*Indian Journal of Agricultural Sciences* **91** (2): 227–31, February 2021/Article

# Wavelet based long memory model for modelling wheat price in India

RANJIT KUMAR PAUL[1]*, SANDIPAN SARKAR[1] and SATISH KUMAR YADAV[2]

*ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012, India*

Received: 12 February 2020; Accepted: 09 October 2020

ABSTRACT

Agricultural time-series data concerning production, prices, export and import of several agricultural commodities is published by Indian government along with other private agricultural sectors every year. The analysis of these factors is necessary to formulate and apply several policies regarding food acquisition and its distribution, quality and quantity of import and export products, pricing structure, MSP of agricultural commodities etc. Box – Jenkins's Autoregressive integrated moving average (ARIMA) model is broadly utilized in the field of time-series. In the field of time-series analysis, it is assumed by most of the researchers that the data points of different time lags do not depend on each other, i.e. absence of long memory process. But in agriculture, market price data exhibits that the observation are dependent on distant past. This is the possible indication of long memory process or long range dependency in the mean model. Autoregressive fractionally integrated autoregressive moving average (ARFIMA) model is generally used to portray the characteristic features of the long memory time series models as well as for the forecasting purposes. In this study wavelet decomposition is used for increasing the forecasting accuracy of the ARFIMA model. Daily wholesale data of wheat of Rewari market of Haryana for the period of January, 2010 to November, 2017 is used for the demonstration of our approach.

**Keywords**: ARIMA, ARFIMA, Long memory, Wavelet analysis, Wheat price

Over last several decades Box Jenkin's Autoregressive integrated moving average (ARIMA) methodology (Box *et al*. 2007) is used for forecasting time series data. An ARIMA model can be specified with three parameters, viz. $p$, $d$, $q$ and it only allows the value of $d$ to be integer. There are certain circumstances where it is not possible to postulate the integer value of $d$, i.e the time series model possess long memory property. A time-series process is called as long memory or Fractional differenced (FD) (Beran 1995) process if the autocorrelation function decays very slowly towards zero unlike the exponential decay in usual ARIMA model. A time-series process exhibiting long memory structure can well be modelled with the help of autoregressive fractionally integrated autoregressive moving average (ARFIMA) model (Granger and Joyeux 1980, Hosking 1981). An ARFIMA ($p$, $d$, $q$) process $\{y_t\}$ may be defined as;

$$\varphi(L)\ (1 - L)^d y_t = \theta\ (L) e_t \qquad (1)$$

where, $\varphi(L) = 1 - \varphi_1 L - ... - \varphi_1 L^p$, $\theta(L) = 1 - \theta_1 L - .. - \theta_q L^q$, are respectively the AR and MA operators, sharing no common roots, $(1 - L)^d$ is the fractional differencing operator and $e_t$ are assumed to be independent and identically distributed (i.i.d) with zero mean and variance $\sigma^2$. Some of the applications of ARFIMA model can be found in Paul (2014, 2017), Paul *et al.* (2015), Paul and Mitra (2020).

Time series data consists of signal and noise part; it is very difficult to separate these two parts from the noisy series. Wavelet decomposition tires to extract the signal from the noisy data. A wavelet (Antoniadis 1997) is a mathematical function which can be used to transform a given function or time-series into different time dependent scales. Wavelet methods provide the dynamics of financial time series unlike usual time series analysis (Renaud *et al.* 2003). Soltani *et al.* (2000) suggested that after wavelet decomposition of the long memory process only the signal part contains the long memory property. In this context an attempt has been made to improve the forecasting ability of ARFIMA model using wavelet decomposition.

## MATERIALS AND METHODS

*Long memory process*: Most of financial time-series analysis assumes that the observations are free from long time span and also independent of each other or nearly so. But in practical situation it has been observed that observations of a financial time-series are dependent on distant past (Paul *et al.* 2014). The statistical dependency of a time-series can be estimated by plotting ACF of the data points. Let $y_t$; ($t = 0, 1, 2,...$) be a stationary time-series process and the autocorrelation function of the time-series

Present address: [1]ICAR-Indian Agricultural Statistics Research Institute, New Delhi; [2]ICAR-National Centre for Integrated Pest Management, New Delhi. *Corresponding author email: ranjitstat@gmail.com.

with a time lag of $k$ is given as;

$$\rho_k = cov\,(y_t, y_{t-k})/var\,(y_t) \quad (2)$$

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

$$W_{j.t} \equiv \sum_{l=0}^{Lj-1} h_{j,l} X_{t-l \bmod N}$$

For a stationary long memory process fractional differencing parameter $d$ lies between o to ½. There are various approaches for estimating $d$, in the present investigation GPH (Geweke and Porter 1983) test and sperio test have been used.

