Indian Journal of Agricultural Sciences **81** (3): 290–2, March 2011

Wavelet methodology for estimation of trend in Indian monsoon rainfall time-series data*

RANJIT KUMAR PAUL[1], HIMADRI GHOSH[2] and PRAJNESHU[3]

Indian Agricultural Statistics Research Institute, New Delhi 110 012

Received: 1 July 2009; Revised accepted: 5 January 2011

Key words: ARIMA, Discrete wavelet transform, Indian monsoon rainfall data, Trend, Wavelet methodology in frequency domain

India's agricultural performance depends to a large extent on the quantum and distribution of rainfall during monsoon months. So, information on its various aspects is vital for planning and policy purposes. Accordingly, several attempts have been made in the past to develop models for describing rainfall data. A review of multiple and power regression models employed in India since 1988 along with various modifications made in these models from time to time has been provided by Rajeevan *et al.* (2003). Recently, Aggarwal (2008) has highlighted that global climatic changes and increasing climatic variability are likely to exert pressure on agricultural systems and may constrain attainment of future food production targets. The main purpose of this paper is to study the trend in Indian monsoon rainfall data during the period 1879 to 2006 using Wavelet methodology.

Wavelets are fundamental building block functions, analogous to the trigonometric sine and cosine functions to represent a function in the form of wavelet transform, which preserves all its information on a time scale basis. As with a sine or cosine wave, a wavelet function oscillates and wiggles out to zero. A variety of wavelet and wavelet-related methods have been developed in recent years (Percival and Walden 2000) with potential applications to an increasing range of statistical problems. Wavelet analysis can be studied in two ways: one is in time domain and another is the "frequency domain". In respect of the former, Kulkarni (2000) applied wavelet analysis for investigating association between the southern oscillation and Indian summer monsoon. Further Sunilkumar and Prajneshu (2004) carried out modelling and forecasting of annual rainfall data of meteorological subdivision of eastern Uttar Pradesh using wavelet thresholding approach. However, in respect of Wavelet analysis in frequency domain, no applied work for Indian agricultural data seems to have been reported.

In order to carry out wavelet analysis in frequency domain for data observed at discrete time-epochs, the concept of discrete wavelet transform (DWT) is employed. DWT of a time-series observation $\left\{ X_t, t = 0,1,..., N-1 \right\}$, where $N = 2^J$ used to capture high and low frequency components of the data using wavelet filter and scaling filter, respectively. This, in turn, would enable modelling of time-series data through computation of inverse DWT. For a DWT based on an orthonormal matrix P whose rows are of even order, shifted wavelet filters may be constructed. In discrete wavelet analysis, the filter $\{h_i\}$ should satisfy the following three basic properties for orthonormality:

$$\sum_{i=0}^{L-1} h_i = 0, \quad \sum_{i=0}^{L-1} h_i^2 = 1 \text{ and } \sum_{i=0}^{L-1} h_i h_{i+2n} = \sum_{i=-\infty}^{\infty} h_i h_{i+2n} = 0 \quad ...(1)$$

The first stage wavelet filters are high pass filters with nominal pass band [1/4, 1/2]. This, in turn, implies that the associated scaling filters $\{g_i\}$ are low pass filters and the jth level wavelet coefficient can be associated with pass band $[1/2^{j+1}, 1/2^j]$. Computation of DWT is carried out by employing the Pyramid algorithm, which is discussed below:

The first stage for computing the DWT simply consists of transforming the time-series $\mathbf{X}$ of length $N = 2^J$ into the N/2 first level wavelet coefficients $\mathbf{W}_1$ and the N/2 first level scaling coefficients $\mathbf{V}_1$. Precisely, to obtain unit scale wavelet coefficients, time-series $\{X_t: t = 0,1, …,N-1\}$ is circularly filtered with filter h_i, i = 0,1, …,L-1, where L, an even integer, denotes width of the filter and must satisfy equation (1). Compute

*Short note

A part of the Ph D thesis of the first author submitted to the Indian Agricultural Research Institute, New Delhi in 2008 (unpublished).

