Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 77(1) 2023 79–88

EEMD-FCR-TDNN: A Hybrid Model for Forecasting Agricultural Commodity Prices

Girish Kumar Jha[1], Kapil Choudhary[2] and Ronit Jaiswal[2]

[1]*ICAR-Indian Agricultural Research Institute, New Delhi*
[2]*ICAR-Indian Agricultural Statistics Research Institute, New Delhi*

Received 16 March 2023; Revised 18 April 2023; Accepted 20 April 2023

SUMMARY

This paper aims to develop a hybrid model using ensemble empirical mode decomposition (EEMD) as a decomposition technique and time-delay neural network (TDNN) as a forecasting technique to predict non-stationary and nonlinear agricultural price series. The EEMD first decomposes the agricultural price series into several intrinsic mode functions (IMFs) and a single residual. Further, the resulting IMFs and residual series are grouped into high frequency, low frequency, and a trend component with similar frequency characteristics to capture numerous coexisting hidden factors using the fine-to-coarse reconstruction (FCR) algorithm. After that, a TDNN with a single hidden layer is built to separately forecast each of the three nonlinear components. Finally, the prediction results of all three components are summed up to obtain a final output as the forecast of the original price series. The performance of the proposed hybrid EEMD-FCR-TDNN model is empirically evaluated by comparing it with several benchmark models, including the TDNN model and decomposition-ensemble hybrid models without reconstruction using monthly international maize and soybean oil price series. The results validate that the EEMD-FCR-TDNN model can significantly outperform the other models in terms of both level and directional prediction accuracy with lower computational cost.

Keywords: Ensemble empirical mode decomposition; Fine-to-coarse reconstruction; Intrinsic mode function; Non-stationarity and non-linearity; Time-delay neural network.

1. INTRODUCTION

Prices of agricultural commodities play a vital role in producers' incentives to produce and consumers' economic access to food, leading to a usual dilemma for policy planners. Accurate forecasts of agricultural commodity prices reflecting cumulative information held by different economic agents can play a crucial role in marketing strategy and investment decisions and offer suggestions for agricultural policy planning (Wang *et al.*, 2020). However, the agricultural commodity market is influenced by several factors such as climate variability, including seasonality of production, the derived nature of demand, market imperfections, economic globalization, and a series of administrative regulations, making the price series extremely complex with nonlinearity, non-stationarity, and chaotic characteristics. All these complexities lead to the price prediction of agricultural commodities, an extremely challenging task.

An extensive review of the literature reveals that considerable efforts have been dedicated to improving price forecasting through various time series models over time (Yu *et al.*, 2017). These models can be broadly divided into three categories: statistical models, artificial intelligence (AI) techniques, and hybrid models combining several techniques systematically (Yu *et al.*, 2015). Among the statistical models, the autoregressive integrated moving average (ARIMA) and their constituent models are the most prevalent (Box *et al.*, 2015; Jha and Sinha, 2014). However, the ARIMA models are established on the linearity assumption in contrast to real-world agricultural price series, which are frequently nonlinear. As a result, these approaches fail to capture the nonlinear patterns concealed in the

Corresponding author: Girish Kumar Jha
E-mail address: girish.stat@gmail.com

original agricultural price series, resulting in poor forecast outcomes. On the other hand, AI techniques, specifically, artificial neural networks (ANNs), with their powerful self-learning capabilities and flexible architectures, have been widely utilised to model a nonlinear time series with few initial assumptions and better forecasting accuracy (Jha and Sinha, 2014; Zhang *et al.*, 1998). However, neural network-based models are weak at modelling seasonality and non-stationarity of time series data besides their intrinsic weaknesses, such as parameter sensitivity, over fitting and local minima.

To remedy the limitations of both statistical and AI models, a plenty of hybrid forecasting models have been proposed by integrating several techniques systematically to explore their effectiveness for forecasting applications (Yu *et al.*, 2019). In particular, utilising the promising idea of "divide and conquer", empirical mode decomposition (EMD) based ensemble models have been developed for complex time series forecasting with powerful prediction capabilities (Qian *et al.*, 2019). The main aim of the EMD technique (Huang *et al.*, 1998; Zhang *et al.*, 2008) is to alleviate the difficulty of modelling by dividing the complex original time series data into a set of relatively stable and simple components, whereas under the ensemble stage the prediction results of each component aggregate into the final prediction.

In recent years, multi scale decomposition based on EMD or its variants have been combined with TDNN models for different applications (crude oil price, wind speed, energy prices) of time series forecasting and obtained enhanced prediction accuracy in comparison to individual TDNN models (Qian *et al.*, 2019; Zhu *et al.*, 2016). However, to the best of our knowledge, limited efforts are made to analyse and forecast agricultural prices using EEMD based neural network model despite the fact that agricultural price forecasting differs from most other time series due to its unique features; thus, the goal of this study is to fill this gap in the literature.

In this study, we propose to improve the EEMD based TDNN model for agricultural price forecasting by utilising a data-driven component reconstruction technique. Our approach involves decomposing the agricultural price series into the short-run fluctuations, a slowly varying part, and long-run trend by exploiting EEMD and fine-to-coarse reconstruction (FCR) methods and predicting only three components instead of several modes; thus, we achieve a better interpretation and lower computational cost. The remaining part of the paper is laid out as follows: Section 2 describes the proposed EEMD-FCR-TDNN hybrid model for agricultural price forecasting in detail. For empirical evaluation of the proposed model, two internationally traded agricultural commodities, namely maize and soybean oil monthly price series, are selected in Section 3. Section 4 concludes the work.

