Indian Journal of Agricultural Sciences **89** (5): 882–6, May 2019/Article
https://doi.org/10.56093/ijas.v89i5.89682

Agricultural commodity price analysis using ensemble empirical mode decomposition: A case study of daily potato price series

KAPIL CHOUDHARY[1], GIRISH K JHA[2], RAJEEV R KUMAR[3] and DWIJESH C MISHRA[4]

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

Received: 06 November 2018; Accepted: 06 December 2018

ABSTRACT

Due to multifaceted nature of agricultural price series, conventional mono-scale smoothing approaches are unable to catch its nonstationary and nonlinear properties. Recently, empirical mode decomposition (EMD) has been proposed as a new tool for time-frequency analysis method, which adaptively represents nonstationary signals as sum of different components. The essence of EMD is to decompose a time series into a sum of intrinsic mode function (IMF) components with individual intrinsic time scale properties. One of the major drawbacks of the EMD is the frequent appearance of mode mixing. Ensemble EMD (EEMD) is a substantial improvement of EMD which can better separate the scales naturally by adding white noise series to the original time series and then treating the ensemble averages as the true intrinsic modes. In this paper, daily price data of potato in Bangalore and Delhi markets are decomposed into eight independent intrinsic modes and one residue with different frequencies, indicating some interesting features of price volatility. Further, decomposed IMFs and residue obtained through EEMD are grouped into high frequency, low frequency and a trend component which has similar frequency characteristics, using the fine-to-coarse reconstruction algorithm. These IMF and residue can be used for prediction using any traditional or artificial intelligence technique.

Key words: Agricultural commodity price analysis, Ensemble empirical mode decomposition, Fine-to-coarse reconstruction, Intrinsic mode functions

Agricultural commodity prices play a crucial role in consumer's affordability to food as they directly influence their real income, especially among the small and marginal farming community who spend a large proportion of their income on food (Jha and Sinha 2013). Among agricultural commodities, vegetable prices are more unstable due to perishable nature, seasonality of production, production uncertainty, etc. Potato (*Solanum tuberosum* L.) is the third most essential food crop after rice and wheat in the world. In India, potato is cultivated on approximately 18–19 lakh ha land. In view of frequent mismatch between demand and supply of tomato, onion and potato (TOP) crops, the Central Government in the Union Budget of 2018–19 launched *Operation Greens*, which will promote processing of these commodities and boost the supplies of these vegetables. Timely and reliable price forecast of these commodities are essential for farmers in order to make production and marketing decisions that may have financial implications many months later in the future (Allen 1994, 2014, Sundaramoorthy *et al*).

Agricultural commodity price analysis is different from non-farm goods and services due to its strong dependence on biological processes. Relative to non-farm goods, agricultural price series is greatly influenced by unpredictable random events such as droughts, floods and attacks by disease and pests, which leads to non linear, non stationary and noisy data. In order to have better forecast from this noisy data, smoothing techniques are employed. The empirical mode decomposition (EMD), proposed by Huang *et al*. (1998) is a self-adaptive decomposition technique, using spectral analysis, to improve forecasting precision for non linear and non stationary time series data. EMD was initially proposed for study of signal processing and has been successfully applied in many areas such as biomedical engineering, health monitoring, energy forecasting, exchange rate prediction etc. (Lin *et al*. 2012, Zhu 2012, Zhu *et al*. 2016, Huang and Wu 2017). A new spectrum analysis method, the EMD based on Hilbert-Huang-transform has emerged as a possible alternative to the traditional methods. The ensemble empirical mode

[1]Ph D Scholar (choudharykapil832@gmail.com), [2]Principal Scientist (gkjha@iari.res.in) , [3,4]Scientist (rrk.uasd@gmail.com, dwijesh.mishra@icar.gov.in).

decomposition (EEMD) method was developed by Wu and Huang (2009), which significantly reduces the chance of mode mixing and represents a substantial improvement over the original EMD.

