{0J} are bias term for output layer and hidden layer; φ is hidden output layer activation function, mainly logistic $\varphi(v_j) = \frac{1}{1+e^{-v_j}}$ and I as identity function. The output of a MLP with p input and h hidden nodes is expressed as

$$Y = I\left[a_0 + \sum_{j=1}^{h} w_j \varphi_j\left[\beta_{0J} + \sum_{i=1}^{p} w_{ij} x_i\right]\right] = I\left[a_0 + \sum_{j=1}^{h} w_j [v_j]\right] = I[y_j]$$

where

$$v_j = \varphi_j\left[\beta_{0J} + \sum_{i=1}^{p} w_{ij} x_i\right],\ y_j = \left[a_0 + \sum_{j=1}^{h} w_j [v_j]\right]$$

3.5 Proposed algorithm for modeling using combination of Wavelet with ARIMA and ANN:

Each of the sub-divisional rainfall series is decomposed by means of MODWT using Haar wavelet filter. The level of decomposition is taken to be $J_0 \leq log_2(N)$, where N is the length of the series. In the present investigation J_0 comes out to be 7. One advantage of this wavelet methodology is that one can capture the multi-scale information at different resolution level. So model precision will be more as chances of losing information is less. This is termed as denoising, because unexplained error part or information loss is checked through decomposition at different frequency level. In the present investigation, MODWT has been considered for modeling purpose as there is no restriction for sample size determination. Furthermore, redundancy of MODWT wavelet coefficients increases effective degrees of freedom on each scale and thus decreases variance of certain wavelet-based statistical estimates.

The algorithm used is as follows:

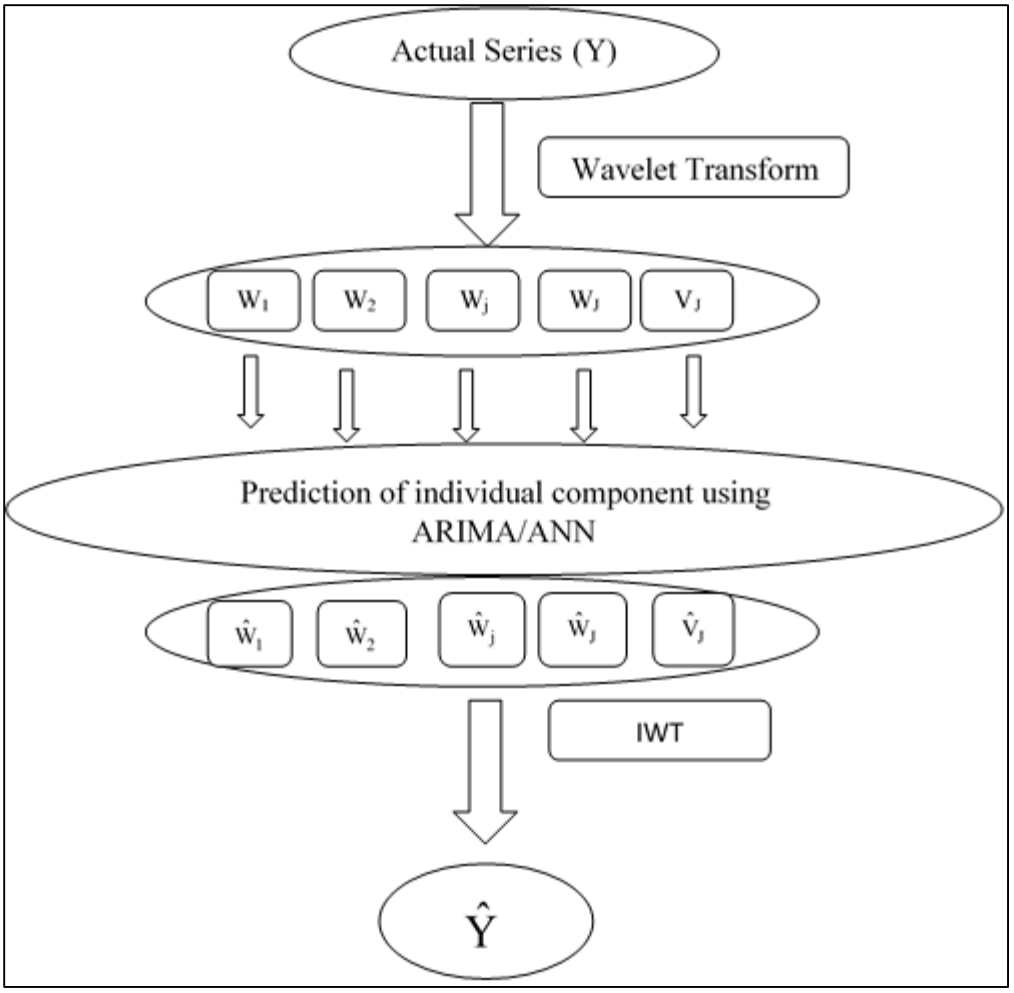
(i) Original time-series is decomposed into a certain number of sub time-series {W_1, W_2, …,W_J,V_J} by non-decimated wavelet transform (MODWT). W_1, W_2 ,…,W_J are wavelet detail component, and V_J is smooth component. These play different role in the original time-series and the behaviour of each sub time-series is distinct from other. Therefore, the contribution to original time-series varies from each other.

(ii) For Wavelet-ARIMA (Wavelet-ANN) Suitable ARIMA (ANN) model is fitted on each of the decomposed sub-series and prediction is computed. The lags of decomposed series are the input for ANN to get the prediction of individual decomposed series as output. The optimum value of input lag and hidden node in case of ANN is obtained based on minimum mean square error. Whereas, the suitable order of ARIMA model is selected based on minimum information criterion.

(iii) By means of inverse wavelet transform (IWT), prediction of actual rainfall series is obtained.

To implement the above algorithm, two R-packages namely WaveletArima (Paul and Samanta, 2018) and WaveletANN (Paul, 2019) have also been developed and uploaded in CRAN

The schematic diagram of the above algorithm for Wavelet-ARIMA and Wavelet-ANN is presented below.



4. Validation of models for hold-out data

One-step ahead forecasts of sub-divisional rainfall for the years 2002 to 2016 in respect of ARIMA, Wavelet-ARIMA and Wavelet-ANN are computed. A comparative study of forecasts by these models is carried out on the basis of Mean Absolute Prediction Error (MAPE) and Root Mean Square Error (RMSE) as

$$\text{MAPE} = 1/15\sum_{i=1}^{15}\left\{\left|y_{t+i}-\hat{y}_{t+i}\right|/y_{t+i}\right\}\times 100 \quad (2)$$

$$\text{RMSE} = \left[1/15\sum_{i=1}^{15}\left[y_{t+i}-\hat{y}_{t+i}\right]^2\right]^{1/2} \quad (3)$$

5. Results and Discussions

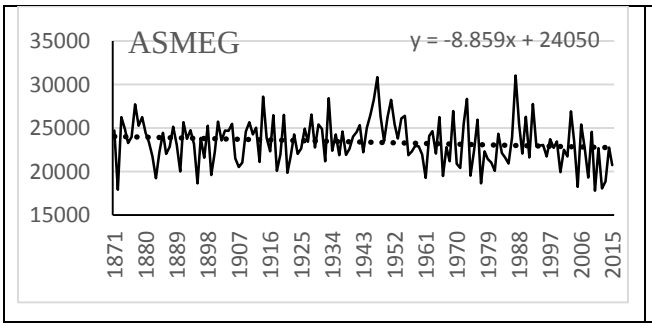
5.1 Summary statistics:

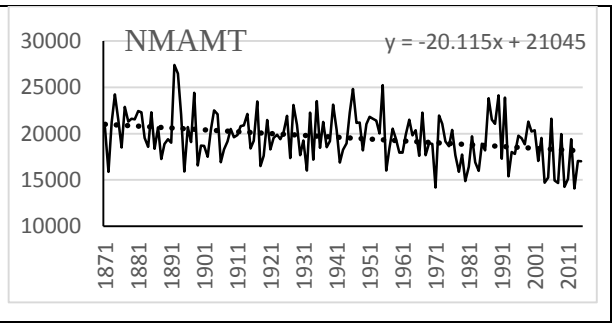
The descriptive statistics of the rainfall data is reported in table 2. Here we consider mean, median (Med), maximum value (Max), minimum value (Min), standard deviation (SD), coefficient of variation (CV), Skewness (Skew) and kurtosis (Kurt) to represent the data set. The values of these statistics are calculated for each of the sub-divisions under consideration. A perusal of table 2 indicates that the average annual rainfall is comparatively higher for Coastal Karnataka Sub-division (COKNT), Konkan and Goa Sub-division (KNGOA), Sub Himalayan West Bengal Sub-division (SHWBL), Kerala sub-division (KERLA) and Assam & Meghalaya sub-division (ASMEG). The variability in annual rainfall is highest in Saurashtra & Kutch Sub-division (SAUKU) and lowest in Assam & Meghalaya sub-division (ASMEG). Almost all the series under consideration are positively skewed and leptokurtic. The Jarque-Bera (J-B) statistic as reported in table 2 indicates that the annual rainfall series in COAPR, COKNT, PUNJB, SAUKU and WRJST subdivisions are nor-normal whereas in all other sub-divisions, the annual rainfall follows normal distribution. The pattern in annual rainfall in all the sub-divisions along with the linear trend equation as estimated by using linear regression model is reported in figure 1.

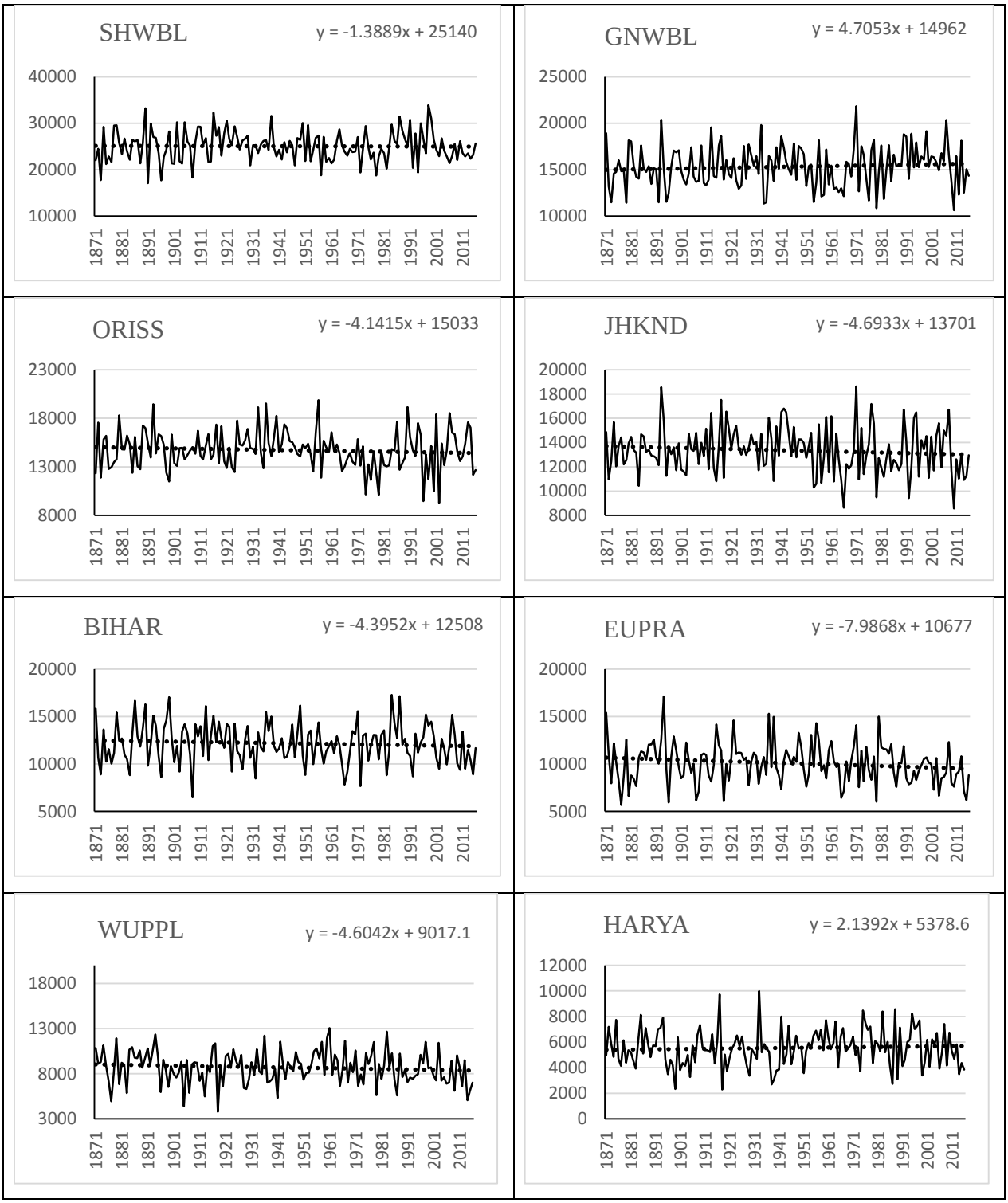
Table 2. Summary statistics of sub-divisional annual rainfall in India