2. METHODOLOGY

An overview of the EEMD based multi scale model for agricultural price forecasting is provided in this section. First, the EEMD, FCR, and TDNN techniques are described briefly. Following that, the suggested EEMD-FCR TDNN forecasting model is thoroughly explained.

2.1 Ensemble empirical mode decomposition (EEMD)

One of the main limitations of the traditional EMD method is that the decomposition results may be affected by mode mixing where a signal of different scales exists in one IMF or a signal of a similar scale resides in different IMFs (Zhu *et al.*, 2016) Wu and Huang (2009) proposed the EEMD technique by combining white noise to the original series leading to homogenization of the scale in time-frequency space so that the EMD can adaptively filter intrinsic local oscillation to the proper scales, which significantly reduces the chance of mode mixing and represents a substantial improvement over the original EMD. Thus, it decomposes a time series into finite simple independent intrinsic mode functions (IMFs) based on the local characteristics of the price series. The generic procedure for the EEMD method is described as follows:

1. Add a Gaussian white noise series with a predefined amplitude to agricultural price series $y(t)$ under consideration

   $$y_i(t) = y(t) + n_i(t)$$

   where $n_i(t)$ denotes the i^{th} added white noise series and $y_i(t)$ represents the noise-added agricultural price of the i^{th} iteration.

2. Find the local minima and maxima points (extrema) and zero crossing for the price series.

3. Connect these extrema by using a spline function to form the upper envelop and lower envelope.
4. Calculate the mean value of lower envelop and upper envelope. If the total number of extrema and the zero crossing points vary by one and the mean value of the envelope is nearly zero, then we subtract the mean from the original price series to obtain the first IMF. Accordingly, agricultural price $y_i(t)$ is decomposed into a set of IMFs $c_{ij}(t)$ and a residual $r_i(t)$ where $c_{ij}(t)$ is the j^{th} IMF for an i^{th} iteration.
5. Repeat the procedure mentioned above for the ensembles and take the ensemble average as the final decomposition result for each IMF and residual.

$$c_j(t) = \frac{1}{w}\sum_{i=1}^{w} c_{ij}(t) \text{ and } r(t) = \frac{1}{w}\sum_{i=1}^{w} r_i(t)$$

where w is the ensemble size.

2.2 Fine to coarse reconstruction (FCR)

The agricultural price series is decomposed into $n+1$ modes by using EEMD. These modes are further grouped into high frequency (HF), low frequency (LF), and trend components according to their inherent data characteristics by employing the fine-to-coarse reconstruction method (Zhang *et al.*, 2008; Zhu *et al.,* 2016). The HF component is made up of all IMFs with a high frequency but a low amplitude, and it shows random and short-term fluctuations. All IMFs with a low frequency and high amplitude show periodic oscillations that make up the LF component. Finally, the residual of a particular price series is known as the trend component. The procedure for the FCR method is as follows:

1) Calculate the average of the IMFs $(\bar{S}_i)$ from $c_1(t)$ to $c_i(t), (1 \le i \le n)$, *i.e.*, $\bar{S}_i = \frac{S_i}{i}$ where $S_i = \sum_{j=1}^{i} c_j(t)$
2) The t-test is used to determine which i^{th} IMF's mean $(\bar{S}_i)$ is significant at α significance level.
3) Once i^{th} IMF is identified, then all IMFs from i to n are grouped as a low-frequency component, and all IMFs from the first IMF to $(i-1)^{th}$ are grouped as a high-frequency component. The residual is a trend component for given price series.

If $\sum_{j=1}^{i-1} c_j(t)$, $\sum_{j=i}^{n} c_j(t)$ and $r(t)$ represent HF, LF and trend components respectivelythen the original price series can be represented as:

$$y(t) = \sum_{j=1}^{i-1} c_j(t) + \sum_{j=i}^{n} c_j(t) + r(t) \text{ i.e.}$$

$$y(t) = \text{HF + LF + Trend}$$

2.3 Time-delay neural network (TDNN)

Artificial neural networks (ANNs) are a type of intelligent learning paradigm that is employed in a wide range of applications. Time series can be modelled with neural networks in two ways: employing a recurrent neural network or constructing short-term memory at the network's input layer (Haykin, 2009). A time-delay neural network (TDNN) is an example of the latter, which uses time delays of a univariate time series to capture the temporal dimension of the series to develop a short-term memory, namely hetero-associative memory, in its network. TDNN with a single hidden layer is employed in this study as a learning tool for modelling the reconstructed component obtained from FCR. A TDNN with a single hidden layer has the following generic expression (Jha and Sinha, 2014):

$$\hat{y}(t) = g\left(\alpha_0 + \sum_{j=1}^{q} \alpha_j f\left(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y(t-i)\right)\right)$$

where $\hat{y}(t)$ is the predicted value, $y(t-i)$ is the i^{th} input (lag), $\alpha_j (j = 0,1,2, ..., q)$ and $\beta_{ij} (i = 0,1,2, ..., p; j = 1,2, ..., q)$ are connection weights, p and q are the number of input and hidden nodes, respectively f and g denote the activation functions at the hidden and output layers, respectively, of the model.