MATERIALS AND METHODS

Empirical mode decomposition: The empirical mode decomposition (EMD) is a form of adaptive time series decomposition technique for non linear and non stationary time series data like agricultural price series. The basic principle of EMD is to decompose a time series into several simple modes, known as intrinsic mode functions (IMFs) and a residue (Huang *et al*. 1998). Every IMF has different amplitude and frequency modulation for a given data set. Each IMF should follow two conditions. First, in the whole data series, the number of extreme values and zero crossing must be equal or differ at most by one and second, the mean value of the envelope defined by the local maxima and minima should be zero at all points. For the given price series y_t, the steps of EMD are described here.

1) Determine all local maximum and minimum points of y_t.
2) Connect all local maximum and minimum points with the help of a spline functions to form the upper envelope y_t^{up} and the lower envelope y_t^{low} respectively.
3) Compute the point by point envelope mean $m_t(11)$ from lower and upper envelopes,

$$m_t(11) = \frac{y_t^{up} + y_t^{low}}{2} \tag{2.1}$$

4) Subtract mean from the input data series, $S_t(11) = y_t - m_t(11)$ (2.2)
5) Evaluate whether St (11) satisfies IMF's conditions or not. If it satisfies then it is first IMF. Otherwise repeat 1 to 4 steps by considering $S_t(11)$ as original series instead of y_t to find out $S_t(12)$. The process is repeated k times until $S_t(1k)$ doesn't satisfy an IMF's conditions.
6) $S_t(1k)$ is first IMF which is represented by $c_t(1)$. Then first residual $r_t(1)$ is

$$r_t(1) = y_t - c_t(1) \tag{2.3}$$

Now instead of y_t, new input data series is $r_t(1)$ in sifting process. All steps from 1 to 6 repeated to find out rest $n - 1$ IMFs ($c_t(2)$, $c_t(3)...c_t(n)$).

So $$y_t = \sum_{i=1}^{n} c_t(i) + r_t \tag{2.4}$$

Although the EMD approach has several evident advantage in processing non stationary and non linear price series, one of its major drawback is the mode mixing problem, which may weaken the physical meaning of IMFs (Zhang *et al*. 2015).

Ensemble empirical mode decomposition: To overcome the problem of mode mixing, Wu and Huang (2009) proposed the ensemble empirical mode decomposition (EEMD) method. The algorithm of EEMD is described here.

1) First of all, generate and add a number of Gaussian white noises $n_t(j)$ into data series y_t, $n_t(j) \sim N(0,\sigma^2)$

$$y_t(j) = y_t + n_t(j) \tag{2.5}$$

2) Apply EMD to newly formed price series $y_t(j)$ and decompose it into a set of IMFs $c_t(ij)$ and residual $r_t(j)$ where $c_t(ij)$ is the $j^{th}(i$ - 12,..., $n)$ IMF decomposed by EMD after adding the $n_t(j)$ for $j^{th}(j = 1,2,...,w)$ time.
3) Repeat first and second step w times to obtain all IMFs and residual. The i^{th} IMF $c_t(i)$ and final residual r_t are given as

$$c_t(i) = \frac{1}{w}\sum_{j=1}^{w} c_t(ij) \text{ and } r_t = \frac{1}{w}\sum_{j=1}^{w} r_t(j) \tag{2.6}$$

Hence, $$y_t = \sum_{i=1}^{n} c_t(i) + r_t \tag{2.7}$$

Fine to coarse reconstruction: The fine-to-coarse reconstruction (FCR) method (Zhang *et al*. 2008) is employed to all the decomposed modes to identify high frequency components (HFs), low frequency components (LFs) and trend component (T). HFs consist of all IMFs whose frequency is high and amplitude is low and show a random and short term fluctuations of price series. LFs consist of all those IMFs which have low frequency and high amplitude and exhibit periodic fluctuation. Trend component is the residual obtained after extracting all IMFs from a given price series. Steps for fine to coarse reconstruction method;

1) Compute the mean of the sum of $c_t(1)$ to $c_t(i)$ ($1 \leq i \leq n$), i.e. $S_t = \sum_{k=1}^{i} c_t(k)$ for each IMF.
2) Employ t-test to identify which i^{th} IMF's mean significantly departs from zero for the first time at selected significance level.
3) Once i^{th} IMF is identified as a significant change point, then from this IMF to last IMF is identified as low frequency component, and from first IMF to $(i - 1)^{th}$ IMF is identified as a high frequency component. The residue is identified as a trend component.