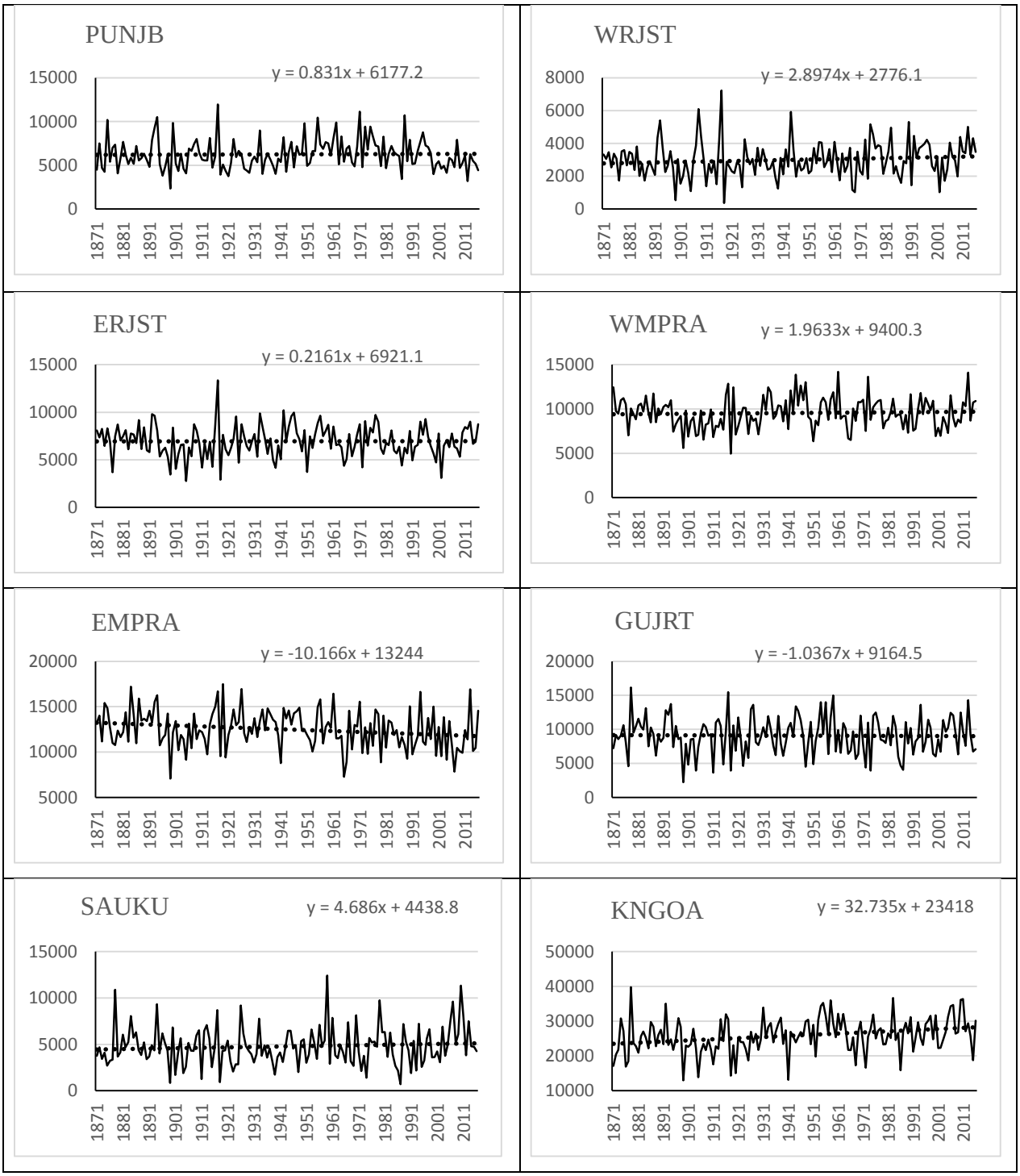
	ASMEG	BIHAR	CHHAT	COAPR	COKNT	EMPRA	ERJST	EUPRA	GNWBL	GUJRT
Mean	23398	12184	13560	9841	32924	12496	6937	10090	15307	9088
Med	23226	12029	13598	9569	32388	12423	6834	10103	15196	9033
Max	31029	17295	20817	17047	55526	17491	13368	17131	21823	16142
Min	17799	6483	8212	5318	20330	7078	2757	5693	10632	2220
SD	2577	2111	2268	1996	5497	2127	1693	2126	2219	2717
Skew	0.18	0.08	0.28	0.55	0.77	0.03	0.10	0.40	0.19	0.02

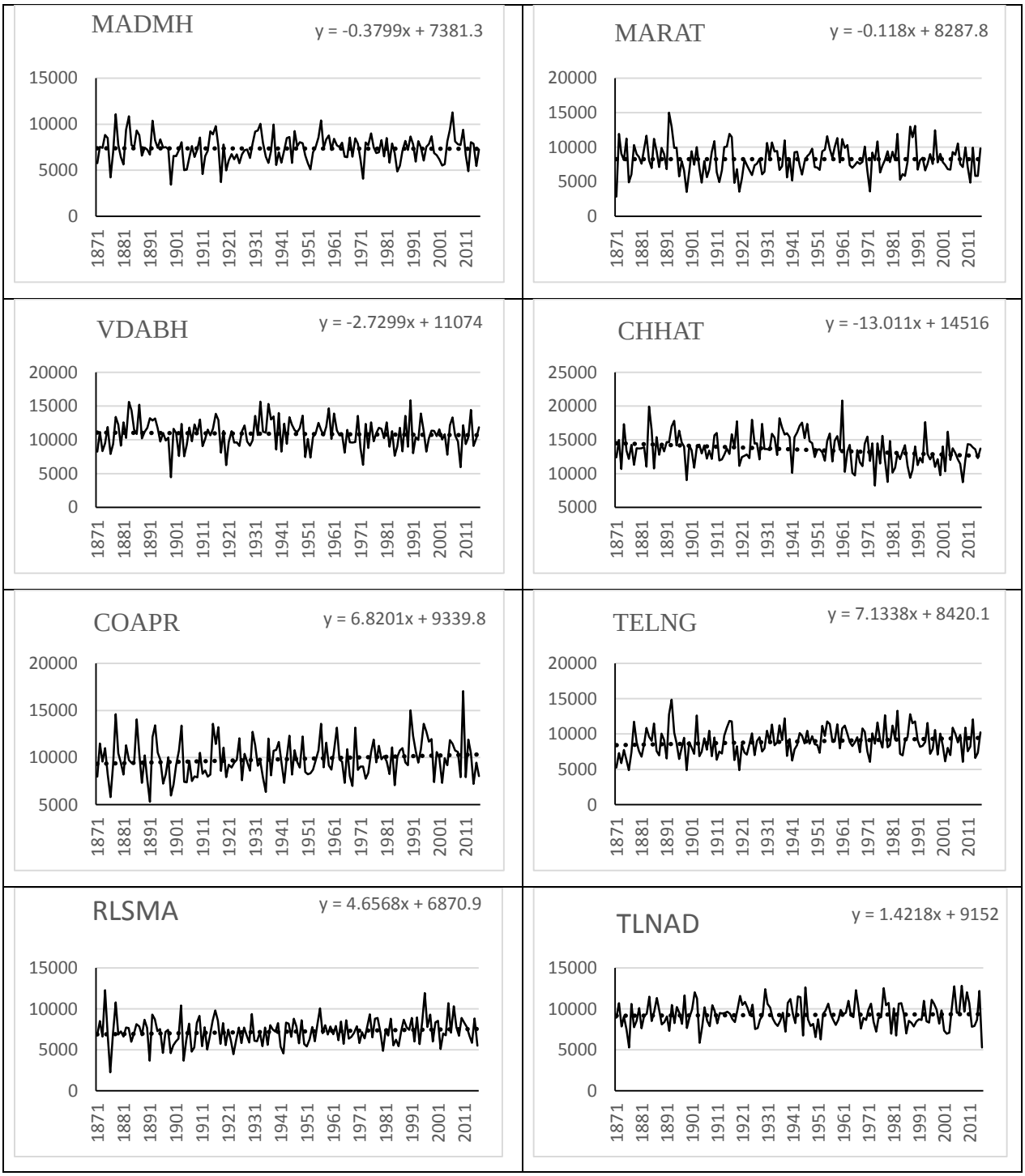
Kurt	3.06	2.79	3.27	3.47	4.72	2.67	3.66	3.40	2.69	2.72
CV (%)	11.01	17.32	16.73	20.28	16.70	17.02	24.41	21.07	14.49	29.89
J-B	0.85	0.43	2.28	8.82	32.37	0.67	2.86	4.95	1.45	0.47
Prob.	0.65	0.81	0.32	0.01	0.00	0.71	0.24	0.08	0.48	0.79
	HARYA	**JHKND**	**KERLA**	**KNGOA**	**MADMH**	**MARAT**	**NIKNT**	**NMAMT**	**ORISS**	**PUNJB**
Mean	5536	13356	28167	25824	7353	8279	8269	19565	14729	6238
Med	5470	13272	28016	25572	7452	8041	8281	19496	14678	5966
Max	9976	18628	39449	39746	11284	15009	12364	27423	19883	11955
Min	2293	8559	18374	13027	3446	2838	4129	14064	9276	2332
SD	1453	1929	4151	5010	1441	2143	1485	2585	2008	1752
Skew	0.31	0.20	0.21	-0.06	0.02	0.21	0.06	0.18	0.08	0.81
Kurt	3.07	2.89	2.86	3.27	3.31	3.07	2.94	3.01	3.28	3.56
CV (%)	26.25	14.44	14.74	19.40	19.60	25.89	17.96	13.21	13.63	28.09
J-B	2.43	1.07	1.18	0.52	0.59	1.07	0.09	0.80	0.63	17.87
Prob.	0.30	0.59	0.56	0.77	0.74	0.59	0.95	0.67	0.73	0.00
	RLSMA	**SAUKU**	**SHWBL**	**SIKNT**	**TELNG**	**TLNAD**	**VDABH**	**WMPRA**	**WRJST**	**WUPPL**
Mean	7213	4783	25038	8851	8944	9256	10874	9545	2989	8679
Med	7171	4366	25128	8891	8835	9245	10986	9402	2918	8744
Max	12274	12407	33936	12323	14846	12792	15858	14209	7223	13051
Min	2254	712	17059	5156	4887	5272	4447	4968	366	3791
SD	1581	2082	3310	1492	1904	1536	2074	1765	1083	1801
Skew	0.19	0.87	0.15	-0.20	0.29	0.07	-0.06	0.19	0.60	-0.05
Kurt	3.74	4.27	2.87	2.82	2.79	2.86	3.15	2.94	4.32	2.74
CV (%)	21.92	43.53	13.22	16.85	21.29	16.60	19.08	18.49	36.24	20.76
J-B	4.22	28.32	0.63	1.13	2.39	0.25	0.25	0.94	19.28	0.47
Prob.	0.12	0.00	0.73	0.57	0.30	0.88	0.88	0.63	0.00	0.79






SHWBL
y = -1.3889x + 25140
40000
30000
20000
10000
GNWBL
y = 4.7053x + 14962
25000
20000
15000
10000
ORISS
y = -4.1415x + 15033
23000
18000
13000
8000
JHKND
y = -4.6933x + 13701
20000
18000
16000
14000
12000
10000
8000
BIHAR
y = -4.3952x + 12508
20000
15000
10000
5000
EUPRA
y = -7.9868x + 10677
20000
15000
10000
5000
WUPPL
y = -4.6042x + 9017.1
18000
13000
8000
3000
HARYA
y = 2.1392x + 5378.6
12000
10000
8000
6000
4000
2000
0
1871
1881
1891
1901
1911
1921
1931
1941
1951
1961
1971
1981
1991
2001
2011


PUNJB
y = 0.831x + 6177.2
WRJST
y = 2.8974x + 2776.1
ERJST
y = 0.2161x + 6921.1
WMPRA
y = 1.9633x + 9400.3
EMPRA
y = -10.166x + 13244
GUJRT
y = -1.0367x + 9164.5
SAUKU
y = 4.686x + 4438.8
KNGOA
y = 32.735x + 23418


MADMH
y = -0.3799x + 7381.3
MARAT
y = -0.118x + 8287.8
VDABH
y = -2.7299x + 11074
CHHAT
y = -13.011x + 14516
COAPR
y = 6.8201x + 9339.8
TELNG
y = 7.1338x + 8420.1
RLSMA
y = 4.6568x + 6870.9
TLNAD
y = 1.4218x + 9152

