Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 67(3) 2013 319-327

Wavelet Frequency Domain Approach for Modelling and Forecasting of Indian Monsoon Rainfall Time-Series Data

Ranjit Kumar Paul, Prajneshu and Himadri Ghosh
Indian Agricultural Statistics Research Institue, New Delhi

Received 15 November 2012; Revised 26 April 2013; Accepted 07 May 2013

SUMMARY

Agricultural performance of a country, generally, depends to a large extent on the quantum and distribution of rainfall. So its accurate forecasting is vital for planning and policy purposes. An attempt is made here for modelling and forecasting of Indian monsoon rainfall time-series data by using the promising nonparametric methodology of 'Wavelet analysis in frequency domain'. Maximal overlap discrete wavelet transform (MODWT) which, unlike discrete wavelet transform (DWT), does not require the number of data points to be a power of two is employed. Haar wavelet filter is used for computing the same in order to analyze the behaviour of time-series data in terms of different times and scales. Wavelet methodology in frequency domain and Autoregressive integrated moving average (ARIMA) methodologies are applied for describing the data and for computing one-step ahead forecasts for hold-out data. Relevant computer programs are developed in SAS, Ver. 9.3 and R, Ver. 2.15.0 software packages and are appended as an Annexure. Comparative study of performance of the two methodologies is carried out from the viewpoint of one-step ahead forecasts on the basis of Root mean square prediction error (RMSPE), Mean absolute prediction error (MAPE) and Relative mean absolute prediction error (RMAPE). It is concluded that, for the data under consideration, Wavelet analysis in frequency domain approach is superior to ARIMA approach.

Keywords: ARIMA, Forecasting, MODWT, Monsoon rainfall, Wavelet, Frequency domain.

1. INTRODUCTION

Autoregressive integrated moving average (ARIMA) methodology (Box *et al*. 2007), which is a parametric approach, has virtually dominated analysis of time-series data during last several decades. Here, role of various explanatory variables enter into the model "implicitly" through response variable observations at past epochs. However, quite often it is not possible to postulate appropriate parametric form for the underlying phenomenon and, in such cases; "Nonparametric" approach is called for. Accordingly, in recent years, an extremely powerful methodology of "Wavelet analysis" is rapidly emerging (Antoniadis 1997, Vidakovic 1999, Percival and Walden 2000). Although, a number of research papers have been published dealing with various theoretical aspects of Wavelets, their application to data is still a difficult task. Wavelet analysis can be studied in two ways: One is in "time domain" and another is in "frequency domain". In respect of the former, Fryzlewicz *et al*. (2003) developed Wavelet process model for forecasting nonstationary time-series. Sunilkumar and Prajneshu (2004) applied Wavelet thresholding approach for modelling and forecasting of monthly meteorological subdivisions rainfall in Eastern Uttar Pradesh, India. Sunilkumar and Prajneshu (2008) carried out modelling and forecasting of marine fish production of India using

Corresponding author*:* Ranjit Kumar Paul
E-mail address: ranjitstat@gmail.com

Wavelet thresholding with autocorrelated errors. For the latter approach, Renaud *et al*. (2003) developed methodology for prediction of time-series data based on multiscale decomposition. Almasri *et al*. (2008) proposed a test statistic by using Wavelet decompositions to test the significance of trend in a time-series data. A 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 regression become inappropriate. In many situations, error autocovariance function exhibits a slow decay reflecting possible presence of long memory process. Wavelet analysis has been extensively used for such purposes, since it suitably matches the structure of these processes. The autocovariance function of the Wavelet transformed series exhibits different behaviour in the sense that autocovariance functions of the transformed series decay hyperbolically fast at a rate much faster than the original process. In general, series that are correlated in the time domain become almost uncorrelated in the Wavelet domain. Aminghafari and Poggi (2007, 2012) used Wavelets and kernel smoothing approach for forecasting nonstationary time-series.