2.4 EEMD-FCR-TDNN model for agricultural price series

For any non-stationary and nonlinear time series, after using EEMD as a decomposition tool, the decomposed modes are grouped into three components using the FCR method. TDNN model is then used to model and predict these components separately. Thus, a multi scale ensemble model, namely, EEMD-FCR-TDNN, has been proposed by integrating EEMD, FCR, and TDNN. The procedure for this model can be separated into four parts:

1) **Data decomposition:** Data-driven decomposition technique EEMD is used to decompose a time series $y(t)$ into $n+1$ modes (IMFs and residual).

These modes have simple structure and stable fluctuation.

2) **Component reconstruction:** The fine to coarse reconstruction (FCR) technique group the decomposed $n+1$ modes into different meaningful components (HF, LF, and Trend) (Yu *et al.,* 2015). Now, instead of $n+1$ modes, the series is divided into three parts only.

3) **Individual prediction:** The TDNN model is well suited for modelling and capturing nonlinear series patterns. So, it has been used for forecasting each of the components.

4) **Ensemble prediction:** The final forecast of the original price series is obtained by adding the predicted value of all three components as:

$$\hat{y}(t) = \widehat{HF} + \widehat{LF} + \widehat{Trend}$$

where, $\widehat{HF}, \widehat{LF}, and\, \widehat{Trend}$ represent predicted values of high frequency, low frequency and trend components respectively.

2.5 Forecasting evaluation criteria

A comprehensive evaluation of each prediction model used in the study has been done in terms of root mean squared error (RMSE), mean absolute percentage error (MAPE), directional prediction statistics (D_{stat}), and Diebold-Mariano (DM) test, since no single accuracy metric can capture the distributional features of the errors completely (Jaiswal *et al.*, 2021; Zhu *et al.*, 2016). Out of the above four indicators, RMSE is a scale-dependent metric, while MAPE benefits being unit-free. These forecasting evaluation criteria for the proposed model are described as:

1) **Root mean squared error (RMSE)**

$$RMSE = \sqrt{\frac{\sum_{t=1}^{h}\left(y(t) - \hat{y}(t)\right)^2}{h}}$$

2) **Mean absolute percentage error (MAPE)**

$$MAPE = \left[\frac{1}{h}\sum_{t=1}^{h}\left|\frac{y(t) - \hat{y}(t)}{y(t)}\right|\right] \times 100\%$$

3) **Directional prediction statistics (D_{stat})**

$$D_{stat} = \frac{1}{h}\sum_{t=1}^{h} a_t \times 100\%$$

where $y(t)$ and $\hat{y}(t)$ are the actual value and predicted value, respectively, h is the size of the testing set and $a_t = \begin{cases} 1, \; if \left[y(t+1) - y(t)\right]\left[\hat{y}(t+1) - y(t)\right] \geq 0 \\ 0, \; otherwise \end{cases}$.

4) **Diebold-Mariano (DM) test**

For a given time series $y(t)$, the DM test (Jiang, 2021) statistics is defined as:

$$z_{DM} = \frac{\bar{d}}{\sqrt{\hat{V}_{\bar{d}}}}$$

where h is the test size, $\{e_{tet}\}_{t=1}^{h}$ and $\{e_{ref}\}_{t=1}^{h}$ are error for test model and reference model respectively, g is loss function, $d_t = g(e_{tet})_t - g(e_{ref})_t$ is the difference between the function prediction error of the two models, $\bar{d} = \frac{1}{h}\sum_{t=1}^{h} d_t$ sample mean, $\hat{V}_{\bar{d}} = \frac{1}{h}\left[\gamma_0 + 2\sum_{j=1}^{l-1}\gamma_j\right]$ estimate of variance using l step forecasts and $\gamma_j = cov(d_t, d_{t-j})$ is the estimate of j^{th} auto covariance of $\left[g(e_{tet}) - g(e_{ref})\right]$.

3. EMPIRICAL RESULTS AND DISCUSSION

In this section, two different agricultural commodity price series are used to explain the process of proposed model empirically. It mainly includes the EEMD process to extract IMFs, component reconstruction using the fine-to-coarse algorithm, and comparison of ensemble prediction results with different models for given price series.

3.1 Data description

To test the effectiveness of the proposed (EEMD-FCR-TDNN) prediction model, this study uses monthly international Maize and Soybean oil price series (dollar per metric tonne, $/MT) obtained from “World Bank Commodity markets ‘Pink Sheet’ data (https://www.worldbank.org/en/research/commodity-markets)” from the period January 1960 to June 2020. The long range of these time series data sets enables extracting more information and analysing agricultural commodity prices long-term. Both price series are divided into two subsets: the training and testing sets. For each time series, the training set consisting of 690 observations is used for the model’s parameters estimation and enhancing the generalization ability, and the last 36 months data is retained for evaluating the developed model. Time plots in Fig. 1 indicate the complex behaviour in both series like that of any typical agricultural price data. Table 1 shows the basic descriptive data for both price series. The international maize price varies from \$38/MT to \$333/MT, whereas soybean oil price ranges from \$157/MT to \$1533/MT during the sample period. The value of the coefficient of variation, a crude measure of volatility, is close to 50 percent which points towards the volatile nature of both series. Both price series are positively skewed and leptokurtic, indicating that they are non-normal. The Jarque-Bera test statistics also confirms the non-normal nature of both the series.