*Wavelet analysis*: Wavelets are the mathematical functions, resembling to the sine and cosine function. A variety of wavelet and wavelet-based methodologies have been developed in the recent decades (Percival and Walden 2000) with potentials applications in various fields. The fluctuating property of wavelet makes the function a wave. There are two types of wavelet decomposition techniques which are widely used in statistical analysis, viz. Continuous wavelet transformation (CWT) and Discrete wavelet transformation (DWT) (Aminghafari and Poggi 2007). CWT is utilized to manage time-series defined on real axis. The DWT handles the series characterized by integers, i.e. univariate time-series data. DWT of a time-series observation is operated to capture the high and low frequency components. In DWT the number of observations must be in the form of $2^N$, where $N$ is an integer. To overcome this situation maximal overlap discrete wavelet transform (MODWT) can be used (Paul *et al.* 2013).

*Maximal overlap Discrete Wavelet Transform (MODWT)*: The MODWT is a linear filtering operation that decomposes a series into coefficients analogous 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. MODWT is all around characterized for all example sizes $N$, while for a total decay of $J$ levels. But DWT requires $N$ in the form of $2^J$, where J is any positive integer (Ogden 1997). MODWT also differs from DWT in the sense that it is a highly redundant, non-orthogonal transform. MODWT coefficients are obtained by applying DWT pyramid algorithm once to X and another to the circularly shifted vector TX. Hence, the first application yields the usual DWT (W) of the time-series vector X computed as W=P (Daubechies 1992) and the second application consists of substituting TX for X obtained as,

$$W = PTX \quad (3)$$

where, W and P can be written as $W= [W_1 W_2 \ldots W_J V_J]$ $P=[P_1 P_2 \ldots P_J Q_J]$. The Mallat algorithm (Mallat 1989) filters the original data series $X = (X_0, X_1, \ldots, X_{N-1})$ using a pair of high-pass and low-pass filters denoted, respectively, as $h = (h_0, h_1, \ldots, h_{L-1})$ and $g = (g_0, g_1, \ldots, g_{L-1})$, each of length L, L < N. The wavelet (Wj) and the scaling coefficients (***C****j*) corresponding to the *j*th level of decomposition, $j = 1, 2, \ldots, J$, $J$ is an positive integer, are obtained by,

$$W_{j.t} \equiv \sum_{l=0}^{L_{j-1}} h_{j,l} X_{t-l \bmod N} \text{ and } C_{j,t} \equiv \sum_{l=0}^{L_{j-1}} g_{j,l} X_{t-l \bmod N}$$

$$X_t = \sum_{j=1}^{J} W_{j,l} + C_{j,t} \quad (4)$$

where $h_{j,l}$ is the $j^{th}$ level MODWT wavelet filter and $g_{j,l}$ is the $j^{th}$ level MODWT scaling filter. A time-series can be totally or partly decomposed (Aminghafari and Poggi 2007) into a number of levels $J_o \le \log^N_2$.

*Wavelet-based prediction*: First of all, time-series process is tested whether the long memory structure exists or not using any of the GPH, semi-parametric method, wavelet method, etc. Then the time-series is decomposed using wavelet transformation (Aminghafari and Poggi 2012) and obtained wavelets and scaling coefficients. Again long memory test is applied in each set of coefficients. It is found that only the scaling coefficient exhibits long memory structure and the wavelet series showed short memory structure. Then ARIMA and ARFIMA models are fitted accordingly to obtain the forecast values of the underlying time-series process.