Agricultural performance of a country, generally, depends to a large extent on the quantum and distribution of rainfall. So its accurate modelling is vital in planning and policy making. Accordingly, several attempts have been made in the past to develop models for describing rainfall. In Indian context, Rajeevan *et al*. (2004) have provided an excellent review of multiple and power regression models employed since 1988 along with various modifications made in these models from time to time, particularly in the identification of relevant explanatory variables. Azad *et al*. (2008) developed a Wavelet-based significance test for periodicities in Indian monsoon rainfall. Ghosh *et al*. (2010) computed size and power of the test for testing significance of trend in Indian monsoon rainfall data using Discrete wavelet transform (DWT). Paul *et al*. (2011) applied Wavelet methodology for detection of trend in Indian monsoon rainfall and found that there is a significant declining trend. However, none of the above mentioned articles dealt with the issue of forecasting Indian monsoon rainfall through Wavelets. Purpose of the present article is to discuss and apply Wavelet methodology in frequency domain for modelling and forecasting purposes and compare its efficiency with that of the usual ARIMA methodology.

2. WAVELETS

The term Wavelet is used to refer to a set of basic functions with a very special structure which is the key to its main fundamental properties. 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).

3. MAXIMAL OVERLAP DISCRETE WAVELET TRANSFORMS (MODWT)

The 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 offers several advantages over DWT. 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 Walden 2000). Finally, redundancy of MODWT wavelet coefficients modestly increases effective degrees of freedom on each scale and thus decreases variance of certain wavelet-based statistical estimates. Since MODWT is energy conserving, it is well suited for analyzing scale dependence of variability in ANOVA studies (Percival and Mofjeld 1997).

3.1 MODWT Filters

A linear filter a_t is a sequence of weights, *i.e.* $a_t \equiv \{...a_{-2}, a_{-1}, a_0, a_1, a_2, ...\}$. Linear filtration of a time-series x_t (or a stochastic process $\{x_t\}$) is defined as

$$a_t \otimes x_t \equiv \sum_m a_m x_{t-m}$$

where $\otimes$ stands for convolution operation. An important characteristic of this filter is its frequency response defined as the Fourier transform of a_t, *i.e.*

$$A(f) \equiv \sum_t a_t \exp(-i2\pi f t), \; -1/2 < f < 1/2,$$

where f is the frequency. MODWT may be thought of as linear filtration of the time-series or the stochastic process with a special set of linear filters: Wavelet filter $\tilde{h}_l$ and scaling filter $\tilde{g}_l$. These are interconnected via the so called quadrature mirror relationship:

$$\tilde{g}_l = (-1)^{l+1}\tilde{h}_{L-1-l}, \; \tilde{h}_l = (-1)^l \tilde{g}_{L-1-l}.$$

and fulfill the conditions:

$$\sum_{l=0}^{L-1} \tilde{h}_l = 0, \sum_{l=0}^{L-1} \tilde{h}_l^2 = 1/2 \text{ and } \sum_{l=-\infty}^{\infty} \tilde{h}_l \tilde{h}_{l+2n} = 0, \text{ for all}$$

nonzero integers n.

$$\sum_{l=0}^{L-1} \tilde{g}_l = 0, \sum_{l=0}^{L-1} \tilde{g}_l^2 = 1/2 \text{ and } \sum_{l=-\infty}^{\infty} \tilde{g}_l \tilde{g}_{l+2n} = 0, \text{ for all}$$

nonzero integers n.

Let $\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$ be respectively j^{th} level MODWT wavelet and scaling filters and let L_j be width of j^{th} level equivalent wavelet and scaling filter. Thus, $\tilde{h}_{j,l} = \tilde{g}_{j,l} = 0$ for $l < 0$ and $l \geq L_j$. Let $\tilde{H}_j(f)$ and $\tilde{G}_j(f)$ be respectively frequency responses of $\tilde{h}_{j,l}$ and. $\tilde{g}_{j,l}$ Then

$$\left|\tilde{H}_j(f)\right| \approx \begin{cases} 1, & 1/2^{j+1} < f \leq 1/2^j \\ 0, & Otherwise \end{cases}$$

and

$$\left|\tilde{G}_j(f)\right| \approx \begin{cases} 1, & 0 < f \leq 1/2^{j+1} \\ 0, & Otherwise \end{cases}$$