Since the empirical mode decomposition is suitable for a non-stationary and nonlinear process, it becomes imperative to test these characteristics of the price series. The Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979) is an effective unit root test to examine the stationarity, and the Brock-Dechert-Scheinkman (BDS) test (Choudhary *et al.*, 2019) can effectively check the nonlinearity characteristics of a time series. The results of the ADF test demonstrate that the probability values of maize and soybean oil price series are not less than 26% and 15%, respectively, indicating that price series are non-stationary.

BDS test examines the spatial dependence of a time series. Here, the embedding dimension is set to 2 and 3 with a length of 4, and the dimensional distance is set to 0.5 times the standard deviation of data. BDS test results (Table 2) exhibit that the probability values for price series are less than 0.001 for both dimensions, which confirms to nonlinearity at a 1% level of significance.

3.2 Data decomposition and component reconstruction

The EEMD decomposes the original price series of maize and soybean oil into a set of IMFs and one residual. The EEMD technique requires three hyperparameters, namely ensemble size (w), noise strength, and stopping criterion (*i.e.*, the maximum number of siftings). The determination of the values of these hyperparameters is based on the experimentation of the data such that the value of θ (evaluation parameter of energy) should be kept as near to zero as possible and ultimately provide better forecasting results. After going through the empirical evaluation of the different combinations of these hyperparameters, the ensemble size is fixed at 250, the noise that has been added to the original series is fixed at the strength of 0.2 times

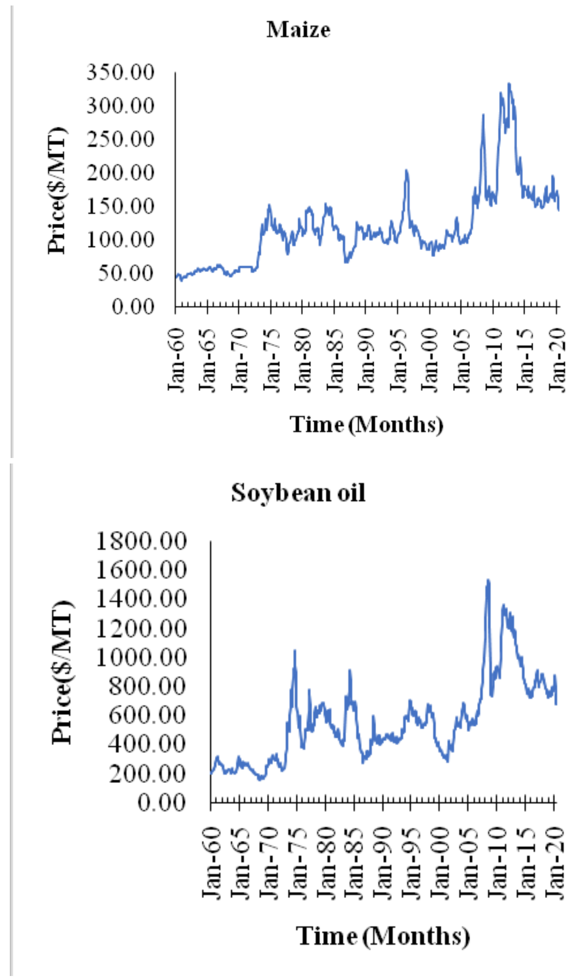

Fig. 1. Time plot for monthly international a) Maize and b) Soybean oil price series (\$/MT)

Table 1. Descriptive statistics of the price (\$/MT) series (from January 1960 to June 2020)

Statistics	Maize	Soybean oil
Mean, Standard deviation	117.56, 57.29	547.66, 275.87
Minimum,Maximum	38.00, 333.05	157.00, 1535.20
Skewness, Kurtosis	1.35, 5.31	0.95, 3.78

Table 2. Brock-Dechert-Scheinkman (BDS) test results

Price Series	Epsilon	Embedding dimension 2		Embedding dimension 3		Conclusion
		Statis-tics	Proba-bility	Statis-tics	Proba-bility	
Maize	0.5 σ	134.21	< 0.001	225.05	< 0.001	Nonlinear
	1.0 σ	69.93	< 0.001	81.06	< 0.001	
	1.5 σ	51.66	< 0.001	52.42	< 0.001	
	2.0 σ	41.22	< 0.001	39.61	< 0.001	

Note: For soybean oil, qualitatively similar results obtained, hence not reported here.

the standard deviation of each series, and the stopping criterion to obtain an IMF is fixed at 50. In practice, the total number of modes is equal to $\log_2 T$, where T is the total number of observations in the price series. Accordingly, both price series are decomposed into nine modes, *i.e.*, eight IMFs and one residual depicted from high frequency to low frequency in Fig. 2. The decomposed modes of EEMD should be orthogonal to each other so that the sum of energies of all modes is equal to that of the original price series. Practically, the decomposed modes do not follow this orthogonal property, and some elusive components along with modes are continuously generated in the sifting process called the end effect (Zhu *et al.*, 2016). This end effect increases the sum of energies of all modes, causing the changes in the degree of energies before and after the decomposition. The energy $E_{y(t)}$ of a time series $y(t)$ is defined as:

$$E_{y(t)} = \sqrt{\frac{\sum_{t=1}^{T} y^2(t)}{T}} \; ; \; \theta = \frac{\left| \sqrt{\sum_{j=1}^{n} E_j^2 + E_{r(t)}^2} - E_{y(t)} \right|}{E_{y(t)}}.$$

where $E_{y(t)}$, E_j, and $E_{r(t)}$ are the energy value of the original time series, j^{th} IMF and residual, respectively. Here, $\theta \geq 0$, as defined above, is used as an evaluation parameter for measuring the influence of the end effect. $\theta = 0$ represents no end effect, whereas the more positive value of θ indicates a greater influence of the end effect. A comparison between EMD and EEMD decomposition algorithms in terms of energy (θ) is presented in Table 3. It is observed that the value of θ using EEMD decomposition is lesser than EMD for both price series. This encouraged us to prefer EEMD over EMD as a decomposition technique for this study to build the proposed model.

The FCR method (Choudhary *et al.*, 2019; Zhang *et al.*, 2008) is used to group all decomposed modes (eight IMFs and one residual) obtained through EEMD into HF, LF, and trend components. These reconstructed HF, LF, and trend components show the different intrinsic properties of the original price series. For grouping decomposed modes, the $\boldsymbol{t}$-test statistic and probability values of the mean for FCR are presented in Table 4 for both price series. Table shows that at the point $\boldsymbol{i} = 4$ modes of maize price significantly change at the significance level of 2% and therefore $\boldsymbol{IMF}_1$ to $\boldsymbol{IMF}_3$ belong to HF components while $\boldsymbol{IMF}_4$ to $\boldsymbol{IMF}_8$ belong to the LF components. Modes of soybean oil price show the significant point at $\boldsymbol{i} = 5$ with a significance level of 1%, so $\boldsymbol{IMF}_1$ to $\boldsymbol{IMF}_4$ form HF components while $\boldsymbol{IMF}_5$ to $\boldsymbol{IMF}_8$ form LF components.

Table 3. Comparison of EMD and EEMD decomposition algorithm in terms of energy (θ)

Price series	EMD	EEMD
Maize	0.0405	0.0066
Soybean Oil	0.0179	0.0019

Table 4. Fine to coarse reconstruction for Maize and Soybean oil price series

Item	Maize		Soybean oil	
	t value	Probability	t value	Probability
S_1	-0.89	0.37	-1.62	0.10
S_2	-1.22	0.22	-0.53	0.59
S_3	-1.28	0.19	-0.45	0.64
S_4	**-2.32**	**0.02**	0.48	0.62
S_5	0.06	0.95	**3.78**	**< 0.01**
S_6	-3.05	0.002	1.86	0.062
S_7	0.50	0.61	5.09	< 0.01
S_8	-0.80	0.42	3.78	< 0.01

In general, the spikes of the HF component show the effects of short-term fluctuations of markets, whereas the LF component spikes represent a particularly significant event. Fig. 3 shows the time plots of all the three components for both price series. Figure 3 observed that nearly six such spikes are found in both price series. These spikes are generally observed due to some significant events (like changes in policy or adverse effects of several biotic and abiotic factors) affecting the demand-supply equilibrium at that time. For example, in our case, the two most significant spikes can be seen in 2008 and 2011 in maize and soybean oil prices, respectively. Reasons behind both events are the 2007-08's world food crisis and the production of biofuels (Trostle, 2011). For ethanol fuel production, usage of maize increased from 15% (2006) to 40% (2012) of total U.S. maize production. The trend component is often treated as the deterministic long-term behaviour. It follows the original series over a long time due to the evolution of price, global population, and the U.S. dollar depreciation value.

Moreover, which component out of the three would be dominant is wholly based on the inherent property of data. Thus it is imperative to find out the correlation and variance of these components with the original price

