doi: 10.17221/268/2015-AGRICECON

An improved EEMD-based hybrid approach for the short-term forecasting of hog price in China

Tao XIONG[1], Chongguang LI[1], Yukun BAO[2]*

[1]*College of Economics and Management, Huazhong Agricultural University, Wuhan, China*
[2]*School of Management, Huazhong University of Science and Technology, Wuhan, China*

*Corresponding author: yukunbao@hust.edu.cn or y.bao@ieee.org

Xiong T., Li Ch., Bao Y.: **An improved EEMD-based hybrid approach for the short-term forecasting of hog price in China**. Agric. Econ. – Czech, 63: 136–148.

Abstract: Short-term forecasting of hog price, which forms the basis for the decision making, is challenging and of great interest for hog producers and market participants. This study develops improved ensemble empirical mode decomposition (EEMD)-based hybrid approach for the short-term hog price forecasting. Specifically, the EEMD is first used to decompose the original hog price series into several intrinsic-mode functions (IMF) and one residue. The fine-to-coarse reconstruction algorithm is then applied to compose the obtained IMFs and residue into the high-frequency fluctuation, the low-frequency fluctuation, and the trend terms which can highlight new features of the hog price fluctuations. Afterwards, the extreme learning machine (ELM) is employed to model the low-frequency fluctuation, while the autoregressive integrated moving average (ARIMA) and the polynomial function are used to fit the high-frequency fluctuation and trend term, respectively, in a multistep-ahead fashion. The commonly used iterated prediction strategy is adopted for the implementation of the multistep-ahead forecasting. The monthly hog price series from January 2000 to May 2015 in China is employed to evaluate the forecasting performance of the proposed approach with the selected counterparts. The numerical results indicate that the improved EEMD-based hybrid approach is a promising alternative for the short-term hog price forecasting.

Keywords: ensemble empirical mode decomposition (EEMD), extreme learning machine (ELM), hog price forecasting, hybrid approach, iterated prediction strategy

According to the *China Agricultural Yearbook 2014*, hog yields in China achieved 83.73 million tons in 2013, accounting for 65.6% of the total livestock yields in China. The livestock production sector in general and the hog production sector in particular, is the primary provider for the agricultural and domestic meat consumption in China.

Hog price has been exceptionally volatile in the recent years, reaching a low of 5.78 RMB/kg in May 2003 and a high of 19.68 RMB/kg in September 2011. The extraordinary fluctuations of the hog price in China exert adverse effects on the people's livelihood and the development of the related industries. Thus, exploring the fluctuation mechanism of the hog price, and then establishing an effective hog price prediction model are of great practical significance for the participants in the hog industry as well as the policy makers.

According to the extant literature review, it is observed that there have been few studies on the hog price forecasting though a considerable body of forecasting research focused on agricultural commodity price. Usually, hog is considered as one of the most important agricultural commodities. Thus, some analysis and discussions about the existing agricultural commodity price forecasting approaches will help to provide some valuable suggestions on the hog price forecasting. Within this field of agricultural

Supported by the Major Program of the National Social Science Foundation of China (Project No.12&ZD048), the Natural Science Foundation of China (Project No. 71501079 and 71571080), the China Postdoctoral Science Foundation (Project No. 2015M570648), the Fundamental Research Funds for the Central Universities (Project No. 2662014BQ045 and 2662015PY026), the MOE (Ministry of Education in China) Project of Humanities and Social Science (Project No. 13YJA630002), and a grant from the Modern Information Management Research Centre at the Huazhong University of Science and Technology (2015AA030).

doi: 10.17221/268/2015-AGRICECON

commodity price forecasting, the traditional statistical techniques, such as the regression analysis and econometric modelling, have already been widely applied (Ramirez and Fadiga, 2003; Hahn 2004; Jumah and Kunst 2008; Adanacioglu and Yercan 2012; Felipe et al. 2012; Li et al. 2012; Martín-Rodríguez and Cáceres-Hernández 2012; Saengwong et al. 2012; Paul et al., 2015). However, the traditional statistical techniques can provide good prediction results only when the price series under study are linear and stationary. Unfortunately, the agricultural commodity price, especially the hog price, may appear nonlinear and non-stationary. In the recent years, the artificial intelligence (AI) tools have been applied to deal with the problem of agricultural commodity price forecasting (Zhang et al. 2005; Shih et al. 2009; Ribeiro and Oliveira 2011; Li et al. 2013; Jha and Sinha 2014; Su et al. 2014). However, AI models have their own shortcomings, such as the local minima, time-consuming, and over-fitting in the neural networks. In the view of the limitations for the traditional and AI techniques in the agricultural commodity price forecasting and the general lack of research in the hog price forecasting, a novel approach is required to deal with the forecasting tasks of the hog price with a high nonlinear and volatility.