In wavelet transformed long-memory time-series, only the scaling coefficient has the long memory property (Soltani *et al*. 2000). This smooth coefficient is the signal component of a time-series model. So after wavelet decomposition, long-memory test is performed in each decomposed series. Then ARIMA and ARFIMA model should be fitted according to the long-memory test.

It has been often seen that the model which is best fitted within the sample data, may not be up to the mark for the out-of-sample forecasting. Hence it is recommended that to select a parsimonious model based on the 90% sample observations and remaining 10% data points should be kept for checking the model accuracy. In this study Mean absolute percentage error (MAPE) and Root mean square error (RMSE) have been used for comporting the accuracy of models.

$MAPE = 100 * \frac{1}{k}\sum_{t=1}^{k} \left|\frac{r_t}{x_t}\right|$, where, $r_t = x_t - \hat{x}_t$ and k is the number of observations.

$RMSE = \sqrt{\frac{1}{k}\sum_{t=1}^{k} r_t^2}$, where, $r_t$ and k are defined as above.

## RESULTS AND DISCUSSIONS

*Data description*: For the present investigation, daily wheat price data is collected from AGMARKNET website for the period January, 2010 to November, 2017 considering Rewari market of Haryana, India. The data is composed of maximum, minimum and model prices of wheat with 1030 data points. In this study, first 900 data points are used for model identification and parameter estimation purpose and rest of 130 data points are used for model validation purpose.

*Descriptive statistics and Test for stationarity*: The average maximum price is higher than the minimum and

Table 1 Descriptive statistics and testing stationarity of Wheat price data

| Series | Min | Max | Mean | St. Dev. | CV (%) | Skewness | Kurtosis |
|---|---|---|---|---|---|---|---|
| Minimum price (₹/q) | 925 | 1950 | 1294 | 218.16 | 16.85 | 0.42 | 2.47 |
| Maximum price (₹/q) | 1000 | 2160 | 1319 | 216.007 | 16.37 | 0.43 | 2.47 |
| Modal price (₹/q) | 970 | 2040 | 1307 | 216.09 | 16.53 | 0.45 | 2.49 |
| Testing stationarity | | | | | | | |
| *Series* | *PP test* | | *ADF test* | | | | |
| | *Test statistic* | *p-value* | *Test statistic* | *p-value* | | | |
| Minimum | -39.512* | <0.01 | -3.604* | 0.032 | | | |
| Maximum | -34.614* | <0.01 | -3.624* | 0.033 | | | |
| Modal | -30.831* | <0.01 | -3.602* | 0.031 | | | |

*denotes the significance at 5% level

the modal prices of wheat (Table 1). Higher value of coefficient of variation in minimum price data indicates the higher variability as compared to other price series. All the series under consideration are positively skewed and platykurtic in nature.

PP (Phillips and Perron) test and ADF (Augmented Dickey Fuller) test are employed to see the presence of unit root in the data set. It is clear from the results (Table 1) of both the tests that the null hypothesis of presence of unit root is rejected at 5% level of significance indicating