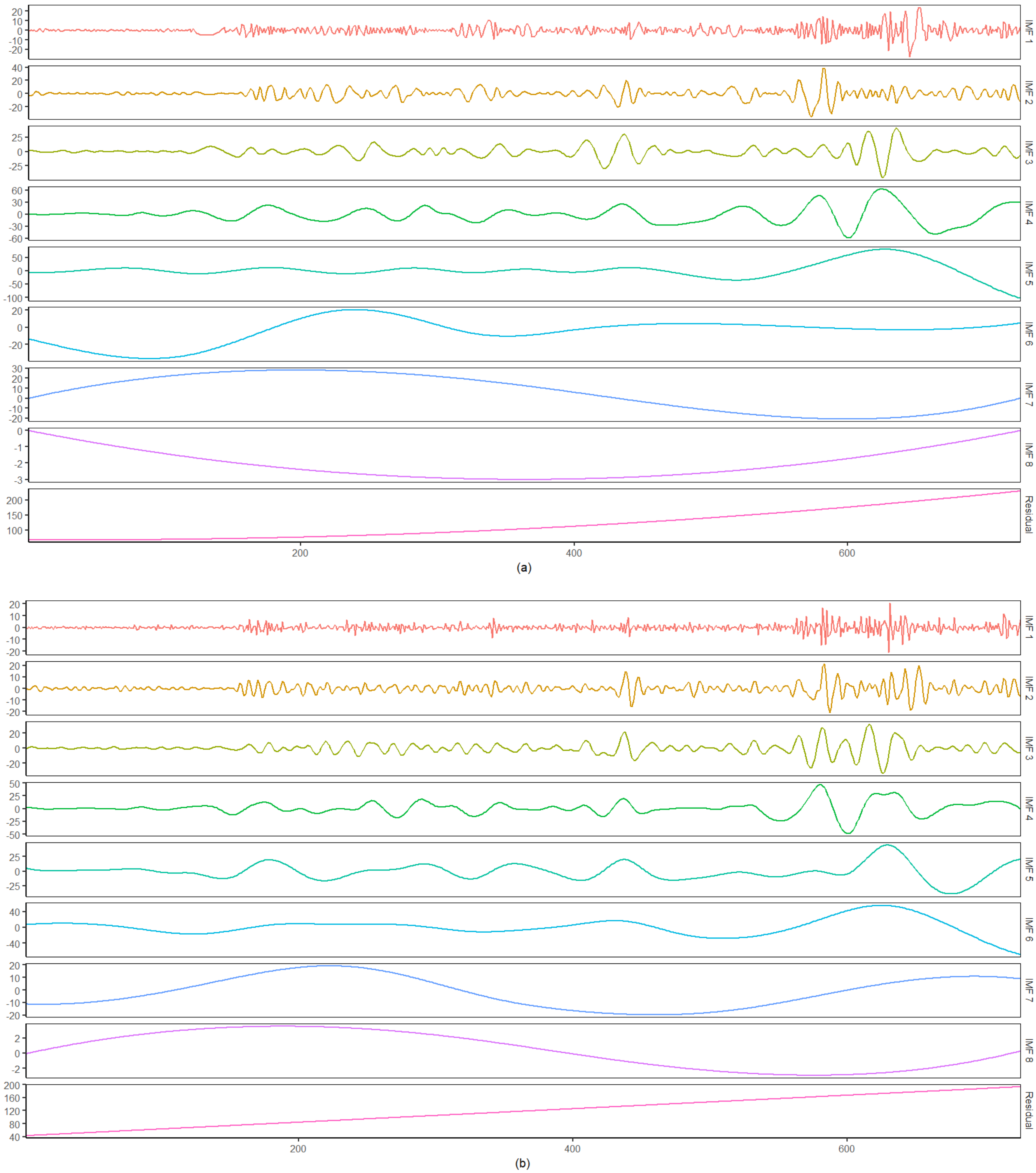

Fig. 2. The IMFs and residual decomposed by (a) EMD and (b) EEMD for Maize price

Note: For soybean oil, similar graph obtained, hence not reported here for brevity.

series. Table 5 shows Pearson's correlation coefficients between individual reconstructed components obtained through FCR and the original price series, along with the variability explained by each component for both price series.

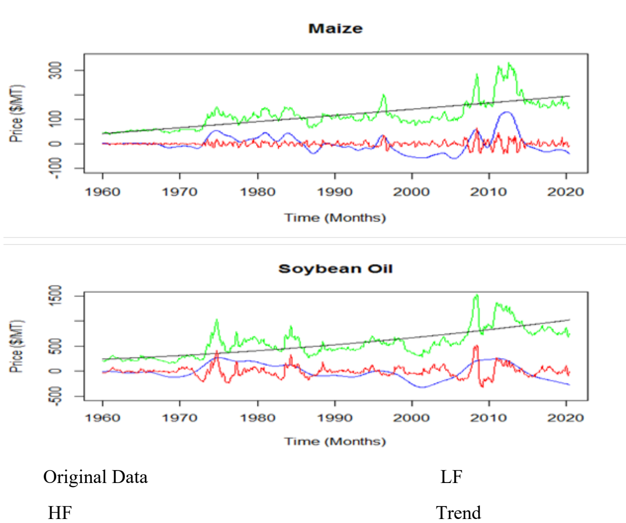

Fig. 3. FCR components for (a) Maize and (b) Soybean oil price series

According to the Table's result, the correlation coefficients between the trend and the original price series are 0.71 and 0.70 in maize and soybean oil, respectively. At the same time, the variability explained by the trend component accounts for more than 72% and 54% of the total variability for maize and soybean oil prices, respectively. Thus, it is observed that the dominant mode among the three decomposed components is the trend in both series, indicating that the trend component primarily governs both the price series. The result for the BDS test for HF, LF, and trend components shows that all the three components in maize and soybean are nonlinear. This encouraged us to employ a nonlinear model, *i.e.* TDNN, for individual prediction.

3.3 Forecasting results and discussion

For this study, we developed a R software package named *eemdTDNN* (Choudhary *et al.*, 2021) published in the *comprehensive R archive network (CRAN)*. Here, the *emdTDNN* and *EEMDTDNN* functions of the above package are used to forecast both price series. The forecasting performance of the proposed EEMD-FCR-TDNN model is compared with the existing monoscale model, *i.e.* TDNN and different multiscale ensemble models like EMD-TDNN (ensemble the value of all IMFs and residual decomposed by EMD and predicted by TDNN) and EEMD-TDNN (ensemble the value of all IMFs and residual decomposed by EEMD and predicted by TDNN) for both price series. Fig. 4 shows the plots of level series and predicted series obtained by all the models for both price series. The figure clearly shows that the EEMD-FCR-TDNN model captures price movement patterns and directions significantly better than conventional models. Moreover, the prediction ability of different models is tested in terms of different forecasting evaluation criteria. For both price series, RMSE, MAPE, directional prediction statistics (D_{Stat}) and Diebold-Mariano (DM) tests have been performed to evaluate each prediction model.