$h_t(t)$, (i = 12,...,m), $l_j(t)$, (j = 1,2,..,$n - m$) and r_t represent IMFs of high frequency, low frequency and residual or trend (T) component respectively, then original price series can be represented as

$$y_t = HFs + LFs + T \tag{2.8}$$

$$y_t = \sum_{i=1}^{m} h_i(t) + \sum_{j=1}^{n-m} l_j(t) + r_t \tag{2.9}$$

RESULTS AND DISCUSSION

In this study, we used the daily wholesale prices (rupees per quintal, ₹/q) of potato of Delhi and Bangalore markets obtained from National horticultural research and development foundation (NHRDF) (http://nhrdf.org/en-us/) for the period 02 January, 2012 to 23 march, 2018. The total number of available data points for Bangalore and Delhi markets were 1668 and 1901 respectively. The time plot of both the series clearly indicates the non linear and non stationary behaviour of the typical agricultural price data. According to basic descriptive statistics for both the price series (Table 1) the average price of potato is higher

Table 1 Descriptive statistics of potato price series in different markets

Statistics	Bangalore	Delhi
Mean	1369.09	996.14
Median	1300.00	905.00
Maximum	2800.00	2765.00
Minimum	500.00	310.00
Standard deviation	445.01	497.52
Skewness	0.92	1.10
Kurtosis	0.41	4.05
Jarque - Bera	245.66	471.05

in Bangalore market in comparison to Delhi market.

Tests for nonstationarity and non linearity: To test the prerequisites for employing EEMD i.e. non-stationarity and nonlinearity, we applied Augmented – Dickey – Fuller (ADF) and Brock-Decher-Scheinkman (BDS) tests to original price series. These tests provide an evidence for the presence or absence of unit root and nonlinearity respectively in the data set. The results of ADF test (Table 2) for two price series demonstrate that all the probability values of price series are not less than 9%, indicating that level price series are nonstationary at 9% level of significance. The results of BDS test are presented in Table 3. Here, embedding dimensions is set to 2 and 3, taking clue from the literature. The probability values are less than 0.001 for both dimensions, which show that both price series are nonlinear at 1% level of significance.

Table 2 Augmented–Dickey–Fuller (ADF) test result

Series	t-statistic	Probability	Conclusion
Bangalore market	-3.17	0.09	Non stationary
Delhi market	-2.80	0.25	Non stationary

Identification of HFs, LFs and trend component: The maximum number of IMFs is restricted to $\log_2 N$, where N is the number of data points (Fig 1). Through sifting process, we obtained 8 IMFs plus one residue for both the price series. All the IMFs are recorded from the highest frequency to the lowest frequency. In the figure, the last one is the residue, slowly varying around the long term average. Since these IMFs are orthogonal to each other, thus the sum of all components is equal to that of the original price series.

The fine-to-coarse reconstruction method (Zhang *et al*. 2008) was employed to all decomposed modes to identify high frequency components, low frequency components, and trend component. These two components and residue carry economic meaning and reveal some interesting features of potato price series. As mentioned earlier, EEMD was employed to decompose daily potato price series of Bangalore and Delhi markets into eight IMFs along with one residue. The test values of *t* statistics and probability of the IMFs from fine to coarse scale method is presented in Table 4. Results demonstrate that modes of Bangalore price begin to significantly deviate from 0 at the point i = 6 at 1% significance level therefore; partial reconstruction of $IMF_1 - IMF_5$ forms high frequency component and the partial reconstruction of $IMF_6 - IMF_8$ represent the