Thus, $\tilde{h}_{j,l}$ is a band pass filter for range of frequencies $1/2^{j+1} < f \leq 1/2^j$ and $\tilde{g}_{j,l}$ is a low pass filter for range of frequencies $0 < f \leq 1/2^{j+1}$. For Haar wavelet filters:

$$\tilde{h}_{j,l} \equiv \begin{cases} 1/2^j, & l = 0,...,2^{j-1}-1 \\ -1/2^j, & l = 2^{j-1},...,2^j-1 \\ 0, & Otherwise, \end{cases}$$

and for Haar scaling filters:

$$\tilde{g}_{j,l} \equiv \begin{cases} 1/2^j, & l = 0,...,2^j-1 \\ 0, & Otherwise. \end{cases}$$

Lengths of Haar filters $\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$ are $L_j = 2^j$.

3.2 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 $(\tilde{\boldsymbol{W}}_j)$ and scaling $(\tilde{\boldsymbol{V}}_j)$ coefficients are defined as:

$$\tilde{W}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} X_{t-l \bmod N},$$

$$\tilde{V}_{j,t} \equiv \sum_{l=0}^{L_j-1} \tilde{g}_{j,l} X_{t-l \bmod N}, \tag{3.1}$$

where $\tilde{h}_{j,l} \equiv h_{j,l}/2^{j/2}$ are j^{th} level MODWT wavelet filters, and $\tilde{g}_{j,l} \equiv g_{j,l}/2^{j/2}$ 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

$$\boldsymbol{X} = \sum_{j=1}^{J_0} \tilde{\boldsymbol{D}}_j + \tilde{\boldsymbol{S}}_{j_0}, \tag{3.2}$$

where

$$\tilde{D}_{j,t} = \sum_{l=0}^{N-1} \tilde{u}_{j,l} \tilde{W}_{j,l+1 \bmod N},$$

$$\tilde{S}_{j,t} = \sum_{l=0}^{N-1} \tilde{v}_{j,l} \tilde{V}_{j,t+l \bmod N}, \tag{3.3}$$

$\tilde{u}_{j,l}$ and $\tilde{v}_{j,l}$ being the filters obtained by periodizing $\tilde{h}_{j,l}$ and $\tilde{g}_{j,l}$. According to eq. (3.2), 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, ..., 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.

3.3 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 $J_0 \le J$ 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 $J_0 \le log_2(N)$ 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. FORECASTING THROUGH WAVELET METHODOLOGY

Following Renaud *et al*. (2003), Azad *et al*. (2008), Aminghafari and Poggi (2007) and Aminghafari and Poggi (2012) for the Haar wavelet, reconstruction formula for $t = N + 1$ can be written as

$$X_{N+1} = \tilde{V}_{J,N+1} + \sum_{j=1}^{J} \tilde{W}_{j,N+1}. \tag{4.1}$$

Then, to predict X_{N+1}, it suffices to predict MODWT approximation and detail coefficients $\tilde{V}_{J,N+1}$ and $\tilde{W}_{j,N+1}$. Hence, idea is to predict for each scale unknown MODWT coefficients by a linear combination of their past values dyadically lagged starting from N:

$$\hat{\tilde{W}}_{j,N+1} = \sum_{k=1}^{A_j} a_{j,k} \hat{\tilde{W}}_{j,N-2^J(K-1)},$$

$$\hat{\tilde{V}}_{J,N+1} = \sum_{k=1}^{A_{J+1}} a_{J+1,k} \hat{\tilde{V}}_{j,N-2^J(K-1)}.$$

It may be noted that the past values appearing on the right hand sides of above equations must depend only on the past observations of the process itself. The explanatory variables are selected using dyadically lagged values of MODWT coefficients in order to extract coefficients of nonredundant (*i.e*. decimated) wavelet transform corresponding to dyadic grid adapted to (*i.e*. ending at) last observed value. Then, complete prediction equation of X_{N+1}, when N observations X_1, X_2, ..., X_N are given, is of the following form:

$$\hat{X}_{N+1} = \sum_{j=1}^{J} \sum_{k=1}^{A_j} a_{j,k} \hat{\tilde{W}}_{j,N-2^J(k-1)} + \sum_{k=1}^{A_{J+1}} a_{J+1,k} \hat{\tilde{V}}_{j,N-2^j(K-1)}$$