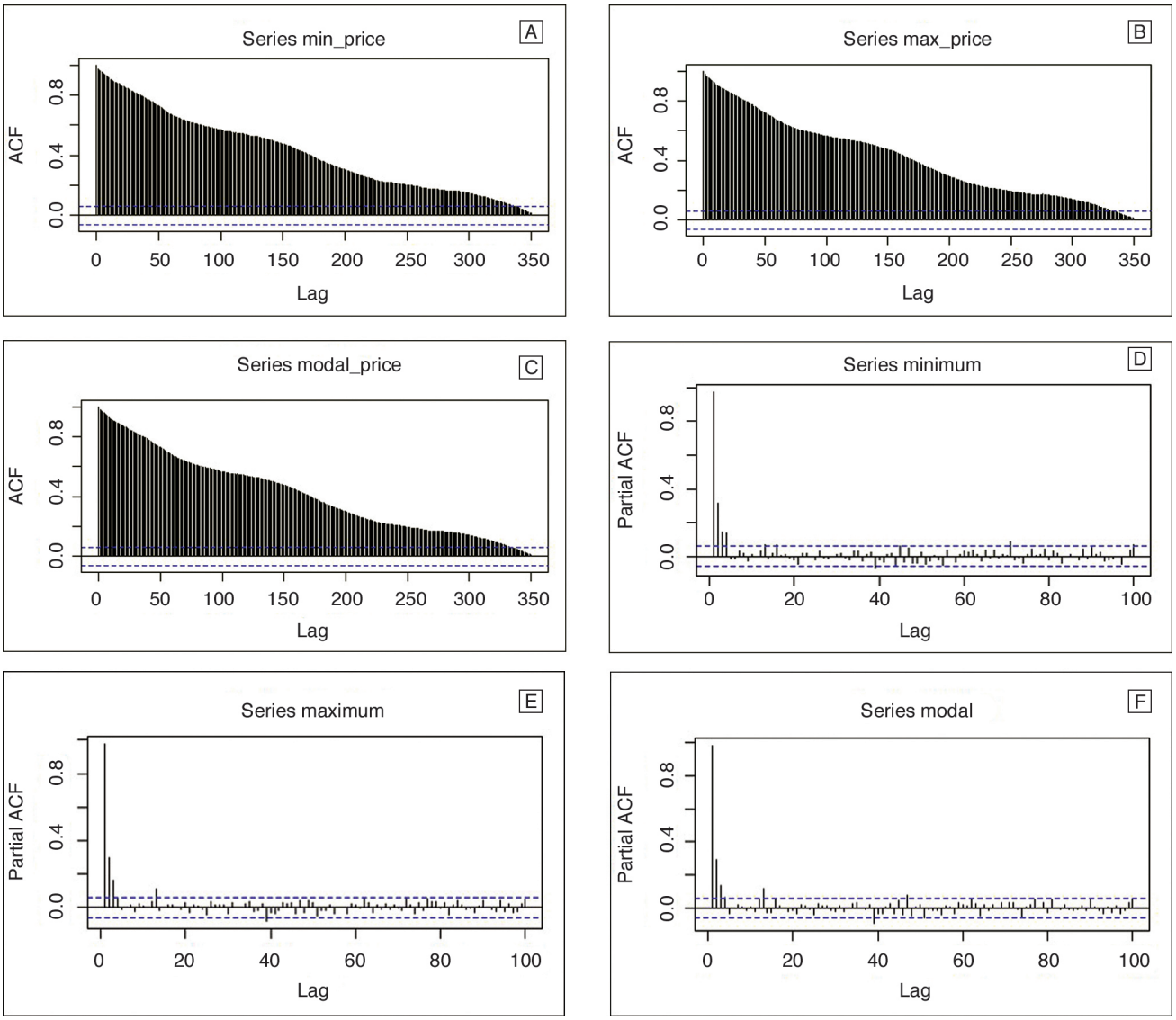


Fig 1 ACF (A-C) and PACF (D-F) plots of Wheat price data.

stationarity of the series.

*ACF and PACF plots*: The Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the original price series are studied to investigate the distributional pattern of the series. The ACF plots of the original series shows (Fig 1 (A-C)) that the autocorrelation functions are decaying very slowly towards zero. Fig 1 (D-F) also portraits the PACF plots of different series under consideration. It is evident that PACF are significant up to 100 lags. Hence various plots are indicating the possible presence of long memory.

*Test for long memory and Fitting of ARFIMA model*: After investigating the ACF plot, the long memory test is conducted to the data set (Table 2). Since the calculated t-value is greater than 1.96 for all the series in case of Sperio and GPH test, the tests is found to be significant. It establishes the presence of long range dependency in price data set.

In this section, on the basis of log likelihood value and AIC value, best fitted ARFIMA model is selected for all series. For minimum series, best fitted model is ARFIMA (2, *d*, 0), accordingly, parameter estimates along with corresponding standard error and P-values are presented in the Table 2. Likewise for maximum series, ARFIMA (2, *d*, 0) and for modal series ARFIMA (1, *d*, 0) has been identified. After that, residuals of best fitted models are investigated. Residuals are found out to be white noise process.