[1]Scientist (e mail: ranjitstat@gmail.com), Fishery Resource and Environmental Monitoring Division, Central Inland Fisheries Research Institute, Barrackpore 743 101;

[2]Senior Scientist (e mail: hghosh@gmail.com), Division of Biometrics and Statistical Modelling; [3]Principal Scientist and Head (e mail: prajneshu@yahoo.co.in), Division of Biometrics and Statistical Modelling

$$2^{1/2}\tilde{W}_{1,t} = \sum_{i=0}^{L-1} h_i X_{(t-i)\ \mathrm{mod}N},\ t = 0,1,...,N-1 \qquad \ldots(2)$$

Now define N/2 wavelet transforms for unit scale as

$$W_{1,t} = 2^{1/2}\,\tilde{W}_{1,2t+1} = \sum_{i=0}^{L-1} h_i X_{(2t+1-i)\ \mathrm{mod}N},\ t = 0,1,...,\ N/2 - 1 \qquad \ldots(3)$$

The above procedure is called "Downsampling" procedure. To obtain first stage scaling coefficients, define "quadrature mirror" scaling filter $g_i = (-1)^{i+1} h_{L-1-i}$. Then the first level scaling coefficients are:

$$V_{1,t} = 2^{1/2}\tilde{V}_{1,2t+1} = \sum_{i=0}^{L-1} g_i X_{(2t+1-i)\ \mathrm{mod}N} \qquad \ldots(4)$$

The second stage of Pyramid algorithm consists of treating $\{V_{1,t}\}$ in the same way as $\{X_t\}$was treated in the first stage. Then we circularly filter$\{V_{1,t}\}$ separately with $\{h_i\}$; and $\{g_i\}$ and subsample to produce two new series for t=0,1,..., N/4–1, namely

$$W_{2,t} = \sum_{i=0}^{L-1} h_i V_{1,(2t+1-i)\mathrm{mod}\ N/2} \qquad \ldots(5)$$

$$V_{2,t} = \sum_{i=0}^{L-1} g_i V_{1,(2t+1-i)\mathrm{mod}N/2} \qquad \ldots(6)$$

These procedures are repeated J times to obtain 2^J DWT's. At the jth stage, the elements of $\mathbf{V}_{j-1}$ are filtered separately with wavelet filter$\{h_i\}$, and scaling filter$\{g_i\}$. The filter outputs are subsampled to form respectively $\mathbf{W}_j$ and $\mathbf{V}_j$. The elements of $\mathbf{V}_j$ with $N/2^j$ elements are called the scaling coefficients for level j, while those of $\mathbf{W}_j$ with $N/2^j$ elements contain the desired wavelet coefficients for level j. At the end of Jth stage, the DWT coefficients **W** with 2^J elements are formed by concatenating the J+1 vectors.

Let P be an $N \times N$ real valued matrix used for obtaining the above DWT, which satisfies orthonormality property $P'P = I_N$, where I_N is the N×N identity matrix and superscript indicates transpose. Then the DWT (**W**) of the time-series **X** may be computed as **W** = P**X**. Now the elements of **W** are decomposed into J+1 subvectors. The first J subvectors contain all the DWT coefficients for scale τ_j. Then **W** can be written as

$$\mathbf{W} = \left[\mathbf{W}_1' \ \ \mathbf{W}_2' \ \ \ldots \mathbf{W}_J' \ \ \mathbf{V}_J'\right] \qquad \ldots(7)$$

It is desirable to decompose a time-series into different components of variations, like low frequency (trend), and high frequency (noise) components. To this end, multiresolution analysis is used on the data on a scale-by-scale basis. Consider the following model for a time-series data $\{X_t\}$:

$$X_t = \mu + T_t + Z_t,\ t = 0,1,...,N-1, \qquad \ldots(8)$$

where μ is a constant term, T_t is an unknown deterministic polynomial trend function of order r, and Z_t is the residual term.

Now, since $\mathbf{W} = \left[\mathbf{W}_1' \ \ \mathbf{W}_2' \ \ \ldots \mathbf{W}_J' \ \ \mathbf{V}_J'\right]'$, the vector **W** can be written as sum of two vectors: $\mathbf{W} = \mathbf{W}_w + \mathbf{W}_s$, where $\mathbf{W}_w$ is an N × 1 vector containing the wavelet coefficients and zeros at all other locations, and $\mathbf{W}_s$ is an N × 1 vector containing the scaling coefficients and zeros at all other locations. From the wavelet coefficients, original function can be reconstructed by using Inverse DWT given by **X** = P′**W**, which can be written as:

$$\mathbf{X} = P'\mathbf{W} = P'\mathbf{W}_s + P'\mathbf{W}_w = \hat{T} + \hat{Z}, \qquad \ldots(9)$$

where $\hat{T} = P'\mathbf{W}_s$ is an estimator of the polynomial trend T at level J, while $\hat{Z} = P'\mathbf{W}_w$ is the estimator of residual Z. The issue of choosing the level k, k = 1, 2,…,J of estimate of trend depends on the goal of application. It should be small for detecting the local trends and cycles while it is set large for detecting the global trend.