Following the "decomposition and ensemble" principle, the empirical mode decomposition (EMD)-based hybrid learning paradigm originally proposed by Yu et al. (2008) has been recently established and justified as a promising alternative to solve such tough forecasting tasks. In (Yu et al. 2008), the original complex crude oil price series is decomposed into several subseries first by the EMD technique, then they model them respectively and ensemble the forecasts obtained using neural networks finally. Since the work of Yu et al. (2008), the EMD-based hybrid learning paradigm has attracted a particular attention in the forecasting community and has been successfully applied to different areas (Yu et al. 2008; Napolitano et al. 2011; Chen et al. 2012; Kisi et al. 2014; Zhou et al. 2014). These works (Yu et al. 2008; Napolitano et al. 2011; Chen et al. 2012; Kisi et al. 2014; Zhou et al. 2014) have some improvements, however, they also suffer from some limitations. The main reasons are twofold. On the one hand, most of these works ((Yu et al. 2008; Napolitano et al. 2011; Chen et al. 2012; Kisi et al. 2014; Zhou et al. 2014) are preoccupied with the one-step-ahead forecasting. For example, Kisi et al. (2014) proposed a nonparametric technique based on the EMD to forecast the next month's monthly streamflows of the Kizilirmak river in Turkey. Zhou et al. (2014) developed an EMD-based general regression neural network model for one-day-ahead prediction of the $PM_{2.5}$ concentrations in Xi'an, China. A novel forecasting model based on the EMD and neural network is proposed by Chen et al. (2012) to forecast the next month's monthly tourist arrivals from Japan, Hong Kong, and Macao to Taiwan. Apparently, however, forecasting over short future horizons for the hog price series is of a great value to decision makers in the hog industry. For instance, the large-scale specialized producers could make a full use of the hog price forecasts over short future horizons when making plan on the number of hogs slaughtered. On the other hand, these works ((Yu et al. 2008; Napolitano et al. 2011; Chen et al. 2012; Kisi et al. 2014; Zhou et al. 2014) were apt to construct different models for each IMFs and residue, resulting in a complex and time-consuming training process, and more importantly, the economic meaning of the obtained IMFs is unclear and thus overlooked in the aforementioned works. Thus, by introducing the iterated strategy, a mainstream multi-step-ahead prediction strategy, and a fine-to-coarse reconstruction algorithm, which can compose the numerous IMFs into three main components, we develop an improved EEMD-based hybrid approach for the short-term forecasting of the hog price in China. By doing so, we hope that this study would fill the two gaps mentioned above.

Inspired by the idea of "decomposition and ensemble" principle (Yu et al. 2008), more specifically, this study proposes an improved ensemble empirical mode decomposition (EEMD)-based hybrid approach for the short-term hog price forecasting. Specifically, the original hog price series is decomposed via EEMD, which is a substantial improvement over the original EMD, into several IMFs and residue firstly. Then, the fine-to-coarse reconstruction algorithm (Zhang et al. 2008) is applied to compose the obtained IMFs and residue into a high-frequency fluctuation, low-frequency fluctuation, and trend terms. Afterwards, three different modelling techniques are used to predict each component based on its own features. To this end, the extreme learning machine, a novel learning algorithm for the feed-forward neural network with a fast learning speed and a high generalization performance, is used to model and forecast the low-frequency fluctuation due to its nonlinear pattern, while the autoregressive integrated moving average (ARIMA), a commonly-used stochastic model,

doi: 10.17221/268/2015-AGRICECON

is employed to solve the high-frequency forecasting. At the same time, the trend term is fitted by a polynomial function. For the implementation of the short-term hog price forecasting, the commonly used iterated prediction strategy is adopted here. The monthly hog price series from January 2000 to May 2015 in China is employed to evaluate the forecasting performance of the proposed approach with the selected counterparts by the goodness of forecast measure and testing approaches.