*Wavelet decomposition*: MODWT is carried out on the basis of "Haar" wavelet filter at level 5 (Fig 2). Wavelet coefficients are plotted as line, up or down. The number of wavelet coefficients at the lowest resolution level (level

Table 2 Long memory parameter and ARFIMA parameter estimate of Wheat price data

| Series | Sperio test | | | GPH | | |
|---|---|---|---|---|---|---|
| | d | S.E | t | d | S.E | Z |
| Minimum | 0.38* | 0.051 | 7.6 | 0.35* | 0.131 | 2.69 |
| Maximum | 0.41* | 0.049 | 8.39 | 0.43* | 0.067 | 6.38 |
| Modal | 0.47* | 0.045 | 10.39 | 0.44* | 0.068 | 6.46 |
| *ARFIMA parameter estimates* | | | | | | |
| *Parameters* | *Estimate* | | *St. Error* | | *p-Value* | |
| *Minimum Series* | | | | | | |
| d | 0.381* | | 0.051 | | <0.01 | |
| ar1 | 0.611* | | 0.047 | | <0.01 | |
| ar2 | 0.372* | | 0.069 | | <0.01 | |
| *Maximum Series* | | | | | | |
| d | 0.412* | | 0.048 | | <0.01 | |
| ar1 | 0.615* | | 0.052 | | <0.01 | |
| ar2 | 0.369* | | 0.061 | | <0.01 | |
| *Modal Series* | | | | | | |
| d | 0.471* | | 0.045 | | <0.01 | |
| ar1 | 0.979* | | 0.078 | | <0.01 | |

* denotes the significance at 5% level

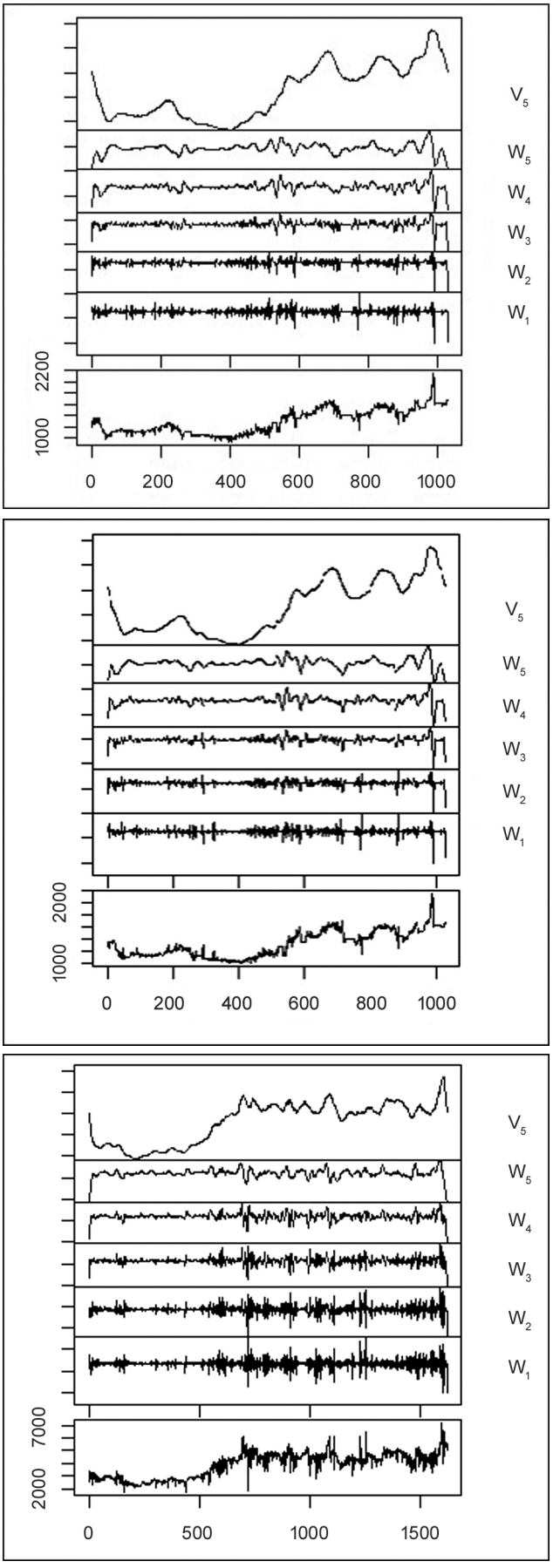


Fig 2 MODWT plots of Wheat price data.