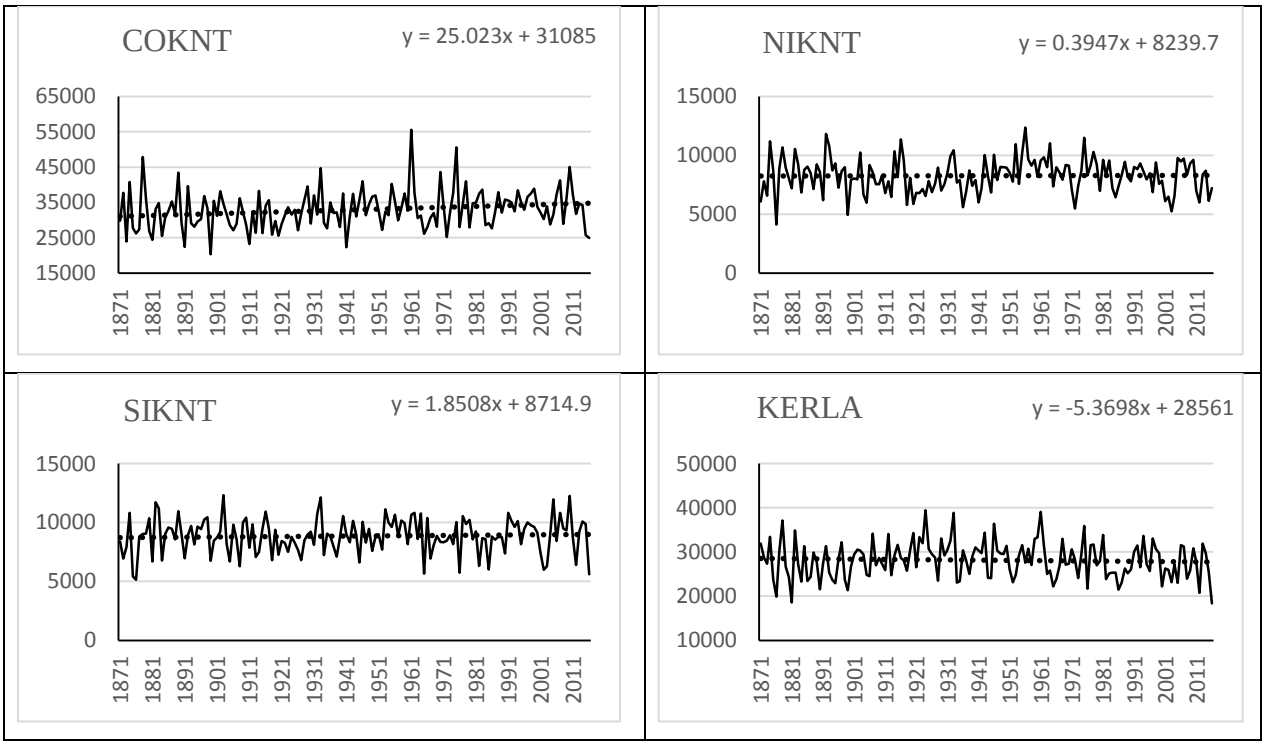


Figure 1. The variability in sub-division wise rainfall along with linear trend

5.2 Test for detection of trend

In this section we test the existence of trend in annual rainfall of 30 sub-divisions of India using Mann-Kendall (M-K) test. The values of test statistics along with corresponding p-values are given in table 3. The linear regression trend test is also conducted and the result is reported in table 3. The M-K test is more robust than that of the linear regression trend test as the former does not depend on distributional assumption of the dataset. Moreover, M-K test is robust against presence of outlier if any in the dataset. A perusal of table 3 implies that there is significant decreasing trend in rainfall in ASMEG, NMANT, EUPRA, WUPPL, EMPRA, CHHAT whereas in KNGOA, COAPR, TELNG and COKNT sub-divisions, significant increasing trend in rainfall is observed. The last column in table 3 implies the direction of the trend in rainfall of individual sub-division. The same pattern can also be seen in the plot and linear trend line depicted in figure 1.

Table 3: Test for detection of trend

	M-K test		Linear regression		Direction
Zone	**Statistics**	**Probability**	**b value**	**Probability**	

ASMEG	**-1.839**	**0.066**	**-8.460**	**0.095**	↓
NMAMT	**-3.570**	**0.000**	**-19.202**	**0.000**	↓
SHWBL	-0.548	0.584	-1.389	0.832	↓
GNWBL	1.145	0.252	4.705	0.282	↓
ORISS	-0.761	0.446	-4.142	0.295	↓
JHKND	-1.152	0.249	-4.693	0.216	↓
BIHAR	-1.093	0.274	-4.395	0.291	↓
EUPRA	**-2.042**	**0.041**	**-7.987**	**0.055**	↓
WUPPL	**-1.875**	**0.061**	-4.604	0.194	↓
HARYA	0.981	0.326	2.139	0.455	↑
PUNJB	0.398	0.691	0.831	0.810	↑
WRJST	1.557	0.120	2.897	0.174	↑
ERJST	0.132	0.895	0.216	0.948	↑
WMPRA	0.259	0.796	1.963	0.573	↑
EMPRA	**-2.528**	**0.011**	**-10.166**	**0.014**	↓
GUJRT	-0.264	0.792	-1.037	0.847	↓
SAUKU	1.101	0.271	4.686	0.253	↑
KNGOA	**3.733**	**0.000**	**32.735**	**0.001**	↑
MADMH	-0.052	0.958	-0.380	0.894	↓
MARAT	-0.078	0.938	-0.118	0.978	↓
VDABH	-0.777	0.437	-2.730	0.505	↓
CHHAT	**-2.870**	**0.004**	**-13.011**	**0.003**	↓
COAPR	**1.736**	**0.083**	**6.820**	**0.082**	↑
TELNG	**1.991**	**0.046**	**7.134**	**0.056**	↑
RLSMA	1.538	0.124	4.657	0.134	↑
TLNAD	0.277	0.781	1.422	0.639	↑
COKNT	**2.606**	**0.009**	**25.023**	**0.020**	↑
NIKNT	0.323	0.747	0.395	0.893	↑
SIKNT	0.766	0.443	1.851	0.529	↑
KERLA	-0.499	0.618	-5.370	0.512	↓

The bold values indicate significance at 10% level of significance.

5.3 Modeling by Wavelet Approach

Decomposition of rainfall series was carried out by means of MODWT. Here, maximum level of decomposition is taken as 7. Haar wavelet filter which is the simplest and oldest, was used for analyzing the data on a dyadic scale in order to capture the inherent characteristic as expressed by MODWT coefficients. The MODWT for all the 30 sub-divisional rainfall was

carried out but to save space plots of ASMEG and NMAMT sub-divisions are reported here in figure 2 and 3 respectively. In figure 2 and 3, lower most plot denotes original time-series, W_1 to W_7 denote the wavelet details components, and V_7 denotes the smoothed component of MODWT. A perusal of figure 2 and 3 indicates that at lower scale, localized variation in the dataset is detected; whereas, at higher scale, global variation in the dataset is detected. Further, the wavelet coefficients are related to differences (of various order) of (weighted) average values of portions of X_t concentrated in time. Coefficients at the top provide "high frequency" information and the coefficients at below provide "low frequency" information. Neither the wavelet coefficients (details) nor the scale coefficients (smooth) remain constant over time rather they help in detecting changes in the data at various time-epochs. The hypothesis tested in the present investigation was that whether modeling wavelet decomposed series improved the prediction accuracy or not.

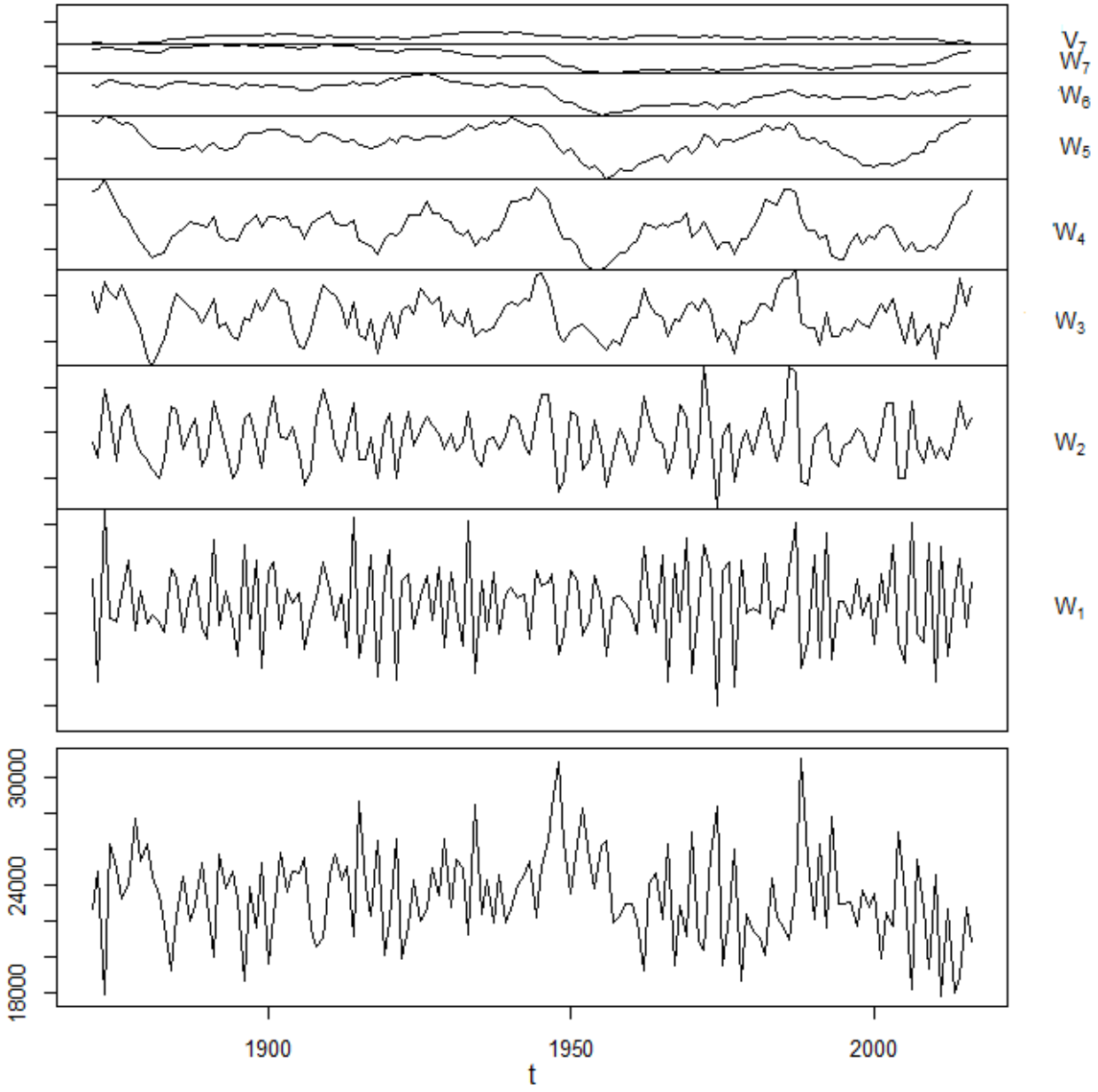


Fig. 2. MODWT of annual rainfall time-series data for ASMEG sub-division

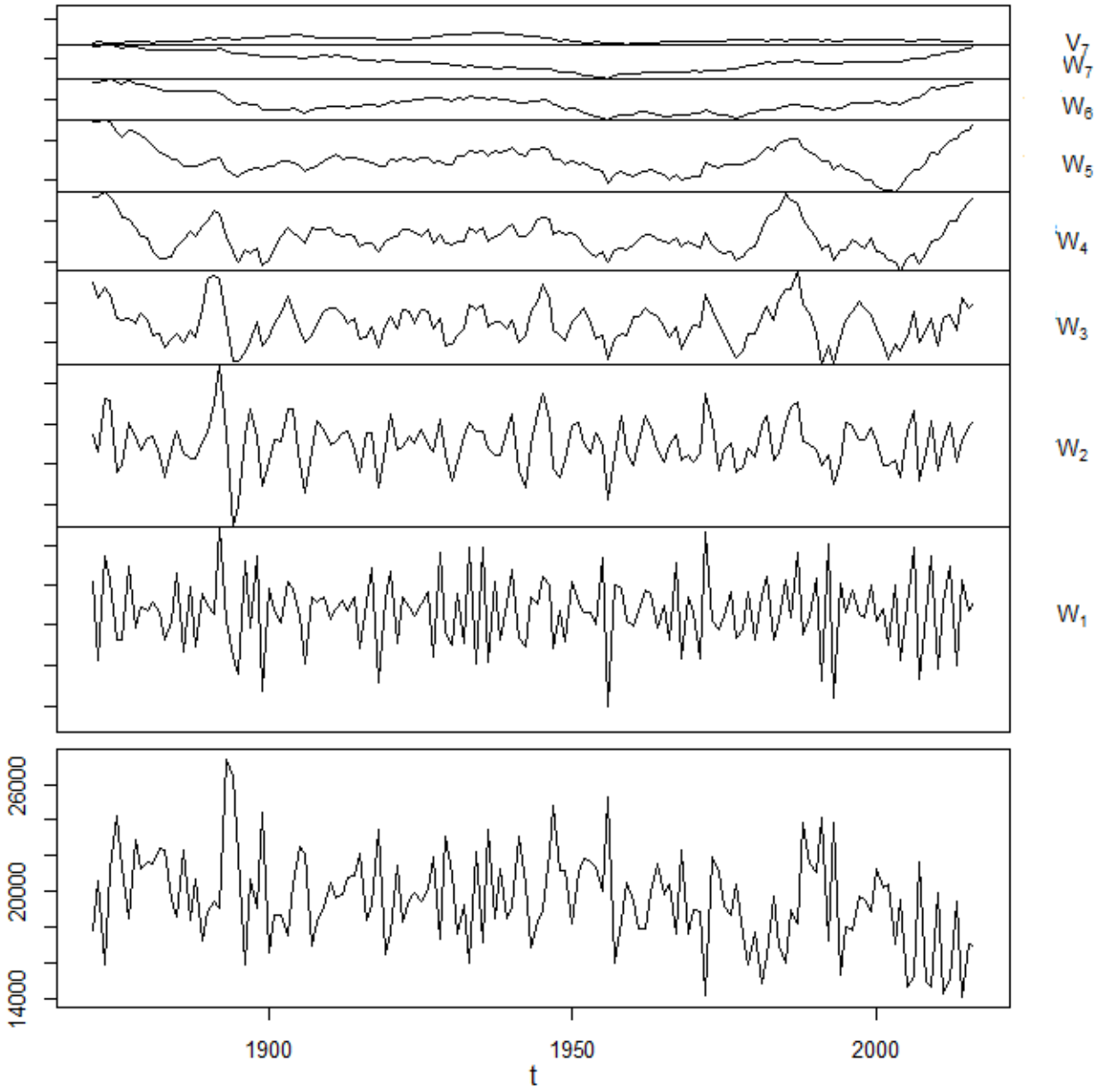


Fig. 3. MODWT of annual rainfall time-series data for NMAMT sub-division

5.4 Validation of models

The dataset during the year 1871 to 2001 was used for model development and remaining data i.e. data during 2002 to 2016 was used for validation purpose. In case of Wavelet-ANN model development, optimum value of input node and hidden layer were obtained based on minimum mean square error. Similarly, in Wavelet-ARIMA model, the optimum value of order of autoregression, differencing and moving average were obtained based on minimum information criterion. For the present investigation, optimum value of the inputs in wavelet-ANN and Wavelet-ARIMA model are presented in table 2.