Monthly rainfall data of India during the years 1879 to 2006, available at the website (*www.tropmet.res.in*) of the Indian Institute of Tropical Meteorology, Pune, India, is considered. From this data, India's monsoon rainfall data for each year is obtained by adding monthly rainfall figures during the four months, ie June till September. The SPLUS WAVELET TOOLKIT software package is used for carrying out wavelet analysis. The DWT coefficients for Haar wavelets are computed. The wavelet coefficients are related to differences (of various orders) of weighted average values of portions of X_t concentrated at time t. Wavelet coefficients have been observed as bars, up or down according to the sign of the coefficients. 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. Coefficients at the top (below) are "high frequency" ("low frequency") information. Wavelet coefficients do not remain constant over time and reflect changes in the data at various time-epochs. The locations of abrupt jumps can be spotted by looking for vertical (between levels) clustering of relatively large coefficients. It may be noted that prominent short-scale fluctuations during 1889–1900 and 1915–18 are represented by the spikes of wavelet coefficients with smaller scale τ_j. It is observed that the large wavelet coefficient occurs at the end of the wavelet coefficients corresponding to scale τ_6, indicating thereby the presence of decreasing trend. The estimate of trend of rainfall data is computed by multiresolution analysis for various scales using Haar wavelets. As the scale increases, the declining global trend present in the data set becomes more and more prominent. For visual display, graph of estimate of trend of rainfall data for scale 6 is exhibited in Fig 1.

Subsequently, trend in the data is estimated through Box-Jenkins Autoregressive moving average (ARIMA) methodology (Box *et al*. 2007) using the SAS software package. Assuming presence of deterministic

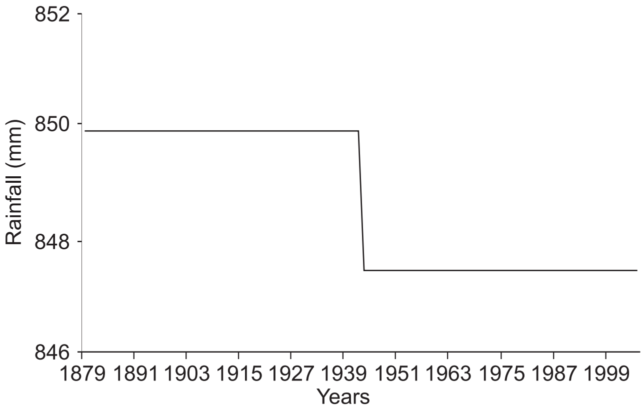

Fig 1 Estimate of trend by Haar wavelet for scale 6

linear trend in the rainfall series, following Autoregressive moving average with exogenous variable (ARMAX) model is fitted:

$$Y_t = \quad + \quad t + (1- \quad B)/(1- \quad B)\,\varepsilon_t\ ,\ t=1,2,\ldots,T \qquad \ldots(10)$$

where ε_t is error term. By using conditional least squares estimation procedure available in SAS software package, it is found that the regression coefficient δ has p-value 0.20, and so trend in the data is not significant at 5% level. Thus, the ARMAX model is not able to pick up any declining trend. Finally, the fitted trend model is

$$T_t = \hat{} + \hat{}\ t + \hat{}\ z_{t-1} - \hat{}\ \hat{\varepsilon}_{t-1}$$

where $z_t = Y_t - \hat{} - \hat{\delta}\, t$, $\hat{\varepsilon}_t = Y_t - T_t$ and $\hat{}, \hat{}, \hat{}, \hat{\theta}$ are 848.62, –0.23, –0.37, –0.26, respectively.

SUMMARY

Trend in Indian monsoon rainfall data during 1879–2006 was modelled through Wavelet analysis in frequency domain. A significant result obtained was that the data showed a declining trend. This is a matter of great concern and may have serious implications from global warming point of view, which in turn, would affect India's agricultural production. It may be emphasized that the traditional Box Jenkins ARIMA methodology could not capture the declining trend present in the data.

REFERENCES

Aggarwal P K. 2008. Global climate change and Indian agriculture: impacts, adaptation and mitigation. *Indian Journal of Agricultural Sciences* **78**: 911–9.

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

Kulkarni J R. 2000. Wavelet analysis of the association between the southern oscillation and Indian summer monsoon. *International Journal of Climatology* **20**: 89–104.

Percival D B and Walden A T. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge Univ. Press, UK

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 Science* **86**: 422–31.

Sunilkumar G and Prajneshu. 2004. Modelling and forecasting meteorological subdivisions rainfall data using wavelet thresholding approach. *Calcutta Statistical Association Bulletin* **54**: 255–68.