= 1) is exactly half the number of original data points and the number of coefficients decreases by half at each level. The coefficients $W_1$, $W_2$, $W_3$, $W_4$, $W_5$ and $V_5$ are calculated using "Haar" filter. 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. This smooth coefficient is the actual signal, hidden in the noisy time series data.

*Long memory test of Wavelet coefficients*: For investigating the presence of long memory property, sperio test has been performed to all the coefficients of MODWT. The analysis reveals that only smooth coefficient namely $V_5$ has the significant long memory property at 5% level, while rest of the wavelet coefficients possess the short memory property. Keeping this result in view ARFIMA model has been fitted in the $V_5$ series and ARIMA model in the rest of the decomposed series.

*Validation of the fitted model*: Last 130 observations of the data set were kept for the validation of the models. After decomposing the series only scaling coefficients ($V_5$) has the long memory property. So, ARFIMA model is fitted in $V_5$ series and ARIMA model is fitted in the rest of the series. After computing the forecasted values of respective coefficients, i.e. $W_1$, $W_2$, $W_3$, $W_4$, $W_5$ and $V_5$, inverse wavelet transformation is applied to get the actual forecast values.

A comparative performance of the results of wavelet approach along with ARFIMA model has been carried out in terms of MAPE and Root mean square error (RMSE). For wavelet method, the MAPE and RMSE values for maximum, minimum and modal price are found out to be (17.73, 103.89); (16.83, 264.14); (17.36, 268.83). Similarly, for ARFIMA model, the MAPE and RMSE values for maximum, minimum and modal price are found out to be (20.54, 118.19); (19.55, 299.47); (22.18, 330.49). It is evident from above figures that wavelet method performs better than the usual ARFIMA model.

The analysis reveals that in wavelet decomposition of the long memory series, only the smoothest part will contain the long memory property and the rest of the series will have short memory. It clearly indicates the outperformance of wavelet approach over usual ARFIMA model. For all the series, forecasting performance of wavelet approach is more efficient than the usual ARFIMA model in terms of lower MAPE and RMSE values.

## REFERENCES


Antoniadis A. 1997. Wavelets in statistics: A review. *Journal of Italian Statistics Society* **6**: 291–304.

Aminghafari M and Poggi J M (2007). Forecasting time series using wavelets. *International Journal of Wavelets, Multiresolution and Information Process* **5**: 709–24.

Beran J. 1995. *Statistics for Long-Memory Processes*, Chapman and Hall Publishing Inc, New York.

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

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

Geweke J and Porter-Hudak S. 1983. The estimation and application of long-memory time series models. *Journal of Time Series Analysis* **4**: 221–38.

Granger C W J and Joyeux R. 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**(1): 15–19.

Hosking J R M. 1981. Fractional differencing. *Biometrika* **68**: 559–67.

Ogden T. 1997. *Essential wavelets for statistical applications and data analysis*, Birkhauser, Boston.

Paul R K, Gurung B, Samanta S and Paul A K. 2014. Modelling long memory in volatility for spot price of lentil with multi-step ahead out-of-sample forecast using AR-FIGARCH model. *Economic Affairs* **60**(3): 457–66.

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): 319–27.

Paul R K. 2014. Forecasting wholesale price of pigeon pea using long memory time-series models. *Agricultural Economics Research Review* **27**(2): 167–76.

Paul R K. 2017. Modelling long memory in maximum and minimum temperature series in India. *Mausam* **68**(2): 317–26.

Paul R K and Mitra D. 2020. Forecasting of price of rice in India using long memory time series model. *National Academy of Science Letter* (Accepted).

Paul R K, Gurung B and Paul A K. 2015. Modelling and forecasting of retail price of arhar dal in Karnal, Haryana. *Indian Journal of Agricultural Science* **85**(1): 69–72.

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

Renaud O, Stark J L and Murtagh F. 2003. Prediction based on a multiscale decomposition. *International Journal of Wavelets, Multiresolution and Information Process* **1**: 217–32.

Soltani S, Boichu D, Simard P and Canu S. 2000. The long-term memory prediction by multiscale decomposition. *Signal Processing* **80**: 2195–05.