The results from Table 6 indicate that all multiscale ensemble forecasting models including, EMD-TDNN, EEMD-TDNN, and EEMD-FCR-TDNN, outperform the single prediction model, *i.e.* TDNN for both price data in terms of RMSE and MAPE. It is mainly due to the "decomposition-ensemble technique" where both decomposition techniques (EMD and EEMD) reveal the hidden patterns and trends of time series and produce stationary and nonlinear modes (IMFs and residual), which improve the forecasting ability of TDNN. Among multiscale models, EEMD-TDNN outperforms EMD-TDNN as expected since EEMD is better than EMD in terms of energy and ability to counter both mode mixing and end effect. The proposed EEMD-

Table 5. Correlation and variance of the components for the price series

	Maize			Soybean oil		
	Pearson correlation	Variance	Variance as % of observed	Pearson correlation	Variance	Variance as % of observed
Observed		3282.92			76109.41	
HF	0.23	158.18	4.82%	0.49	12010.9	15.78%
LF	0.50	1462.08	44.54%	0.48	25727.2	33.80%
Trend	0.71	2393.71	72.91%	0.70	41603.5	54.66%

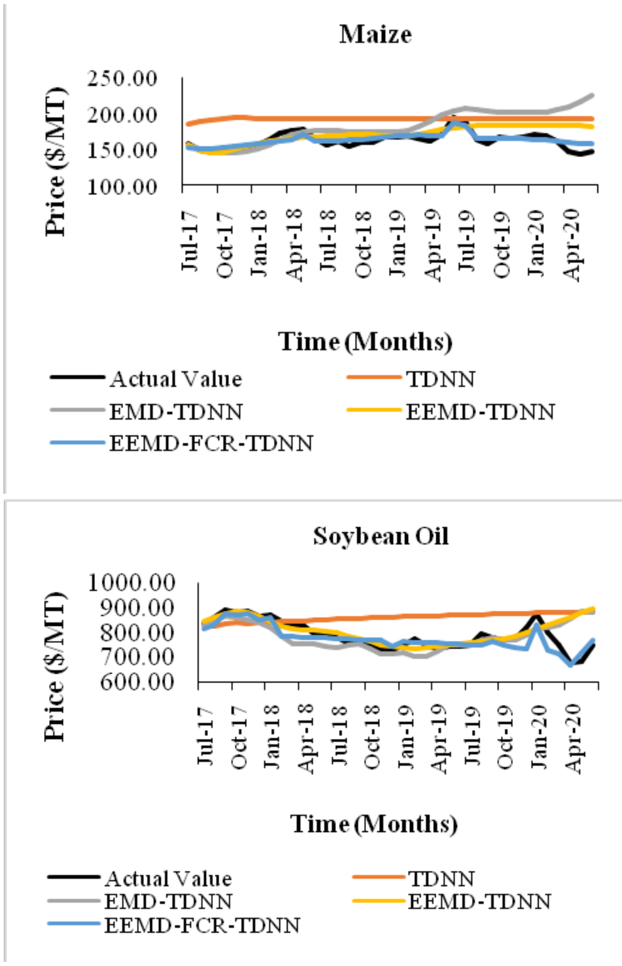

Fig. 4. The predicted results of different models for Maize and Soybean oil price series

FCR-TDNN model attains minimum RMSE and MAPE for both price series compared with individual and multiscale models. It is possible because of the algorithm of the FCR method utilized in the proposed model for grouping the IMFs and residual into HF, LF, and trend components. Each component obtained by FCR has apparent uniformity and holds centralized feature information. It means these components are able to better identify different characteristics of data such as short-term fluctuations, significant events, both short-term and long-term recurrent events, and their cumulative effects as individual IMFs fail to recognize them due to insignificant amplitude. After that, the most appropriate forecasting model, *i.e.* TDNN, is used for the three nonlinear components, which improves the forecasting ability of the proposed model.

With regards to D_{Stat} for direction prediction, the results are the same as RMSE and MAPE. As shown in Table 6, EEMD-FCR-TDNN significantly outperforms EEMD-TDNN, EMD-TDNN and TDNN models for both series. In both price series, the EEMD-FCR-TDNN model has the lowest RMSE and MAPE values, whereas the highest D_{Stat} values, as seen in the diagrams.

Table 6. Forecasting performance of TDNN, EMD-TDNN, EEMD-TDNN and EEMD-FCR-TDNN models for Maize and Soybean oil price series

Forecasting Models	Maize			Soybean oil		
	MAPE	RMSE	D_{stat}	MAPE	RMSE	D_{stat}
TDNN	0.19	32.31	57.14	0.11	99.43	42.85
EMD-TDNN	0.13	29.32	48.57	0.05	61.56	60.00
EEMD-TDNN	0.07	15.15	60.00	0.04	56.50	60.00
EEMD-FCR-TDNN	0.03	08.89	62.85	0.04	50.73	74.28

Other than these evaluation criteria, the Diebold-Mariano (DM) test is also used to evaluate the forecasting accuracy of different models statistically. Results of the DM test for each prediction model are presented in Table 7. Firstly, all the multiscale ensemble models (EMD-TDNN, EEMD-TDNN, EEMD-FCR-TDNN) perform better than the monoscale model (TDNN) at the significance level of less than 1% except for maize price where EMD-TDNN outperform TDNN at a 6 % significance level. Secondly, EEMD-TDNN outperforms EMD-TDNN for both price series, at the level of 1% and 4% significance, respectively. Thirdly, the proposed EEMD-FCR-TDNN model outperforms EMD-TDNN at a 1% and 2% significance level for the maize and soybean oil price series. Finally, the proposed EEMD-FCR-TDNN model significantly outperforms EEMD-TDNN at 1% and less than 1% significance level for maize and soybean oil price series, respectively.