where $W = (\boldsymbol{w}_1, ..., \boldsymbol{w}_J, \boldsymbol{v}_J)$ represents Haar MODWT of $X(X = \tilde{V}_J + \sum_{j=1}^{J} \tilde{W}_j)$. For example choosing $A_j = 1$ for all resolution levels j, leads to the prediction

$$\hat{X}_{N+1} = a_j \hat{\tilde{W}}_{j,N} + a_{J+1} \hat{\tilde{V}}_{j,N}.$$

To further link this method with prediction based on a regular Autoregressive (AR) process, note that if on each scale the lagged coefficients follow an AR(A_j), addition of the prediction on each level would lead to same prediction formula (4.1). This multiresolution prediction model is actually linear. To estimate $Q = \sum_{j=1}^{J+1} A_j$ unknown parameters grouped in a vector α, normal equation $A'A\alpha = A'\boldsymbol{S}$ is to be solved, where

$$A' = (L_{N-1}, ..., L_{N-M}),$$

$$L_t' = \left[\tilde{w}_{1,t}, ..., \tilde{w}_{1,t-2A_1}, ..., \tilde{w}_{J,t}, ..., \tilde{w}_{J,t-2^J A_J}, ..., \tilde{v}_{J,t}, ..., \tilde{v}_{J,t-2^J A_{J+1}}\right],$$

$$a' = \left[a_{1,1}, ..., a_{1,A_1}, ..., a_{J,1} ..., a_{J,A_J}, ..., a_{J+1,1}, ..., a_{J+1,A_{J+1}}\right],$$

$$\boldsymbol{S}' = [X_N, ..., X_{t+1}, ..., X_{N-M+1}],$$

A is $Q \times M$ matrix, $\boldsymbol{\alpha}$ and $\boldsymbol{S}$ are respectively Q and M ($< Q$) size vectors.

5. RESULTS AND DISCUSSION

Indian monthly rainfall time-series data during the years 1871 to 2012, available at the website (*www.tropmet.res.in*) of the Indian Institute of Tropical Meteorology, Pune, India, is considered. Indian monsoon rainfall amount each year is calculated as the sum of monthly rainfalls from June to September of that year. Data during 1871 to 2001 comprising 131 data points are used for model building and data during 2002-2012 comprising remaining 11 data points are used for model validation. Computer programs for computation of MODWT and One-step ahead forecast through Wavelet method are respectively developed in R, Ver. 2.15.0 and SAS, Ver. 9.3 software packages and are appended as an Annexure.

5.1 Modelling of Rainfall Data by ARIMA Approach

In order to examine stationarity of the data, unit root test proposed by Dickey *et al*. (1979) is applied for parameter in the auxiliary regression

$$\Delta_1 y_t = \rho y_{t-1} + \alpha_1 \Delta_1 y_{t-1} + \varepsilon_t.$$

The relevant H_0: $\rho = 0$ against H_1: $\rho < 0$. The estimate of ρ is computed as –0.006. Calculated *t*-statistic is –0.58 which is greater than the critical value –1.95 of *t* at 5% level of significance (Franses 1998). Therefore, H_0 is rejected at 5% level and so $\rho < 0$, implying thereby that the data is stationary. The best ARIMA model, selected on the basis of minimum AIC and BIC values, is ARIMA(1,0,0) model. Parameter estimates along with corresponding standard errors of this model are reported in Table 1.

Table 1. Parameter estimates and other statistics.

Parameter	Estimate	Standard error	*t*-value	*p*-value
Constant	85.01	0.65	130.78	< 0.001
ARI	–0.12	0.08	–1.32	0.18

5.2 Modelling of Rainfall Data by Wavelet Approach

For computation of MODWT and forecasting of Indian monsoon rainfall by Wavelet approach, methodology discussed in Sections 3.2 and 4 is followed. 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 Fig. 1. Here, *X* denotes original time-series plot, *W1* to *W7* denote the wavelet details components, and *V7* denotes the smoothed component of MODWT. A perusal indicates that localized variation in the data is detected at lower scale, whereas global variation is detected at higher scale. Further, Multiresolution analysis (MRA) is also carried out on the basis of "Haar" wavelet and graph of the same is exhibited in Fig. 2. 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. Above mentioned pattern can also be verified from the MRA exhibited in Fig. 2.