Table 2: Optimum values of different input parameters for ANN and ARIMA model for wavelet decomposed series

Series	ANN		ARIMA			Series	ANN		ARIMA			Series	ANN		ARIMA		
ASMEG	**input**	**hidden**	**p**	**d**	**q**	**PUNJB**	**input**	**hidden**	**p**	**d**	**q**	**VDABH**	**input**	**hidden**	**p**	**d**	**q**
W_1	4	2	0	0	1	W_1	8	4	0	0	1	W_1	4	2	4	0	0
W_2	14	8	0	0	1	W_2	14	8	2	0	0	W_2	10	6	5	0	0
W_3	13	7	1	0	0	W_3	21	11	5	0	0	W_3	18	10	1	0	0
W_4	9	5	2	0	2	W_4	17	9	2	0	2	W_4	17	9	1	0	2
W_5	18	10	1	0	1	W_5	17	9	0	1	1	W_5	21	11	1	0	2
W_6	4	2	1	1	0	W_6	4	2	3	2	1	W_6	6	4	1	1	2
W_7	1	1	1	2	2	W_7	1	1	2	2	2	W_7	6	4	1	1	0
V_7	19	10	1	1	3	V_7	19	10	1	1	0	V_7	4	2	1	1	0
NMAMT						**WRJST**						**CHHAT**					
W_1	11	6	0	0	1	W_1	16	8	2	0	0	W_1	9	5	0	0	1
W_2	11	6	5	0	0	W_2	20	10	5	0	0	W_2	17	9	4	0	0
W_3	13	7	5	0	1	W_3	17	9	5	0	0	W_3	13	7	5	0	0
W_4	21	11	1	1	0	W_4	17	9	2	0	2	W_4	17	9	1	0	0
W_5	17	9	1	1	1	W_5	17	9	1	0	0	W_5	17	9	1	1	0
W_6	12	6	1	1	1	W_6	5	3	0	1	0	W_6	4	2	1	1	1
W_7	1	1	0	2	1	W_7	1	1	0	2	1	W_7	1	1	0	2	1
V_7	3	2	1	1	2	V_7	19	10	0	1	1	V_7	19	10	0	1	3
SHWBL						**ERJST**						**COAPR**					
W_1	11	6	4	0	0	W_1	6	4	3	0	0	W_1	7	4	4	0	1
W_2	10	6	4	0	0	W_2	14	8	4	0	0	W_2	12	6	2	0	0
W_3	20	10	1	0	0	W_3	17	9	5	0	0	W_3	8	4	5	0	0
W_4	20	10	1	0	0	W_4	12	6	2	0	2	W_4	17	9	2	0	1
W_5	17	9	1	0	0	W_5	17	9	1	0	0	W_5	19	10	1	1	0
W_6	13	7	1	0	0	W_6	1	1	1	1	0	W_6	1	1	1	1	0
W_7	1	1	0	2	1	W_7	1	1	0	1	1	W_7	1	1	0	2	1
V_7	19	10	1	1	0	V_7	19	10	1	1	0	V_7	19	10	1	1	0

GNWBL						**WMPRA**						**TELNG**					
W_1	6	4	0	0	2	W_1	6	4	0	0	1	W_1	7	4	0	0	1
W_2	10	6	1	0	0	W_2	12	6	5	0	0	W_2	8	4	2	0	0
W_3	17	9	5	0	0	W_3	18	10	5	0	0	W_3	18	10	5	0	0
W_4	17	9	1	0	0	W_4	12	6	1	0	0	W_4	17	9	2	0	1
W_5	21	11	0	1	1	W_5	17	9	2	0	2	W_5	17	9	1	0	2
W_6	1	1	2	2	1	W_6	3	2	2	0	2	W_6	6	4	1	1	0
W_7	1	1	1	1	0	W_7	1	1	2	1	2	W_7	1	1	0	2	1
V_7	19	10	2	1	2	V_7	19	10	2	1	2	V_7	5	3	1	1	0
ORISS						**EMPRA**						**RLSMA**					
W_1	3	2	0	0	1	W_1	7	4	0	0	1	W_1	8	4	2	0	1
W_2	6	4	1	0	0	W_2	6	4	4	0	0	W_2	16	8	2	0	0
W_3	16	8	1	0	0	W_3	16	8	5	0	0	W_3	15	8	1	0	0
W_4	18	10	2	0	2	W_4	11	6	1	0	0	W_4	19	10	2	0	2
W_5	17	9	2	0	0	W_5	17	9	0	1	1	W_5	17	9	1	1	0
W_6	1	1	0	1	1	W_6	1	1	2	1	2	W_6	1	1	0	1	1
W_7	1	1	0	2	1	W_7	1	1	0	2	1	W_7	1	1	2	2	1
V_7	21	11	1	1	0	V_7	19	10	2	1	2	V_7	1	1	1	1	0
JHKND						**GUJRT**						**TLNAD**					
W_1	20	10	0	0	1	W_1	4	2	4	0	0	W_1	6	4	4	0	0
W_2	10	6	5	0	0	W_2	14	8	5	0	0	W_2	8	4	5	0	0
W_3	17	9	5	0	0	W_3	13	7	5	0	0	W_3	15	8	5	0	0
W_4	19	10	1	0	0	W_4	17	9	2	0	2	W_4	10	6	2	0	0
W_5	19	10	1	0	0	W_5	17	9	1	0	0	W_5	17	9	2	0	0
W_6	1	1	0	1	1	W_6	1	1	0	1	1	W_6	1	1	2	0	1
W_7	1	1	2	2	1	W_7	1	1	1	0	0	W_7	1	1	1	1	0
V_7	19	10	2	1	2	V_7	19	10	1	0	0	V_7	19	10	2	0	1
BIHAR						**SAUKU**						**COKNT**					
W_1	11	6	5	0	0	W_1	8	4	3	0	0	W_1	6	4	1	0	2
W_2	10	6	4	0	0	W_2	17	9	2	0	2	W_2	12	6	5	0	0

W_3	13	7	5	0	0	W_3	12	6	5	0	0	W_3	10	6	5	0	0
W_4	17	9	1	0	0	W_4	20	10	0	1	1	W_4	17	9	2	0	1
W_5	17	9	1	0	1	W_5	17	9	1	0	0	W_5	17	9	1	1	1
W_6	2	2	2	0	0	W_6	5	3	1	1	0	W_6	1	1	1	1	0
W_7	1	1	1	1	0	W_7	1	1	0	2	2	W_7	1	1	0	2	1
V_7	19	10	2	1	2	V_7	5	3	1	1	0	V_7	19	10	1	1	0
EUPRA						**KNGOA**						**NIKNT**					
W_1	6	4	0	0	1	W_1	5	3	5	0	1	W_1	4	2	1	0	2
W_2	12	6	4	0	0	W_2	12	6	2	0	0	W_2	10	6	5	0	0
W_3	18	10	5	0	0	W_3	12	6	5	0	0	W_3	21	11	5	0	0
W_4	17	9	1	0	2	W_4	18	10	2	0	1	W_4	19	10	2	0	2
W_5	17	9	1	1	0	W_5	17	9	0	1	1	W_5	17	9	3	0	0
W_6	1	1	0	1	1	W_6	7	4	1	1	1	W_6	2	2	1	2	2
W_7	1	1	0	2	1	W_7	1	1	0	2	2	W_7	1	1	0	2	1
V_7	19	10	1	1	0	V_7	19	10	0	1	1	V_7	2	2	1	1	0
WUPPL						**MADMH**						**SIKNT**					
W_1	6	4	0	0	1	W_1	7	4	4	0	0	W_1	4	2	5	0	0
W_2	10	6	2	0	0	W_2	14	8	5	0	0	W_2	12	6	5	0	0
W_3	9	5	5	0	0	W_3	14	8	5	0	0	W_3	21	11	1	0	0
W_4	17	9	1	0	0	W_4	19	10	2	0	1	W_4	19	10	2	0	2
W_5	21	11	1	0	2	W_5	18	10	2	0	1	W_5	17	9	1	0	0
W_6	1	1	1	0	0	W_6	2	2	1	0	1	W_6	1	1	1	0	0
W_7	1	1	0	1	1	W_7	1	1	1	1	0	W_7	1	1	1	1	0
V_7	19	10	1	1	0	V_7	2	2	1	0	1	V_7	19	10	1	0	0
HARYA						**MARAT**						**KERLA**					
W_1	6	4	0	0	1	W_1	8	4	2	0	0	W_1	6	4	0	0	3
W_2	14	8	2	0	0	W_2	6	4	2	0	0	W_2	11	6	5	0	0
W_3	12	6	5	0	0	W_3	17	9	5	0	1	W_3	21	11	5	0	0
W_4	19	10	2	0	2	W_4	18	10	1	0	2	W_4	17	9	1	0	1
W_5	20	10	1	0	2	W_5	18	10	2	0	0	W_5	18	10	2	1	0

W_6	5	3	1	1	0	W_6	7	4	1	1	0	W_6	1	1	1	1	1
W_7	1	1	0	2	1	W_7	1	1	1	1	0	W_7	2	2	0	2	1
V_7	19	10	1	1	0	V_7	2	2	2	0	0	V_7	19	10	1	1	0

1 Input: no of input lags; Hidden: Number of nodes in the hidden layer; p: order of Autoregression; d: order of differencing; q: order of Moving average

One-step ahead forecasts of annual rainfall for hold-out data i.e. the rainfall data for the years 2002 to 2016 by using the three approaches i.e. Wavelet-ARIMA, Wavelet-ANN and ARIMA were computed. For comparison of predictive accuracy of two approaches, MAPE and RMSE was used as described in equation 2 and 3. Both the statistics value for all the sub-divisions are reported in table 3. It is clear from table 3 that for all the sub-divisions, Wavelet-ANN and Wavelet-ARIMA approach perform better than the ARIMA model. This is because of denoising the series using wavelet and modeling decomposed series independently. Between Wavelet-ANN and Wavelet-ARIMA model, Wavelet-ANN is little better than Wavelet-ARIMA for the dataset under consideration.

Table 3. One- step ahead prediction performance of Wavelet-ANN, Wavelet-ARIMA and ARIMA model.