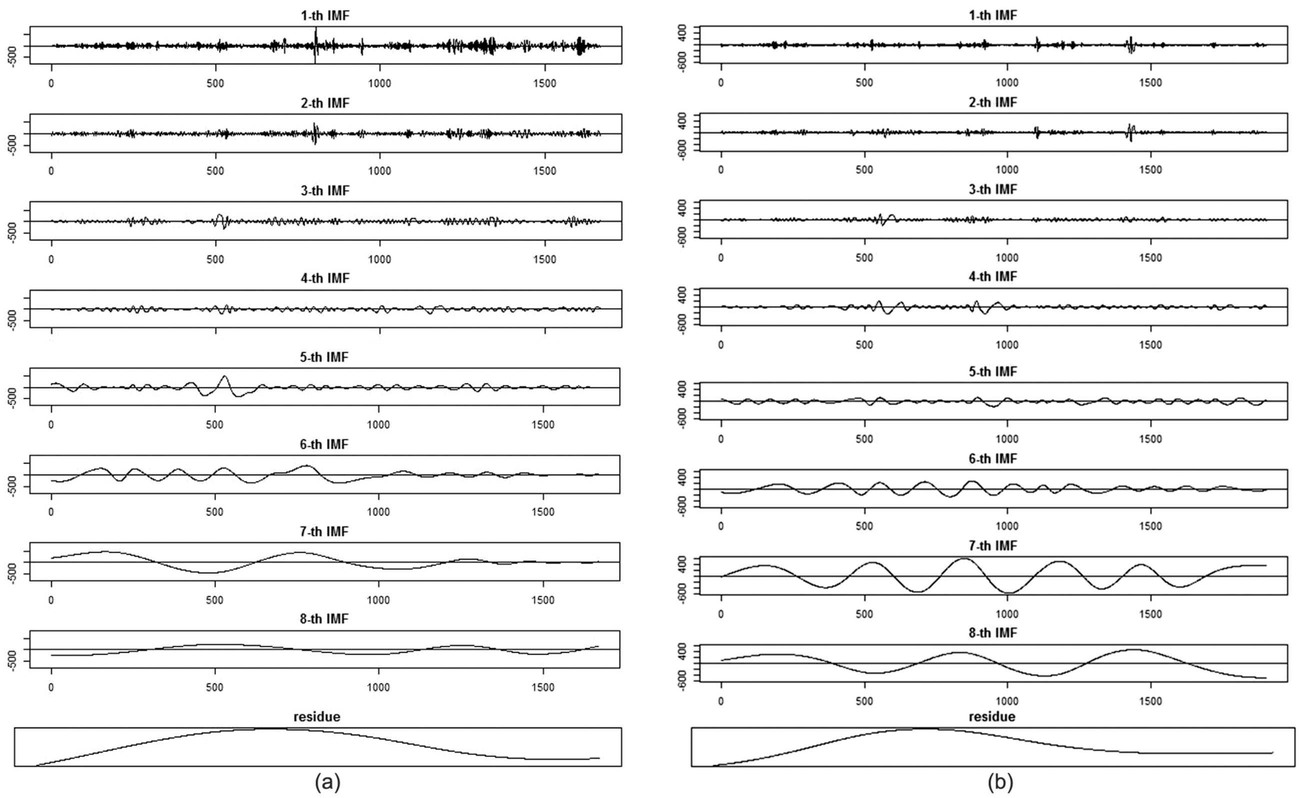

Fig 1 IMFs and residue for daily Potato wholesale price of (a) Banagalore and (b) Delhi markets.