The Root mean square error (RMSE) values for fitted models by ARIMA and Wavelet methods using observations during 1871 to 2001 are respectively computed as 81.75 and 62.60 respectively. Further, one-step ahead forecasts of Indian monsoon rainfall hold-out data for the years 2002 to 2012 by using the above two methods are also computed. The Root mean square prediction error (RMSPE), Mean absolute prediction

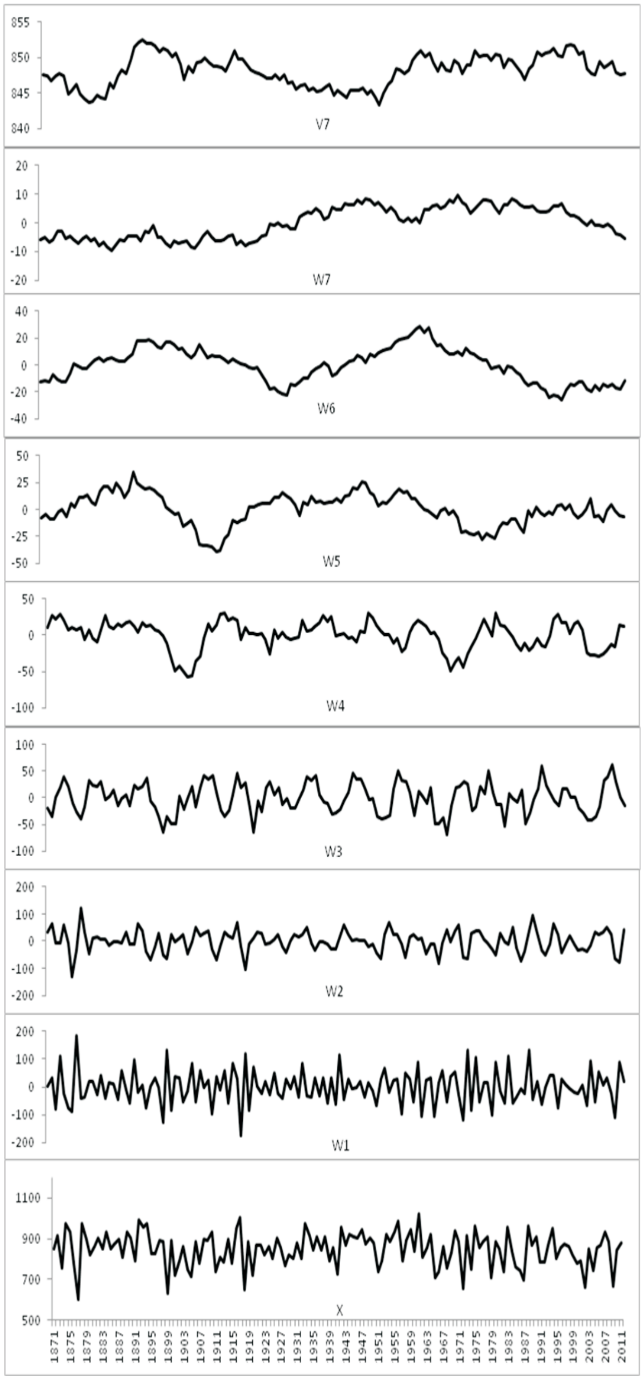

Fig. 1. MODWT of Indian monsoon rainfall time-series data.