	Wavelet-ANN		Wavelet-ARIMA		ARIMA	
Sub-division	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
ASMEG	13.46	3216.79	14.07	3475.72	15.12	3514.01
NMAMT	17.62	3657.784	20.93	3789.62	21.64	3861.79
SHWBL	11.45	3072.763	11.41	3035.37	11.43	2759.75
GNWBL	12.65	2273.959	13.43	2468.98	14.01	2472.62
ORISS	11.46	1934.119	11.20	1922.05	13.32	2192.99
JHKND	13.18	1996.482	14.51	2055.50	15.64	2162.56
BIHAR	16.60	2039.144	17.88	2144.05	18.76	2154.21
EUPRA	22.36	2013.251	23.24	2059.59	25.80	2261.78
WUPPL	25.83	1996.638	25.61	1977.52	26.02	2006.33
HARYA	19.90	1117.654	21.10	1243.07	25.61	1291.25
PUNJB	26.89	1481.363	27.60	1539.40	33.07	1701.38
WRJST	25.75	764.8907	26.39	867.36	35.22	1053.34
ERJST	13.86	1107.872	14.17	1146.67	19.86	1461.28
WMPRA	13.82	1663.771	14.51	1681.81	14.90	1737.52
EMPRA	16.83	2274.272	17.62	2338.78	22.20	2623.64
GUJRT	20.08	2265.432	20.05	2397.86	21.81	2483.07
SAUKU	23.91	2325.803	25.87	2535.86	27.31	2683.84

KNGOA	13.12	5310.639	13.76	5392.85	14.02	5341.57
MADMH	16.15	1703.062	17.72	1719.36	18.62	1671.41
MARAT	18.43	1385.009	20.21	1588.56	22.72	1766.66
VDABH	17.20	2075.468	17.31	2078.53	17.78	2082.81
CHHAT	11.21	1722.553	10.50	1631.91	11.55	1799.76
COAPR	16.77	1656.19	16.64	1789.28	17.74	2442.65
TELNG	18.33	1722.471	21.58	1913.72	22.61	1966.40
RLSMA	14.34	1218.503	16.28	1341.41	16.52	1620.77
TLNAD	18.48	2050.329	18.33	2084.31	20.86	2218.15
COKNT	12.73	5011.439	12.70	5087.11	13.35	5354.23
NIKNT	19.14	1565.29	18.58	1536.23	19.22	1567.93
SIKNT	19.43	2049.332	20.15	2063.94	20.22	2003.27
KERLA	13.59	3808.633	13.70	3814.90	15.63	4402.48

Conclusions

The overconfidence and lack of reliability for regional rainfall forecasts is a common problem amongst the researchers. Wavelet decomposition approach is one of the most important approach to smooth the series as well as to extract the actual signal from the noisy and utilize the strengths of a range of independent models to increase the reliability and accuracy of climate predictions. Modeling of rainfall series assume prime importance both in the local as well as in global level. In the present investigation, the advantage of powerful nonparametric Wavelet methodology in frequency domain for modeling and forecasting of sub-divisional rainfall in India employing Haar wavelet filter is highlighted. Moreover, the feature of rainfall in a location may not be always linear so that it can be modelled through the classical ARIMA model. To accommodate the pattern of nonlinearity and complexity, decomposition of the series under consideration is required. When the original series has much nonlinearity as its property, the MODWT has simplified it by breaking it into its sub-frequencies. Therefore, the ANN can now model the details and approximate components sufficiently so that the accuracy of the forecasting process is improved up to a marked extent. Therefore, the combination of wavelet approach along with classical time series model i.e. ARIMA model and promising machine learning technique i.e. ANN is applied for forecasting annual rainfall in 30 subdivisions of India. Superiority of Wavelet-ARIMA and Wavelet-

ANN approach over traditional ARIMA model is demonstrated in terms of RMSE and MAPE. In Wavelet-ANN and Wavelet-ARIMA approach the minimum and maximum MAPE has been found in CHHAT and PUNJB sub-division respectively. The study has revealed that the Wavelet-ARIMA model and Wavelet-ANN approach could be used successfully for modeling as well as forecasting of annual rainfall in different sub-divisions of India.

Chapter VI

Out-Of-Sample Forecasting Of ARFIMA-GARCH Model And Periodicity In Rainfall

Out-of-sample forecasting by ARFIMA-GARCH model

Formulae for multi-step ahead out-of-sample forecast and forecast error variance for ARFIMA-GARCH model has been developed by recursive use of conditional expectation in the same line of Ghosh *et al*. (2011) and Paul *et al*. (2014).

Let, $\{y_t\}$ follows an ARFIMA $(1, \mathrm{d}, 1)$ process with error, $\{e_t\}$ follows GARCH (1,1) model, can be expressed as,

$$(1-\rho_1 L)y_t = (1-\theta_1 L)(1-L)^{-d}e_t \qquad (1)$$

$$e_t = h_t^{1/2}\zeta_t \text{ where } h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 h_{t-1}$$

ζ_t is a i.i.d. random variable with zero mean and constant variance σ^2. After expanding the term $(1-L)^{-d}$ with Taylor series expansion and by ignoring the higher order terms, equation (1) can be written as,

$$\Rightarrow y_t - \rho_1 y_{t-1} = (e_t - \theta_1 e_{t-1})(1 + dL + \frac{d(d-1)}{2}L^2)$$

$$\Rightarrow y_t = \rho_1 y_{t-1} + e_t + (d-\theta_1)e_{t-1} + \left[\frac{d(d-1)}{2} - \theta_1\right]e_{t-2} - \frac{d(d-1)}{2}e_{t-3}$$

Let, t data points are utilized for modelling and the parameter estimation purpose and k data points are conserved for model validation purpose. The i-step ahead out-of-sample forecast and forecast conditional error variance are denoted by $\hat{y}_{t+i|1,2,..,t}$ and $\hat{h}_{t+i|1,2,\ldots,t}$, $i = 1,2,3$ respectively.

So, One-step ahead prediction, i.e. $\hat{y}_{t+1|1,2,\ldots,t}$ is computed as,

$$\hat{y}_{t+1|1,2,\ldots,t} = \hat{\rho}_1 y_t + (\hat{d} - \hat{\theta}_1)\hat{e}_t + \left[\frac{\hat{d}(\hat{d}-1)}{2} - \hat{\theta}\right]\hat{e}_{t-1} - \frac{\hat{d}(\hat{d}-1)}{2}\hat{e}_{t-2}$$

where $\hat{e}_t$ is the residual in the fitted ARFIMA $(1, \mathrm{d}, 1)$- GARCH (1,1) model at time t.

And the corresponding one-step ahead prediction of error variance is calculated as,

$$\hat{h}_{t+1} = E[\{e_{t+1} - \hat{e}_{t+1|1,2,\ldots,t}\}^2 | e_1, \ldots, e_t]$$

$$= \hat{\alpha}_0 + \hat{\alpha}_1 {\hat{e}_t}^2 + \hat{\beta}_1 \hat{h}_t$$

Second-step ahead prediction, i.e. $\hat{y}_{t+2|1,2,\ldots,t}$ is computed as,

$$\hat{y}_{t+2|1,2,\ldots,t} = \hat{\rho}_1 \hat{y}_{t+1} + \left[\frac{\hat{d}(\hat{d}-1)}{2} - \hat{\theta}\right] \hat{e}_t - \frac{\hat{d}(\hat{d}-1)}{2} \hat{e}_{t-1}$$

where $E(\hat{e}_{t+j} | y_1, y_2, \ldots, y_2) = \begin{cases} 0, j > 0 \\ \hat{e}_{t+j}, \ j \leq 0 \end{cases}$

Corresponding second-step ahead forecast error variance is calculated as,

$$\hat{h}_{t+2|1,2,\ldots,t} = E[\{e_{t+2} - \hat{e}_{t+2|1,2,\ldots,t+1}\}^2 | e_1, \ldots, e_{t+1}]$$

$$= E\left[E\{e_{t+2} - \hat{e}_{t+2|1,2,\ldots,t+1}\}^2 \middle| e_1, \ldots, e_{t+1}\right] + V[\{\hat{e}_{t+2|1,\ldots,t+1}\} | e_1, \ldots, e_{t+1}]$$

$$\hat{h}_{t+2|1,2,\ldots,t} = \hat{\alpha}_0 + \hat{\alpha}_1 {\hat{e}_{t+1}}^2 + \hat{\beta}_1 \hat{h}_{t+1}$$

we know that, $\mathrm{E}(e_{t+1}^2 | e_1, \ldots, e_t) = h_{t+1}$ so, $\hat{e}_{t+1}^2 = h_{t+1}$

Periodicity in Rainfall

The periodogram was first introduced by Schuster (1898) and was used to search for hidden periodicities and, naturally, it became a tool for testing the presence of periodic components. Consider a stationary process Z_t which we assume has zero mean and absolutely continuous spectral density function (SDF) given by S(f) with autocovariance sequence (acvs), the following relationship between the spectral density and autocovariance sequence holds

$$S(f) = \sum_{\tau=-\infty}^{\infty} S_\tau e^{-i2\pi f\tau}, \quad for \quad f \in (-1/2, 1/2]$$

This relation holds if the auto-covariance sequence (acvs) is square summable, i.e., if $\sum_{\tau=-\infty}^{\infty} S_\tau < \infty$
For LM processes the acvs is not necessarily a square summable sequence, so that the relation does not always hold. A naïve estimator of the spectrum is given by the periodogram, in which the autocovariance is replaced by a biased estimator

$$\hat{S}_\tau = \begin{cases} \frac{1}{N}\sum_{t=1}^{N-|\tau|} Z_t Z_{t+|\tau|}, & for\ |\tau| \le N-1 \\ 0, & for \tau > N-1 \end{cases}$$

The periodogram is hence given by

$$\hat{S}(f) = \sum_{\tau=-(N-1)}^{N-1} \hat{S}_\tau e^{-i2\pi f\tau} = \frac{1}{N}\left|\sum_{t=1}^{N} Z_t e^{-i2\pi f\tau}\right|^2, \quad f \in (-1/2, 1/2]$$

The periodogram is often evaluated at the Fourier frequencies k, where is the greatest integer less than or equal to x.

$$\varphi(k) = k/N, k = -\left\lfloor \frac{N-1}{2} \right\rfloor, \ldots, \left\lfloor \frac{N}{2} \right\rfloor$$

Fisher (1929) proposed a test for the periodicity when d = 0, based on the quotient of the maximum ordinate of the periodogram and the sum over all periodogram ordinates at the Fourier frequencies. Let us have a sample size N where N = 2m + 1 if N odd and N = 2m + 2 if N even. The Fisher's test is based on the statistic

$$g = \frac{\max_{1\le k\le m} \hat{S}(\varphi_k)}{\sum_{q=1}^{m} \hat{S}(q/N)}$$

The distribution of g under the null hypothesis H0: A1= 0 (i.e., that there is no periodicity) is given by

$$P[g > g_0] = \sum_{j=1}^{M} (-1)^{j-1} \binom{m}{j} (1 - jg_0)^{m-1}$$

where M is the largest integer satisfying both $M < 1/g_0$ and $M \le m$; see Fisher (1929). Fisher's test has often been used to detect periodic components. However, as observed by Siegel (1980), Fisher's test is quite conservative in detecting "compound" periodicities. Another extension of Fisher's test which may be used to detect "compound" periodicities was proposed by Siegel (1980). Siegel's test is based on all large periodogram ordinates exceeding a certain threshold value. He suggested the following rescaled periodogram

$$\tilde{S}(\varphi_k) \equiv \frac{\hat{S}(\varphi_k)}{\sum_{q=1}^{m} \hat{S}(j/N)}$$

For a given value (≤ 1) Siegel's statistic consists of the positive excess of g above

$$T_\lambda = \sum_{k=1}^{m} \left(\tilde{S}(\varphi_k) - \lambda_{gF}\right)_+$$

where $a_+ = \max(a\ 0)$ is the positive-part function. Note that, when = 1 we have Fisher's test. Siegel (1980) obtained the distribution function for T as

$$P[T_\lambda > t] = \sum_{j=1}^{M} \sum_{k=0}^{j-1} (-1)^{j+k+1} \binom{m}{j} \binom{j-1}{k} t^k \left(1 - j\lambda_{gF} - t\right)_+^{m-k-1}$$

Algorithm by Araghi et al (2014)

- Each dataset was decomposed through DWT using the Daubechies (db) wavelet family as the mother wavelet. This split each series into A and D components.
- The Mann-Kendall test was applied to the original time series, to the decomposed components (i.e. A and D components), and also to combinations of A plus one or two D components.
- The MK test was performed on the original time series, the decomposed components, and the combinations of A and D components.
- The dominant component(s) that had the greatest impact on the rainfall time series were determined by matching the MK Z-values for the original series as well as with the combination of the decomposed series.

The above algorithm has been used for computation of periodicities in rainfall in different sub divisions. The dataset considered was annual rainfall during the period 1887 to 2014 for all 30 subdivisions of India

It is observed that the periodicities of 2-4 years are significant in most of the subdivisions.