Table 3 Brock-Decher-Scheikman (BDS) test result

Series	Embedding dimension 2		Embedding dimension 3		Conclusion
	Statistics	Probability	Statistics	Probability	
Bengaluru market	190.32	< 0.001	307.38	< 0.001	Non linear
	104.90	< 0.001	128.19	< 0.001	
	76.82	< 0.001	81.76	< 0.001	
	69.60	< 0.001	70.05	< 0.001	
Delhi market	308.63	< 0.001	537.12	< 0.001	Non linear
	152.62	< 0.001	188.08	< 0.001	
	100.66	< 0.001	105.94	< 0.001	
	082.19	< 0.001	079.76	< 0.001	

Table 4 Fine to coarse reconstruction for different market price series

Item	Bangalore market		Delhi market	
	t value	Probability	t value	Probability
S_1	0.19	0.84	0.16	0.87
S_2	0.12	0.89	0.42	0.66
S_3	1.36	0.17	0.71	0.47
S_4	1.88	0.06	-0.63	0.52
S_5	-0.78	0.43	-2.55	0.01
S_6	-3.15	<0.01	-2.48	0.01
S_7	0.10	0.91	-1.52	0.12
S_8	-1.28	0.20	0.95	0.33

low frequency component. Modes of Delhi price show a significant distance to 0 at the point $i = 5$ at 1% significance level hence IMF_1 to IMF_4 are HFs while $IMF_5 – IMF_8$ are LFs. Fig 2 shows the three components and each component has some characteristic features. In general, the residue is often treated as the deterministic long term behaviour. Trend component follows original series over a long period of time during the evolution of potato price.Each spike of the low frequency corresponds to a significant event while the high frequency with the characteristics of small amplitudes exhibits the short term fluctuation of potato market.

Table 5 provides Pearson correlation between each component and the original price series as well as variance of each component. The results clearly reveal that the dominant mode among three decompositions of the observed data is low frequency component for both price series. The Pearson's correlation coefficient between the low frequency component and the original price is 0.80 and 0.77 for Bangalore and Delhi markets respectively. At the same time, the low frequency component accounts for more than 67% and 74% of variability for Bangalore and Delhi markets respectively, suggesting that the effects of some significant events on potato price may be very crucial. Hence, the effect of every significant event can be measured by isolating the low frequency component from the original price series and the result can be used for predicting the next significant event of the same type. A significant event in 2014 for both markets, when the potato price soared 28-30% compared to last year mainly due to low production in major potato producing states like Uttar Pradesh, Bihar and Punjab due to heavy rain as well as significant increase in demand from the export market. In fact, the return (first difference) for the low frequency component often changes slowly, hence the curve generated by the significant event looks like a smooth handed curve of price series.

The trend component accounts for 24% and 38% of variability in Bangalore and Delhi markets respectively and holds correlation in the range of 0.45–0.50 with the original price series. In fact, by comparing the trend component with the observed price, it can be observed that the potato prices rise due to some significant event and it would return to the trend after the effect of the event is over. Due to rise in potato price in 2014, trend component moved upwards and gradually returned to the trend in both the markets after the influence of the event is over.

Besides significant events and the long term trend, agricultural commodity prices are also influenced by many high frequency events, such as fluctuation in fuel prices, strikes etc, with short-lived effects. The effects of short term fluctuations of markets are contained in the high frequency component. It is interesting to note that the high frequency component does not contribute significantly in the price determination of potato in Delhi market while it accounts for more than 10% of variability of potato price in case of Bangalore market (Table 5). Hence, high frequency component may be neglected for long term forecasting but not for short term forecasting.

Table 5 Correlation and variance of the component for different market price series

	Bangalore market			Delhi market		
	Pearson correlation	Variance	Variance as % of observed	Pearson correlation	Variance	Variance as % of observed
Observed		198037.10			247530.2	
High frequency component	0.35	21352.33	10.78%	0.09	6389.33	2.58%
Low frequency component	0.80	133091.00	67.21%	0.77	184139.20	74.39%
Trend	0.45	47516.99	23.99%	0.50	93646.55	37.83%