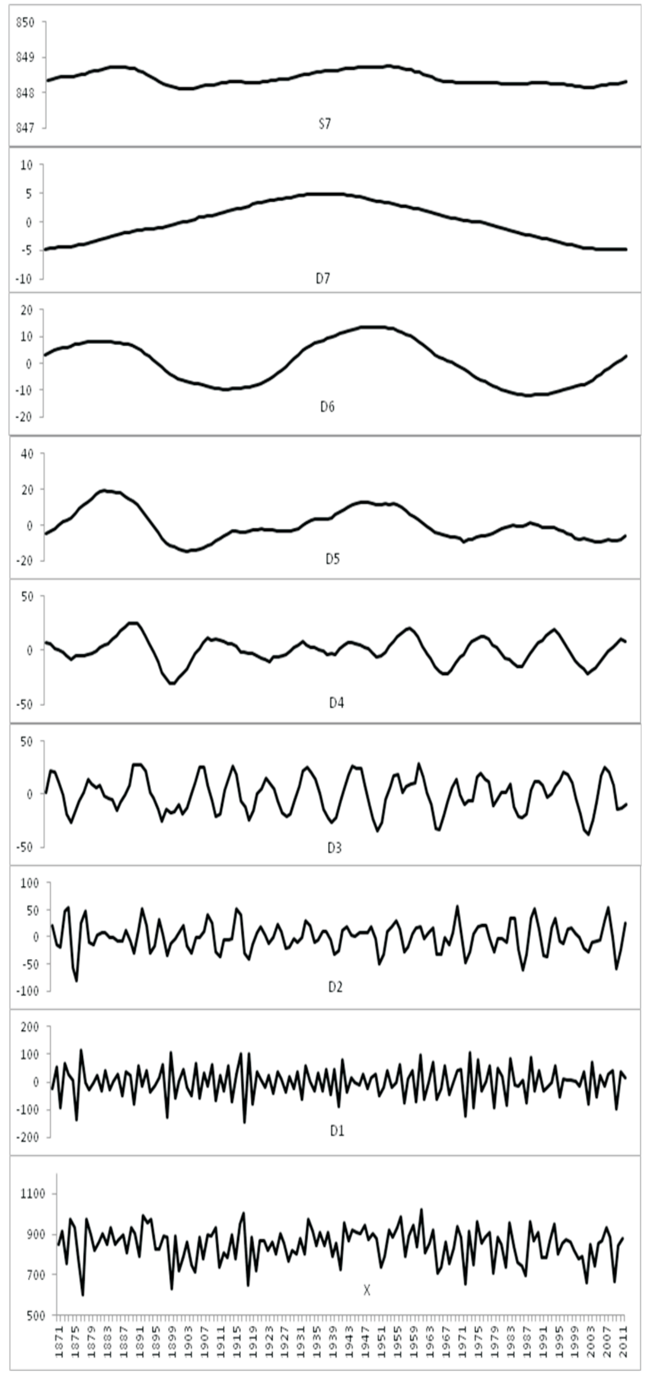

Fig. 2. MRA of Indian monsoon rainfall time-series data.

error (MAPE), and Relative mean absolute prediction error (RMAPE) values for fitted ARIMA model, viz. 90.61, 61.35, and 8.37 are found to be higher than the corresponding ones for fitted Wavelet method, viz. 67.79, 46.54, and 6.19. respectively, thereby reflecting superiority of Wavelet approach over ARIMA approach for prediction purposes also for the data under consideration. For validation of fitted models, more replicates with moving windows are also used. In the first window, observations during 1871-2001 are used for model building and those during 2002-2006 for prediction; in the second window observations during 1872-2002 are used for model building and those during 2003-2007 for prediction, and so on. Finally, data during 1876-2006 are employed for model building and those during 2008-2012 for prediction. The corresponding values of MAPE, RMAPE, and RMSPE are computed and are reported in Table 2. Again, it is

Table 2. One- step ahead prediction performance based on window prediction.

PredictionWindow	Wavelet			ARIMA		
	MAPE	RMAPE(%)	RMSPE	MAPE	RMAPE(%)	RMSPE
2002-2006	55.961	7.81	77.2455	64.176	9.25	97.120
2003-2007	42.716	5.12	54.167	43.299	5.36	61.234
2004-2008	47.567	5.90	57.802	50.145	6.01	63.299
2005-2009	60.964	7.91	82.240	65.567	8.78	92.070
2006-2010	62.250	8.12	81.640	65.522	8.74	92.068
2007-2011	59.292	7.70	80.965	67.722	8.91	92.647
2008-2012	50.448	6.91	74.209	55.295	7.60	85.264

concluded that Wavelet method outperforms the ARIMA method. In order to get a visual insight, fitted and predicted values of Indian monsoon rainfall by ARIMA and Wavelet methods along with data are exhibited in Fig. 3. A perusal indicates that latter is