Combinations of wavelet coefficients and smooth coefficient

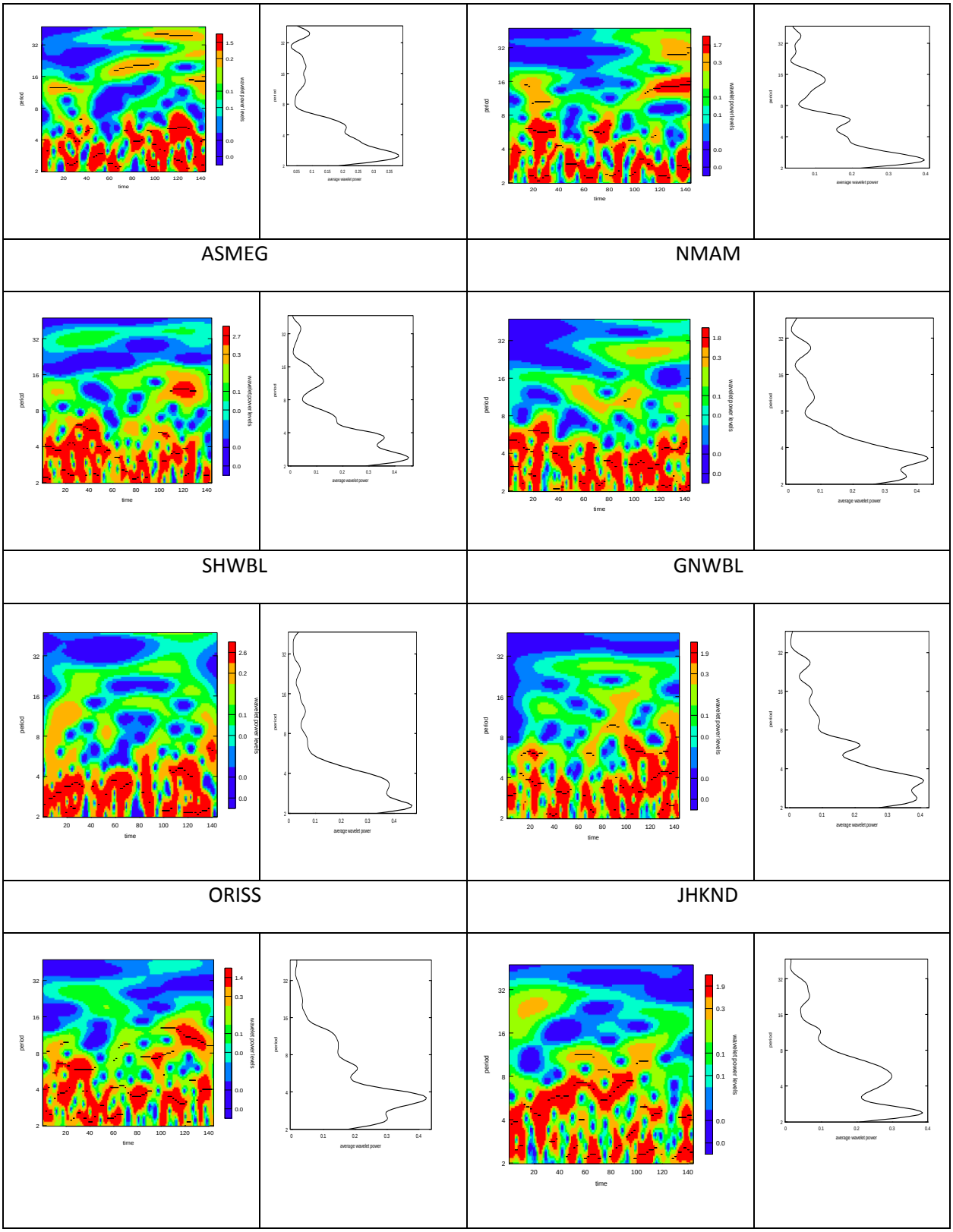
1	D1			19	D1	D5	A
2	D2			20	D1	D6	A
3	D3			21	D1	D7	A
4	D4			22	D2	D3	A
5	D5			23	D2	D4	A
6	D6			24	D2	D5	A
7	D7			25	D2	D6	A
8	A			26	D2	D7	A
9	D1	A		27	D3	D4	A
10	D2	A		28	D3	D5	A
11	D3	A		29	D3	D6	A
12	D4	A		30	D3	D7	A
13	D5	A		31	D4	D5	A
14	D6	A		32	D4	D6	A

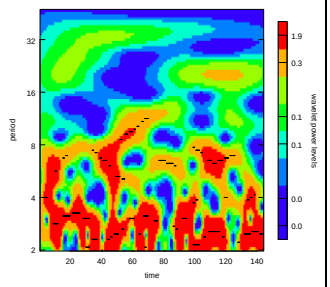
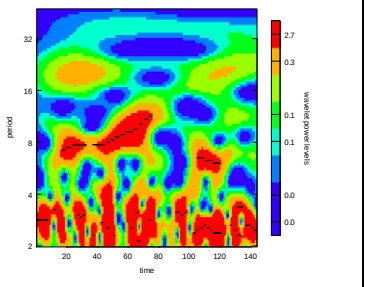
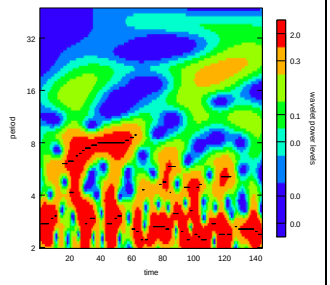
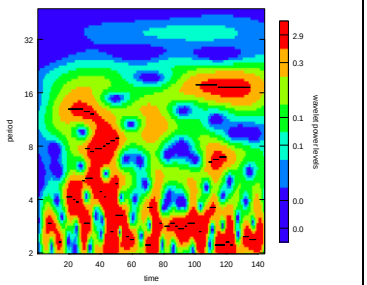
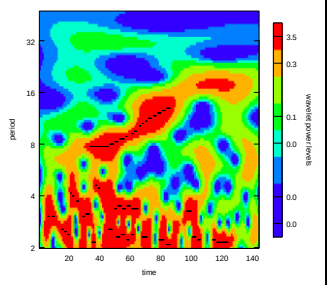
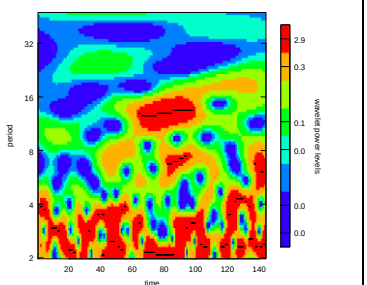
15	D7	A		33	D4	D7	A
16	D1	D2	A	34	D5	D6	A
17	D1	D3	A	35	D5	D7	A
18	D1	D4	A	36	D6	D7	A

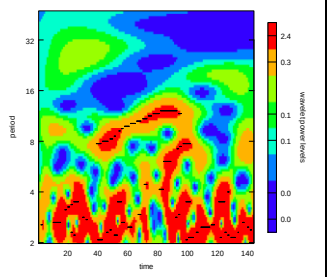
Periodicity

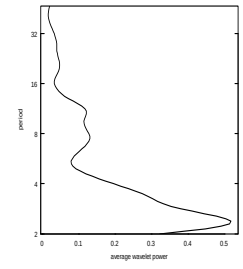
Sub-Division	Periodicity range	Peak periodicity	Sub-Division	Periodicity range	Peak periodicity
ASMEG	2	2.4	GUJRT	2-4	2.4
NMAMT	2-4	2.5	SAUKU	2-4	2.5
SHWBL	4	2.5	KNGOA	2-6	2.6
GNWBL	2-6	3.5	MADMH	2-8	3.2
ORISS	2-4	2.5	MARAT	2-8	2.5
JHKND	2-8	3.6	VDABH	2-8	4.8
BIHAR	2-6	2.5	CHHAT	2-8	2.8, 7.8
EUPRA	2-8	3.5	COAPR	2-6	2.5
WUPPL	2-4	3.8	TELNG	2-4	2.6
HARYA	2-8	2.5	RLSMA	2-8	2.6
PUNJB	2-4	2.8	TLNAD	2-12	2.8
WRJST	2-4	2.7	COKNT	2-6	3.0
ERJST	2-4	2.6	NIKNT	2-6	2.8
WMPRA	2-6	2.6	SIKNT	2-6	2.6
EMPRA	2-4	2.4	KERLA	2-6	3.0