Table 7. DM test result of several models for Maize price series (p-values in bracket)

Series	Tested Model	Benchmark Models		
		TDNN	EMD-TDNN	EEMD-TDNN
Maize	EMD-TDNN	02.64 (0.006)		
	EEMD-TDNN	08.59 (<0.001)	05. 19(<0.001)	
	EEMD-FCR-TDNN	09.84(<0.001)	04.39(<0.001)	03.08(0.001)

Note: For soybean oil, qualitatively similar results obtained, hence not reported here.

4. CONCLUSIONS

In this study, a new multiscale data-adaptive model, namely EEMD-FCR-TDNN, is proposed to predict non-stationary and nonlinear properties of time-series data. The proposed model is compared with existing mono scale model and multi scale ensemble models using the world market's monthly maize and soybean oil price series. Empirical results clearly reveal that all multiscale ensemble models outperform the monoscale model in terms of RMSE, MAPE and D_{Stat} value. Further, the proposed EEMD-FCR-TDNN model significantly outperforms all other models with respect to different forecasting evaluation criteria of both level and directional predictions. Finally, the proposed EEMD-FCR-TDNN model can serve as a competitive model for agricultural price forecasting.

REFERENCES

Box, G.E.P., Reinsel, G.C., Jenkins, M. and Ljung, G.M. (2015). *Time series analysis: forecasting and control.* Wiley, Germany.

Choudhary, K., Jha, G.K., Kumar, R.R. and Jaiswal, R. (2021). *eemdTDNN:EEMD and its variant based time delay neural network model*. https://cran.r-project.org/web/packages/eemdTDNN/eemdTDNN.pdf

Choudhary, K., Jha, G.K., Kumar, R.R. and Mishra, D.C. (2019). Agricultural commodity price analysis using ensemble empirical mode decomposition: A case study of daily potato price series. *Indian Journal of Agricultural Sciences*, **89(5)**, 882-886.

Dickey, D.A. and Fuller, W.A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, **74(366)**, 427. https://doi.org/10.2307/2286348

Haykin, S. (2009). *Neural Networks and Learning Machines*. PHI Learning, India

Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Snin, H.H., Zheng, Q., Yen, N.C., Tung, C.C. and Liu, H.H. (1998). The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **454(1971)**, 903-995. https://doi.org/10.1098/rspa.1998.0193

Jaiswal, R., Jha, G.K., Kumar, R.R. and Choudhary, K. (2021). Deep long short-term memory based model for agricultural price forecasting. *Neural Computing and Applications,* **34(6)**, 4661-4676.https://doi.org/10.1007/S00521-021-06621-3

Jha, G.K. and Sinha, K. (2014). Time-delay neural networks for time series prediction: An application to the monthly wholesale price of oilseeds in India. *Neural Computing and Applications*, **24(3-4)**, 563-571. https://doi.org/10.1007/s00521-012-1264-z

Jiang, W. (2021). Applications of deep learning in stock market prediction: Recent progress. *Expert Systems with Applications*, **184**, 115537. https://doi.org/10.1016/J.ESWA.2021.115537

Qian, Z., Pei, Y., Zareipour, H. and Chen, N. (2019). A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. *Applied Energy*, **235**, 939–953. https://doi.org/10.1016/j.apenergy.2018.10.080

Trostle, R. (2011). Why another food commodity price spike? *Amber Waves*, **9(3)**,18.https://doi.org/10.22004/ag.econ.120970

Wang, L., Feng, J., Sui, X., Chu, X. and Mu, W. (2020). Agricultural product price forecasting methods: research advances and trend. *British Food Journal,* **122(7)**, 2121–2138.https://doi.org/10.1108/BFJ-09-2019-0683

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. https://doi.org/10.1142/S1793536909000047

Yu, L., Liang, S., Chen, R. and Lai, K.K. (2019). Predicting monthly biofuel production using a hybrid ensemble forecasting methodology. *International Journal of Forecasting*, **38(1)**, 3-20. https://doi.org/10.1016/j.ijforecast.2019.08.014

Yu, L., Wang, Z. and Tang, L. (2015). A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. *Applied Energy*, **156**, 251–267. https://doi.org/10.1016/J.APENERGY.2015.07.025

Yu, L., Zhao, Y. and Tang, L. (2017). Ensemble Forecasting for Complex Time Series Using Sparse Representation and Neural Networks. *Journal of Forecasting*, **36(2)**, 122–138. https://doi.org/10.1002/for.2418

Zhang, G., Eddy Patuwo, B. and Y. Hu, M. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, **14(1)**, 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7

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)**, 905-918. https://doi.org/10.1016/j.eneco.2007.02.012

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)**, 633–651. https://doi.org/10.1002/for.2395