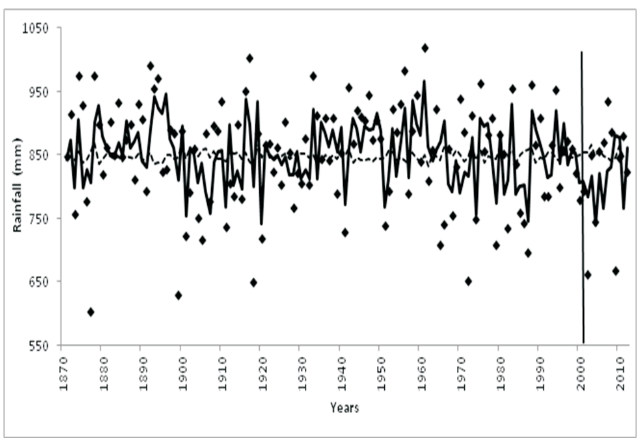

Fig. 3. Predicted values of Indian monsoon rainfall by Wavelet (Solid line) and ARIMA (Dashed line) methods along with the data points

better than former for both modelling and prediction purposes. However, it may be pointed out that, although Wavelet methodology is able to capture the pattern in monsoon rainfall unlike ARIMA method, yet it could not predict satisfactorily very low rainfalls during 2002, 2004, 2007 and 2009.

6. CONCLUDING REMARKS

The utility of nonparametric Wavelet methodology in frequency domain for modelling and forecasting purposes employing Haar wavelet is highlighted. Superiority of this approach over traditional Autoregressive integrated moving average methodology is demonstrated for Indian monsoon rainfall time-series data. It is hoped that research workers would start employing Wavelet analysis, which is a very promising and versatile methodology, for analyzing their data sets. The prediction of monsoon rainfall may be used by farmers for selecting appropriate crops to be sown in future years as well as by policy makers for proper planning, pertaining to exports and imports, fixing of minimum support prices, etc., particularly for rainfed crops. In this paper, computer program for computing one-step ahead predictions was written using SAS-IML. As future work, attempt is being made to write it in R to ensure wider usage. Further, there is a need to improve the one-step ahead prediction accuracy of Indian monsoon rainfall. It is believed that it depends on a number of predictors, like North Atlantic sea surface temperature, and North Pacific mean sea level pressure. In this article, the effect of these predictors on Indian monsoon rainfall was implicitly assumed but for making more efficient predictions, extension of the Wavelet and ARIMA methodologies capable of incorporating the effect of above predictors explicitly in the models need to be developed. Work is also in progress to study the utility of other wavelets, like Daubechies, Least asymmetric, and Symmlets and shall be reported separately in due course of time.

ACKNOWLEDGEMENTS

Authors are grateful to the Associate Editor and referees for their valuable comments.

REFERENCES

Almasri, A., Locking, H. and Shukur, G. (2008). Testing for climate warming in Sweden during 1850–1999, using wavelets analysis. *J. Appl. Stat.*, **35**, 431-443.

Aminghafari, M. and Poggi, J.M. (2012). Nonstationary time series forecasting using wavelets and kernel smoothing, *Comm. Statist. – Theory 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. *J. Ital. Statist. Soc.*, **6**, 97-144.

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.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (2007). *Time-Series Analysis: Forecasting and Control*. 3rd ed. Pearson Education, India.

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

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.

Ghosh, H., Paul, R.K. and Prajneshu (2010). Wavelet frequency domain approach for statistical modelling of rainfall time-series data. *J. Stat. Theo. Prac.*, **4**, 813-825.

Kulkarni, J.R. (2000). Wavelet analysis of the association between the southern oscillation and Indian summer monsoon. *Int. J. Clim.*, **20**, 89-104.

Nason, G.P. and von Sachs, R. (1999). Wavelet analysis in time-series analysis. *Phil. Trans. Roy. Soc. Lon.,* **357**, 2511-2526.