Periodicities in rainfall plot for all the sub-division

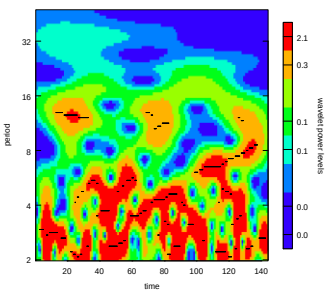


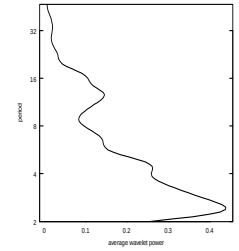
BIHAR	EUPRA
	
WUPPL	HARYA
	
PUNJB	WRJST
	
ERJST	WMPRA



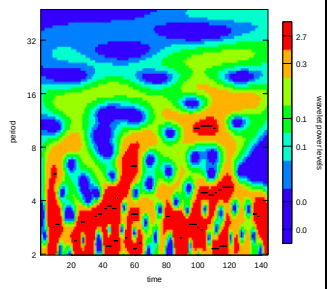


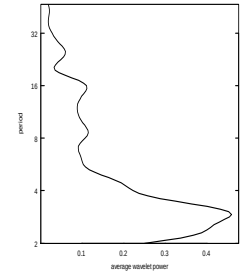
EMPRA



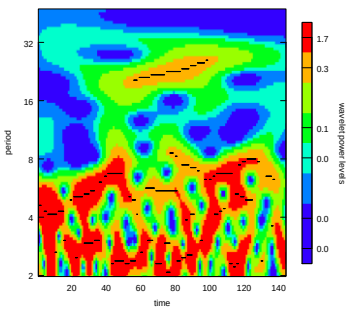


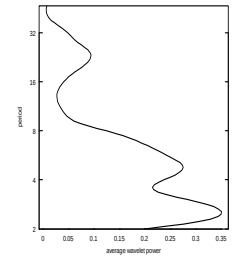
GUJRT



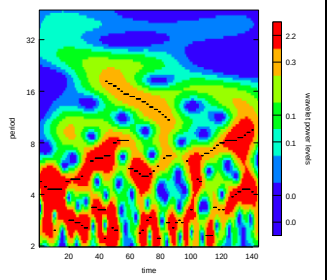


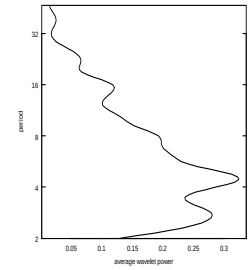
SAUKU



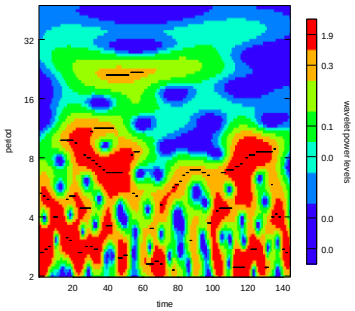


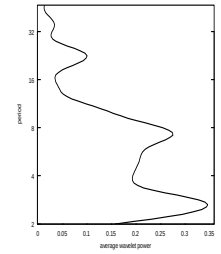
KNGOA



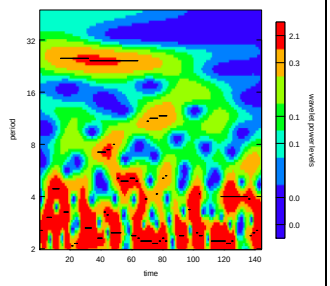


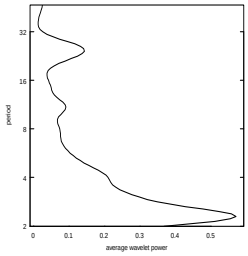
MADMH



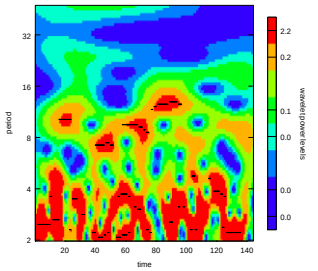


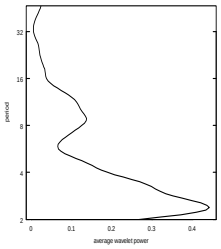
MARAT



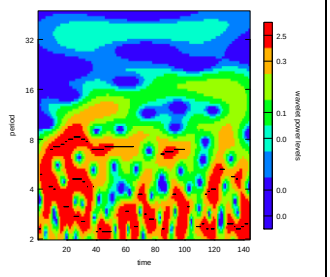


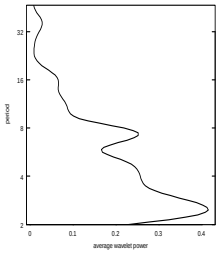
VDABH



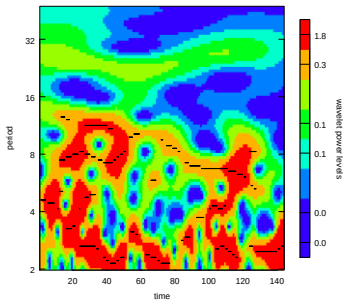


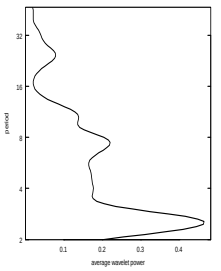
CHHAT



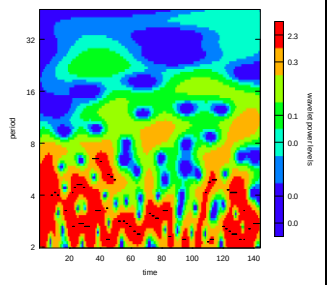


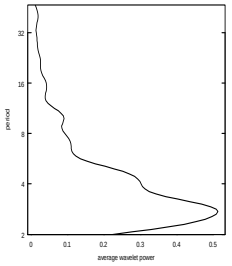
COAPR



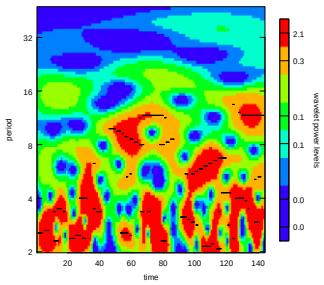


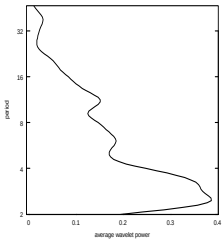
TELNG



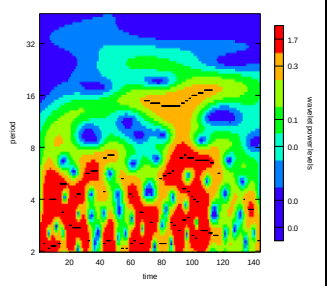


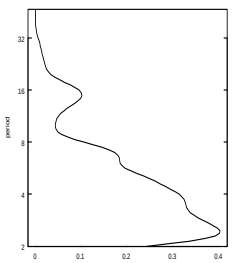
RLSMA

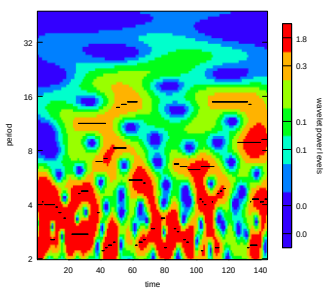


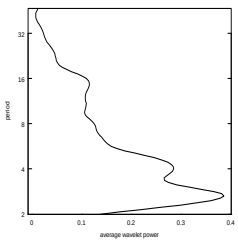


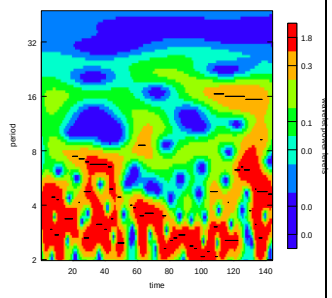
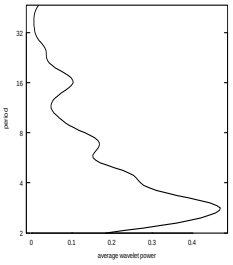
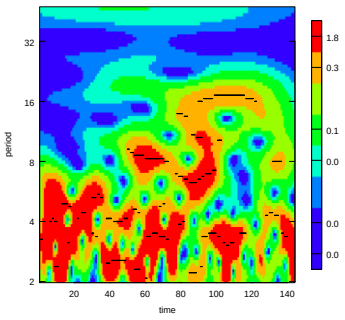
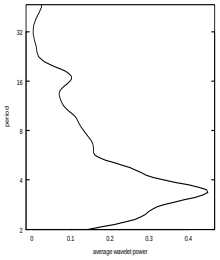
TLNAD









COKNT		NIKNT	
			

SUMMARY

Presence of long memory in climatic variables is frequently observed. The trend assessment becomes difficult in the presence of long-memory as the usual methods are not capable to take care of this property during trend estimation. In order to estimate the trend in presence of long memory, the non-parametric wavelet method has become popular in the recent time. The discrete wavelet transformation (DWT) re-expresses a time-series in terms of coefficients that are associated with a particular time and a particular scale. In the present study, DWT has been applied to estimate the monthly rainfall trend for the monsoon months: June-September in ten selected sub-divisions of India using "Haar" wavelet filter. The results from DWT were cross checked with the non-parametric Mann-Kendall (M-K) test. The investigation reveals that the monthly rainfall trend for the monsoon months of different sub-divisions in India are significantly decreasing over the years. However, in some of the sub-divisions, rainfall trend is increasing. DWT reveals significant trend in most of the sub-divisions whereas M-K test reveals that most of the trends are not significant at 5% level. The variation in monthly rainfall for June-September for the studied zones were also investigated by wavelet using Haar filter. The local as well as global variation in rainfall is observed by DWT plot of the respective months in the respective locations. The variability in rainfall is evident in the recent decades in all the locations.

The surface temperatures over a given region vary seasonally and annually depending upon latitude, altitude and location with respect to geographical features such as a water body (river, lake or sea), mountains, etc. Probably one of the most widely quoted aspects of climatic change is the significant increase in global mean air temperature during the past century. The modeling issue of maximum temperature series in India which possesses characteristic of long-term memory is addressed and TSF has proven to be a robust approach of capturing such long memory even in the presence of detected structural break. Time series modeling of temperature series assume prime importance both in the local and global levels. Wavelet transformation was applied to decompose the temperature series into time–frequency domain in order to study the local as well as global variation over different scale and time epochs. It is found that there is significant increase in maximum temperature over the years in India. The shift in maximum temperature in India occurred during mid of 1963 as detected by the statistical test. Temperature data series exhibiting long-range dependence property combined with structural break required to be modeled with a concrete and valid technique

which can overcome the issue of loss of information, biased estimates and inaccurate forecast. In this regard, TSF approach has been found to serve the better results. For long memory processes with a change in mean level, the present study on maximum temperature data has established the outperformance of TSF methodology in terms of SSE, MSE and RMAPE criterions over the AR-Truncation approach.

Time series analysis of weather data can be a very valuable tool to investigate its variability pattern and, maybe, even to predict short- and long-term changes in the time series. In this study, the long memory behaviour of monthly minimum and maximum temperature of India for the period 1901 to 2007 by means of fractional integration techniques has been investigated. The results show that the time series can be specified in terms of autoregressive fractionally integrated moving average (ARFIMA) process. Both the series were found to be integrated with orders of integration smaller than 0.5 ensuring the long memory stationarity. Wavelet methodology in frequency domain with Haar wavelet filter was applied in order to see the oscillation at different scale and at different time epochs of the series. Multiresolution analysis (MRA) was carried out to explore the local as well as global variations in both the temperature series over the years. The variability in minimum temperature is found to be more than maximum temperature. Though there is no clear significance trend in the temperature series in the long run, but there are pockets of change in the temperature pattern. The predictive ability of ARFIMA model was investigated in terms of relative mean absolute percentage error.

Long memory time series have been analysed by using ARFIMA models. Model parameter *d* reflects the long memory in the maximum and minimum temperature series. It is found that in the both the series long memory parameter is significant. The study has revealed that the ARFIMA model could be used successfully for modelling the temperature series. The predictive ability of ARFIMA model was investigated in terms of relative mean absolute percentage error. The variability in minimum temperature is found to be more than maximum temperature. The study reveals that there are pockets of change in the temperature pattern (both in maximum as well as in minimum temperature) which may be clearly visible by vertical clustering of coefficients in MRA.

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.

The overconfidence and lack of reliability for regional rainfall forecasts is a common problem amongst the researchers. Moreover, the feature of rainfall in a location may not be always linear so that it can be modelled through the classical ARIMA model. To accommodate the pattern of nonlinearity and complexity, decomposition of the series under consideration is required. When the original series has much nonlinearity as its property, the MODWT has simplified it by breaking it into its sub-frequencies. Therefore, the ANN can now model the details and approximate components sufficiently so that the accuracy of the forecasting process is improved up to a marked extent. Therefore, the combination of wavelet approach along with classical time series model i.e. ARIMA model and promising machine learning technique i.e. ANN is applied for forecasting annual rainfall in 30 subdivisions of India. Superiority of Wavelet-ARIMA and Wavelet-ANN approach over traditional ARIMA model is demonstrated in terms of RMSE and MAPE. In Wavelet-ANN and Wavelet-ARIMA approach the minimum and maximum MAPE has been found in CHHAT and PUNJB sub-division respectively.

REFERENCES

Abry, P. and Veitch, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE Trans. Info. Theory*, **44**, 2-15.

Adamowski, K., Prokoph, A. and Adamowski, J. (2009) Development of a new method of wavelet aided trend detection and estimation. Hydrological Processes, 23, 2686–2696

Aggarwal PK. 2009. Vulnerability of Indian agriculture to climate change: current state of knowledge, paper presented at the National Workshop – *Review of Implementation of Work Programme towards Indian Network of Climate Change Assessment, October 14*. Ministry of Environment and Forests, New Delhi, http://moef.nic.in/downloads/others/Vulnerability_PK%20Aggarwal.pdf.

Almasri A, Locking H and Shukur, G. 2008. Testing for climate warming in Sweden during 1850–1999, using wavelets analysis. *Journal of Applied Statistics,* **35**, 431-443.

Almasri A. (2011). A New Approach for Testing Periodicity. Communications in Statistics - Theory and Methods. 40: 1196–1217

Almasri, A., 2010, "Tests for Trend: A Simulation Study", *Communications in Statistics - Simulation and Computation",* **39,** 3, 598-611.

Aminghafari M and Poggi JM. 2007. Forecasting time series using wavelets. *Int. J. Wave. Mult. Inf.* Proc., **5**, 709-724.

Aminghafari M and Poggi JM. 2012. Nonstationary Time series forecasting using wavelets and kernel smoothing, *Communications in Statistics - Theory and Methods,* **41**, 485-499.

Aminghafari, M. and Poggi, J.-M. (2007). Forecasting time series using wavelets. *Int. J. Wave. Mult. Inf. Proc*., **5**, 709–724.

Antoniadis A. 1997. Wavelets in statistics: A review. *Journal of Italian Statistical Society*, **6**:97-144

Antoniadis, A., 1997, "Wavelets in statistics: A review", *Journal of the Italian Statistical Society,* **6**, 291-304.

Azad, S., Narasimha, R. and Sett, S.K. (2008). A wavelet based significance test for periodicities in Indian monsoon rainfall. *Int. J. Wave. Mult. Inf. Proc*., **6**, 291–304.

Beran J. (1995). *Statistics for long memory processes*. Chapman & Hall.

Beran, J. (1994). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society*, **B 57** (4) 659-672

Beran, J. (1994). *Statistics for Long-Memory Processes.* Chapman and Hall Publishing Inc. New York.

Birthal PS, Negi, DS, Kumar, S, Aggarwal S, Suresh A. and Khan T. 2014. How sensitive is Indian agriculture to climate change? *Indian Journal of Agricultural Economics*, **69** (4), 474-487.

Booth, G.G., F.R. Kaen, and P.E. Koveos. (1982). R/S analyses of foreign exchange rates under two international monetary regimes. *Journal of Monetary Economics*, **10**, 407-415.

Box GEP, Jenkins GM and Reinsel GC. 2007. *Time-Series Analysis: Forecasting and Control*. 3rd ed. Pearson Education, India.

Brockwell, P.J. and Davis, R.A. 1991. *Time Series: Theory and Methods*, 2nd Edition. Springer, New York

Daubechies, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia.

De Salvo M, Raffael R, and R Moser. 2013. The Impact of climate change on permanent crops in an Alpine region: A Ricardian Analysis, *Agricultural Systems,* **118**, 23-32.

Dhimri, A. P., Mohanty, U. C. and Rathore, L. S. (2005). Minimum temperature forecast at Manali, India. *Current Science*, **88** (4), 927-934

Diebold, F.X., & Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, **105**, 131159

Eichner JF, Koscielny-Bunde E, Bunde A, Havlin S and Schellnhuber HJ. 2003. Power-law persistence and trends in the atmosphere: A detailed study of long temperature records. *Physical Review* E, **68**, 046133

Engle, R.F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom in1ation. *Econometrica*, **50**, 987–1007

Fisher, R. A. (1929). Tests of significance in harmonic analysis. *Proc. Roy. Soc. London, Ser.* A **125**:54–59.

Fryzlewicz P, Bellegem SV and von Sachs R. 2003. Forecasting non-stationary time series by wavelet process modeling. *Annals of the Institute of Statistical Mathematics*, **55**, 737-764.

Fryzlewicz, P., Bellegem, S.V. and von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Ann. Inst. Stat. Math*., **55**, 737-764.

Fung, H.G., W.C. Lo, and J.E. Peterson. (1994). Examining the Dependence in Intra-Day Stock Index Futures. *The Journal of Futures Markets*. **14,** 405-419.

Gan T. 1995. Trends in air temperature and precipitation for Canada and North Eastern United States. *International Journal of Climatology*, **15**: 1115–1134.

Gan T. 1998. Hydroclimatic trends and possible climatic warming in the Canadian Prairies. *Water Resources Research*, **34**: 3009–3015.

Gedalof Z, Smith DJ. 2001. Interdecadal climate variability and regimescale shifts in Pacific North America. *Geophysical Research Letters*, **28**: 1515–1518.

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

Ghosh, H., Paul, R. K. and Prajneshu. (2013). Nonlinear Time Series Modeling and Forecasting for Periodic and ARCH Effects. *Journal of Statistics Theory and Practice*, **4**(1), 27-44

Ghosh, H., Paul, R.K. and Prajneshu, (2010). Wavelet frequency domain approach for statistical modelling of rainfall time-series data. *Journal of Statistics Theory and Practice*, **4**, 813-825.

Gil-Alana LA, 2005. Statistical modeling of the temperatures in the Northern Hemisphere using fractional integration techniques. *Journal of Climate*, **18**, 5357–5369.

Gil-Alana LA, 2008. Time trend estimation with breaks in temperature time series. Climatic Change, **89**, 325–337

Gil-Alana, L.A. (2004). Long memory behaviour in the daily maximum and minimum temperatures in Melbourne, Australia. *Meteorological Applications*, **11**, 319–328.

Gilbert CG. 1953. An aid for forecasting the minimum temperature at Denver, Colo. *Monthly Weather Review*, **81**: 233–245.

Gobena A, Gan T. 2006. Low-frequency variability in South Western Canadian streamflow: links to large-scale climate anomalies. *International Journal of Climatology*, **26**: 1843–1869.