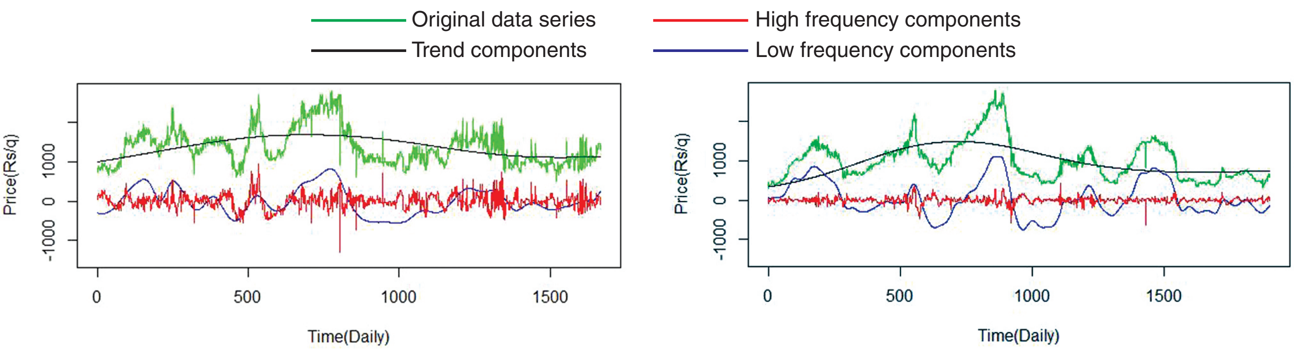

Fig 2 Three components of price series through FCR method for (a) Bangalore and (b) Delhi market price series.

In this study, ensemble empirical mode decomposition has been employed to decompose the non stationary and non linear daily potato price data of Bangalore and Delhi markets into several independent intrinsic modes with varying frequencies in order to analyse some interesting features of agricultural commodity price volatility. The IMFs and residue were composed into high and low frequency component, and trend component based on fine-to-coarse reconstruction. The economic interpretation of the 3 components are identified as short term fluctuation, effects of significant events and a long term deterministic trend. Results clearly indicate that potato price is mainly determined by the disturbances caused by significant events followed by the long term trend. The low frequency component accounted for more than 67% and 74% of variability for Bangalore and Delhi markets respectively. The main advantage of such analysis is that the effect of every significant event can be measured by isolating the low frequency component from the original price series and it can be used for predicting the next significant event of the same type. Several forecasting strategies can be planned after analyzing the composition of price series.

REFERENCES

Allen P G. 1994. Economic forecasting in agriculture. *International Journal of Forecasting* **10**: 81–135.

Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q and Liu H H. 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences* **454**: 903–95.

Huang D and Wu Z. 2017. Forecasting outpatient visits using empirical mode decomposition coupled with back-propagation artificial neural networks optimized by particle swarm optimization. *PLos ONE* **12**(2): e0172539. doi:10.1371/journal.pone.0172539.

Jha G K and Sinha K. 2013. Agricultural price forecasting using neural network model: Aninnovative information delivery system. *Agricultural Economics Research Review* **26**(2): 229–39.

Lin C S, Chiu S H and Lin T Y. 2012. Empirical mode decomposition–based least squares support vector regression for foreign exchange rate forecasting. *Economic Modelling* **29**(6): 2583–90.

Sundaramoorthy C, Jha G K, Pal, S and Mathur V C. 2014. Market integration and volatility in edible oil sector in India. *Journal of the Indian Society of Agricultural Statistics* **68**(1): 67–76.

Wu Z and Huang N E. 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis* **1**(1): 1–41.

Zhang X, Lai K K and Wang S Y. 2008. A new approach for crude oil price analysis based on empirical mode decomposition. *Energy economics* **30**(3): 905–18.

Zhang J L, Zhang Y J and Zhang L. 2015. A novel hybrid method for crude oil price forecasting. *Energy Economics* **49**: 649–59.

Zhu B. 2012. A novel multiscale ensemble carbon price prediction model integrating empirical mode decomposition, genetic algorithm and artificial neural network. *Energies* **5**(2): 355–70.

Zhu B, Shi X, Chevallier J, Wang P and Wei Y M. 2016. An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. *Journal of Forecasting* **35**(7): 633–51.