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

Paul, R.K., Prajneshu and Ghosh, H. (2011). Wavelet methodology for estimation of trend in Indian monsoon rainfall time-series data. *Ind. J. Agric. Sci.*, **81**, 96-98.

Percival, D.B. and Mofjeld, H.O. (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 Univ. Press, U.K.

Rajeevan, M., Pai, D.S., Dikshit, S.K. and Kelkar, R.R. (2004). IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003. *Current Sci.*, **86**, 422-431.

Renaud, O., Stark, J.L. and Murtagh, F. (2003). Prediction based on a multiscale decomposition. *Int. J. Wave. Mult. Inf. Proc.*, **1**, 217-232.

Sunilkumar, G. and Prajneshu (2004). Modelling and forecasting meteorological subdivisions rainfall data using wavelet thresholding approach. *Cal. Stat. Assoc. Bull.*, **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.

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

ANNEXURE

(i) R PROGRAM FOR COMPUTING MODWT

```
library(wavelets)
library(wavethresh)
mydata <-read.table("E:/Project_IASRI/
Monsoon rain/monsoon.txt",header=TRUE)
y<-mydata["monsoon"]
# computing MODWT and MRA
modwt<-modwt(y, filter="haar",
n.levels=7,boundary="periodic", fast=TRUE)
mra<-mra(y, filter="haar", n.levels=7,
boundary="periodic", fast=TRUE,
method="modwt")
#extracting MODWT coefiicient vector
w1<-modwt@W[[1]]
w2<-modwt@W[[2]]
w3<-modwt@W[[3]]
w4<-modwt@W[[4]]
w5<-modwt@W[[5]]
w6<-modwt@W[[6]]
w7<-modwt@W[[7]]
v7<-modwt@V[[7]]
#extracting MRA coefiicient vector
d1<-mra@D[[1]]
d2<-mra@D[[2]]
d3<-mra@D[[3]]
d4<-mra@D[[4]]
d5<-mra@D[[5]]
d6<-mra@D[[6]]
d7<-mra@D[[7]]
s7<-mra@S[[7]]
```

(ii) COMPUTING ONE-STEP AHEAD FORECAST BY USING MODWT

```
data wavelet;
input w1 w2 w3 w4 w5 w6 w7 v7 x;
/* w1-w7 are wavelet coefficients and v7
contains smooth coefficients*/
/* x is original time series*/
cards;
;
proc iml;
use wavelet;
read all into w;
close wavelet;
n=nrow(w);
x=w[,ncol(w)];
level=7; /* specifying the level of
decompositions*/
rj=2;
/* computing one-step ahead forecast*/
%macro forecast;
%do no_matrix = 1 %to level;
c&no_matrix. = j(rj,1,0);
a&no_matrix. = j(rj,j,0);
  do i= 1 to j;
        do k= 1 to rj;
              a&no_matrix.[k,i]=w[(n-
              &no_matrix.)-2**i*(k-1),i];
              c&no_matrix.[k]=w[(n-
              &no_matrix.)-2**j*(k-1),j+1];
        end;
  end;
l&no_matrix.=a&no_matrix.[,1];
do i = 2 to level;
l&no_matrix.=l&no_matrix.//
a&no_matrix.[,i];
end;
l&no_matrix.=l&no_matrix.//c&no_matrix.;
%if &no_matrix. ^= 1 %then %do;
    a = l1||l&no_matrix.;
    l1 = a;
%end;
%end;
%mend;
%forecast;
beta=j(10,1,0);
s=j(level,1,0);
do i = 1 to level;
        s[i]=x[n-(i-1)];
  end;
run;
C_=j(rj,1,0);
A_=j(rj,level,0);
  do k= 1 to rj;
        do i= 1 to level;
          A_[k,i]=w[(n)-2**i*(k-1),i];
        end;
          C_[k]=w[(n)-2**J*(k-), level+1];
  end;
L_=a_[,1];
do i = 2 to level;
L_=L_//a_[,i];
end;
L_=L_//C_;
beta=inv(a*a')*a*s;
xn1=beta'*L_; /* xn1 denotes one-step ahead
forecast*/
print xn1;
run;
quit;
```