Goswami, B. N., Venugopal, V., Sengupta, D., Madhusoodanan, M. S. and Xavier, P. K., 2006,

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. Journal of Econometrics, 14, 227-238.

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.

Guhathakurta, P. and Rajeevan, M., 2006, "*Trends in Rainfall Pattern Over India*", National Climate Centre. Report 2. India Meteorological Department, Pune, India.

Helms, B.P., F.R. Kaen, and R.E. Rosenman. (1984). Memory in commodity futures contracts. *The Journal of Futures Markets*, **10**, 559-567.

Hirsch R, Slack J, Smith R. 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research, 18: 107–121.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, **68**, 165–176.

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770-99.

Hurvich, C.M., Deo, R. and Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time-series. *Journal of Time series Analysis*, **19**, 19-46.

Huybers P and Curry W. 2006. Links between annual, Milankovitch and continuum temperature variability. *Nature*, **441**, 329–332.

Jain, S. K. and Kumr, V., 2012, "Trend analysis of rainfall and temperature data for India", *Current Science*, **102**, 37-49.

Jayawardene, H. K. W. I., Sonnadara, D. U. J. and Jayawardene, D. R., 2005, "Trends of rainfall in Sri-Lanka over the last century, *Sri Lankan Journal of Physics*, **6**, 7-17.

Jensen, M. J. (1999). Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter, Journal of Forecasting, 18, 17-32.

Jin, H. J., and Frechette, D. (2004). Fractional integration in agricultural futures price volatilities. *American Journal of Agricultural Economics*, 86, 432-443.

Kakatkar R, Gnanaseelan C, Chowdary JS, Parekh A and Deepa JS. 2017. Indian summer monsoon rainfall variability during 2014 and 2015 and associated Indo-Pacific upper ocean temperature patterns. *Theoretical and Applied Climatology*, **131**, 1235–1247.

Kallache M, Rust HW, Kropp J. (2005). Trend assessment: applications for hydrology and climate research. Nonlinear Processes in Geophysics, 12: 2001–2210.

Kangieser PC. 1959. Forecasting minimum temperatures on clear winter nights in an arid region. *Monthly Weather Review*, **87**: 19–28.

Kendall M. (1975). *Rank Correlation Methods*. Charles Griffin: London.

Killick R and Eckley IA. 2014. Changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software*, **58**(3), 1-19

Kisi O. 2010 Wavelet regression model for short-term stream flow forecasting. *Journal of Hydrology*, **389**:344–353

Kisi O. 2011. Wavelet regression model as an alternative to neural networks for river stage forecasting. *Water Resources Management*, **25**(2):579–600

Kothawale DR, and Rupa Kumar K. 2005. On the recent changes in surface temperature trends over India. *Geophysical Research Letters*, **32**, L18714

Kumar KK, Kumar KR and Pant GB. 1997. Pre-monsoon maximum and minimum temperatures over India in relation to the summer monsoon rainfall. *International Journal of Climatology*, **17**: 1115–1127.

Kumar, K. K., Kumar, K. R., Ashrit, R. G., Deshpande, N. R. and Hansen, J. W., 2004, "Climate impacts on Indian agriculture", *International Journal of Climatology*, **24**, 1375–1393.

Lennartz S and Bunde A. 2009. Trend evaluation in records with long-term memory: Application to global warming. *Geophysical Research Letters*, **36**, L16706

Lettenmaier D. (1988). Multivariate nonparametric tests for trend in water quality. Water Resources Bulletin, 24: 505–512.

Libiseller C, Grimvall A. 2002. Performance of partial Mann-Kendall tests for trend detection in the presence of covariates. *Environmetrics*, **13**: 71–84.

Lo, A. (1991). Long-Term Memory in Stock Market Prices. *Econometrica*. **59**, 1279-1313.

Lopes, S. R. C., and Mendes, B. V. M. (2006). Bandwidth selection in classical and robust estimation of long memory. *International Journal of Statistics and Systems*, **1**, 107-190.

Malamud BD and Turcotte DL. 1999. Advances in Geophysics: Long Range Persistence in Geophysical Time Series, Self-affine time series: I. Generation and analysis, Dmowska R and Saltzman B (ed.), pp 1–87. Academic Press, San Diego.

Mallows, C L. 1973. Some comments on Cp. *Technometrics*, **15**, 661-675.

Mann H. 1945. Non-parametric tests against trend. *Econometrica*, **13**: 245–259.

Meena HM, Machiwal D, Santra P. 2019. Trends and homogeneity of monthly, seasonal, and annual rainfall over arid region of Rajasthan, India. *Theoretical and Applied Climatology*, **136**: 795-811

Mendelsohn R, Dinar A and Williams L. 2006. The Distributional Impact of Climate Change on Rich and Poor Countries. *Environment and Development Economics,* **11,** 159–178.

Mills CT. 2014. Time series modelling of temperatures: an example from Kefalonia *Meteorological Applications*, **21**: 578–584.

Mohan, V., Jargle, N.K. and Kulkarni, P.D. (1989). Numerical prediction of daily maximum temperature over Ozar. *Mausam,* **40**, 227-28.

Monetti RA, Havlin S and Bunde A. 2003. Long-term persistence in the sea surface temperature fluctuations. *Physica* A, **320**, 581-589

Nagarajan, R. (2009). *Drought Assessment*. Springer, The Netherland.

Nason, G. P. and von Sachs, R., 1999, "Wavelets in time series analysis". *Philosophical Transaction of the Royal Society London A*, **357**, 2511–2526.

Ogden T. 1997. *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhauser, Boston.

Papailias, F. and Dias, G. F. 2015. Forecasting long memory series subject to structural change: A two-stage approach. *International Journal of Forecasting,* **31**, 1056-1066

Partal T, Kucuk M. 2006. Long term trend analysis using discrete wavelet components of annual precipitation measurements in Marmara region (Turkey). *Physics and Chemistry of the Earth*, **31**: 1189–1200.

Parthasarathy B, Munot AA and Kothawale D R. 1995. All India monthly and seasonal rainfall series: 1871–1993. *Theoretical and Applied Climatology*, **49**, 217–224.

Pattantyús-Ábrahám, M, Király A and Jánosi IM. 2004. Nonuniversal atmospheric persistence: Different scaling of daily minimum and maximum temperatures. *Physical Review* E, 69:021110.

Paul RK 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 RK and Birthal PS. 2015. Investigating rainfall trend over India using wavelet technique. *Journal of Water and Climate Change*, **7**(2), 365-378.

Paul RK and Samanata S. 2018. WaveletArima: Wavelet ARIMA Model. R package version 0.1.0. https://CRAN.R-project.org/package=WaveletArima

Paul RK, Birthal PS 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 RK, Prajneshu and Ghosh H. 2013. Wavelet Frequency Domain Approach for Modeling and Forecasting of Indian Monsoon Rainfall Time-Series Data. *Journal of the Indian Society of Agricultural Statistics*, **67** (3): 319-327

Paul RK, Samanta S and Gurung, B. 2015. Monte Carlo simulation for comparison of different estimators of long memory parameter: An application of ARFIMA model for forecasting commodity price. *Model Assisted Statistics and Application*, **10**(2), 116-127.

Paul RK, Sarkar S, Mitra D, Panwar S, Paul AK and Bhar LM. 2019. Wavelets based estimation of trend in sub-divisional rainfall in India. *Mausam* (Accepted)

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

Paul RK. 2019. WaveletANN: Wavelet ANN Model. R package version 0.1.0. https://CRAN.R-project.org/package=WaveletANN

Paul, R. K. (2014). Forecasting Wholesale Price of Pigeon Pea using Long Memory Time-Series Models. Agricultural Economics Research Review, **27(2),** 167-176

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., Gurung, B., Samanta, S. and Paul, A. K. (2015). Modeling Long Memory in Volatility for Spot Price of Lentil with Multi-step Ahead Out-of-sample Forecast Using AR-FIGARCH Model. *Economics Affairs*

Paul, R. K., Prajneshu, and Ghosh, H. (2011). Wavelet methodology for estimation of trend in Indian monsoon rainfall time-series data. *Indian Journal of Agricultural Science*, **81** (3), 96-98.

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.

Paul, R.K. and Birthal, P.S. (2015). Investigating rainfall trend over India using wavelet technique. *Journal of Water and Climate Change*, **7(2)**, 365-378.

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

Pelletier JD. 1997. Analysis and Modeling of the Natural Variability of Climate. *Journal of Climate*, **10**, 1331 – 1342,

Percival DB and Mofjeld HO. 1997. Analysis of subtidal coastal sea level fluctuations using wavelets. *J. Amer. Stat. Assoc.*, **92**, 868-880.

Percival, D. B. and Walden, A. T., 2000, "*Wavelet methods for time series analysis"*, Cambridge University Press, U.K.

Peters, E.E. (1989). Fractal Structure in the Capital Markets. *Financial Analysts Journal.* July-August: 32-37.

Peters, E.E. (1991). *Chaos and Order in the Capital Markets.* John Wiley & Sons, Inc. New York.

Peters, E.E. (1994). *Fractal Market Analysis*. John Wiley & Sons, Inc. New York.

Phillips, P.C.B. and P. Perron (1988). Testing for unit roots in time series regression. *Biometrika*, **75**, 335-346.

Raj, Y.E.A. (1998). Prediction of winter minimum temperature at Pune by analogue and regression method. *Mausam*, **40**, 175-80.

Rajeevan M, Pai DS, Dikshit SK and Kelkar RR. 2004. IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003. *Current Science*, **86**, 422-431.

Rapach, D. E. and Strauss, J. K. (2008). Structural breaks and GARCH models of exchange rate volatility. *Journal of Applied Econometrics*, **23**, 65-90.

Reisen, V. A., 1994, "Estimation of the fractional difference parameter in the ARFIMA(p,d,q) model using the smoothed periodogram", *Journal Time Series Analysis*, **15**, 1, 335–350.

Renaud O, Stark JL and Murtagh F. 2003. Prediction based on a multiscale decomposition. *International Journal of Wavelets, Multiresolution and Information Processing*, **1**, 217-232.

Robinson, P.M. (1995). Log-periodogram regression of time-series with long-range dependence. *The Annals of Statistics*, **23**, 1048–1072.

Robinson, P.M. (2003). *Time series with long memory*. Oxford university press, Oxford.

Rohini P, Rajeevan M, Srivastava AK. 2016. On the Variability and Increasing Trends of Heat Waves over India, *Scientific Reports*, **6**, 26153.

Said, S.E. and D. Dickey (1984). Testing for unit roots in autoregressive moving-average models with unknown order. *Biometrika*, **71**, 599-607.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. Terrest. Magnest. 3:13–41.

Shimotsu, K. (2010). Exact local whittle estimation of fractional integration with unknown mean and time trend. *Econometric Theory*, **26**(02), 501-540.

Siegel, A. F. (1980) Testing for periodicity in a time series. *J. Amer. Statist. Assoc*. **75**:345–8.

Sowell FB. 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53**, 165-188.

Spreen WC. 1956. Empirically determined distributions of hourly temperatures. *Journal of Atmospheric Sciences*, **13**: 351–355.

Sunilkumar G and Prajneshu 2004. Modeling and forecasting meteorological subdivisions rainfall data using wavelet thresholding approach. *Calcutta Statistical Association Bulletin*, **54**, 255-268.

Sunilkumar, G. and Prajneshu (2008). Modelling and forecasting the marine fish production of India using Wavelet thresholding for autocorrelated errors. *Ind. J. Fish*., **55**, 291-294.

Tanaka, K. (1999). The nonstationary fractional unit root. *Econometric Theory*, **15**, 549- 582.

Ustaoglu B, Cigizoglub HK and Karaca M. 2008. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. *Meteorological Applications*, **15**: 431–445.

Van Loon H, Jenne RL. 1975. Estimates of seasonal mean temperature, using persistence between seasons. *Monthly Weather Review*, **103**: 1121–1128.

Venkata Ramana R, Krishna B and Kumar SR. 2013. Monthly Rainfall Prediction Using Wavelet Neural Network Analysis. Water Resources Management, **27**: 3697-3711

Vidakovic B. 1999. *Statistical Modeling by Wavelets.* John Wiley, New York.

Vyushin DI and Kushner PJ. 2009. Power-law and long-memory characteristics of the atmospheric general circulation. *Journal of Climate*, **22**, 2890-2904.

Wang CSH, Bauwens L and Hsiao C. 2013. Forecasting a long memory process subject to structural breaks. *Journal of Econometrics*, **177**, 171-184.

Weiss, A. (1986). Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory*, **2**, 101-103.

Werner R, Valev D, Danov D and Guineva V. 2015. Study of structural break points in global and hemispheric temperature series by piecewise regression. *Advances in Space Research,* **56** (11), 2323-2334.

Woodcock, F. (1984). Australian experimental model output statistics forecasts of daily maximum and minimum temperature. *Monthly Weather Review*, **112**, 2112-2121

Yajima, Y., 1988, "On estimation of regression model with long memory stationary errors", *Annals of Statistics,* **16**, 791-807.

Yuan N, Fu Z, Liu S. 2014. Extracting climate memory using Fractional Integrated Statistical Model: A new perspective on climate prediction, *Scientific Reports*, **4**, 6577.