METHODOLOGY

Ensemble empirical mode decomposition

The ensemble empirical mode decomposition (EEMD) (Wu and Huang 2009) is a substantial improvement over the original EMD method developed by Huang et al. (1998) because it avoids the problem of mode mixing.

The two key parts of the original EMD as well as the EEMD method are the intrinsic-mode functions (IMF) and the sifting process. The term "intrinsic-mode function" is applied because it stands for the oscillation mode embedded in the series. To be specific, an IMF is a function that satisfies two conditions: (1) the number of zero crossing and the number of extremes must either be equal or differ at most by one in the whole series; (2) at any point, the mean value of the envelope determined by the local maxima and the envelope determined by the local minima is zero.

Upon the above definition, the IMFs can be extracted step by step from the series $x(t)$ on the basis of the following sifting process:

Step 1: Identify all local maxima and local minima;

Step 2: Connect the local maxima using a cubic spline to produce the upper envelope, $x_{up}(t)$, and the local minima using another cubic spline to produce the lower envelope, $x_{low}(t)$;

Step 3: Calculate the point-by-point local envelope mean $m(t)$ according to the upper and lower envelopes as $m(t) = (x_{up}(t) + x_{low}(t))/2$;

Step 4: Generate the component h by subtracting the local envelope mean from the data, $h = x(t) - m(t)$;

Step 5: Repeat *Step* 1 to *Step* 4 as many times as it is required while treating h as a series until the stop criterion is satisfied;

Step 6: The final h is assigned as an extracted IMF component c;

Step 7: Obtain the residue r as $r = x(t) - c$;

Step 8: Repeat *Step* 1 to *Step* 7 while treating r as a new series and until the final residue becomes a monotonic function.

Generally speaking, the IMF extraction process is made up of the process from *Step 1* to *Step 6*, and the whole sifting process is composed by the process from *Step 1* to *Step 8*. After finishing, the data series $x(t)$ can be decomposed into the IMFs and a residue, i.e.,

$$x(t) = \sum_{i=1}^{n} c_j + r_n \tag{1}$$

where n is the number of IMFs, c_j ($j = 1, 2, ..., n$) are the IMFs, and r_n is the residue, which represents the overall trend of the data series $x(t)$.

The underlying idea of the EEMD technique is that the observed series are amalgamations of the true series and noise; the observed series are closer to the true one by averaging the series with different noise levels. Thus, an additional step of adding the white noise is incorporated into the original EMD algorithm to solve the problem of mode mixing.

For the given data series $x(t)$, the EEMD procedure can be described as follows:

Step 1: Generate series with the added white noise, $x_i(t) = x(t) + w_i(t)$;

Step 2: Decompose the $x_i(t)$ by the sifting process, and then obtain the IMF components, $\sum_{j=1}^{n} c_{ij} + r_{in}$;

Step 3: Repeat *Step* 1 and *Step* 2 m times with a different white noise series each time; then generate a set of IMF components, $\sum_{i=1}^{m}(\sum_{j=1}^{n} c_{ij} + j_{in})$, where m denotes the ensemble number;

Step 4: Obtain the means of the corresponding IMFs as the final result, i.e., the j_{th} ensemble IMF, $\overline{c_j} = \frac{1}{m}\sum_{i=1}^{m} c_{ij}$, and the ensemble residue, $\overline{r_n} = \frac{1}{m}\sum_{i=1}^{m} r_{in}$.

Fine-to-coarse reconstruction algorithm

The original hog price series is decomposed into several IMFs and one residue via the EEMD. Then, a fine-to-coarse reconstruction algorithm is applied to compose the obtained IMFs and residue into the high-frequency fluctuation, the low-frequency fluctuation, and trend terms (Yu et al. 2015; Yu et al. 2016).

Given the IMFs and residue of hog price decomposed by EEMD, the fine-to-coarse reconstruction algorithm can be described as follows:

Step 1: Calculating the mean of the sum of IMF_1 to IMF_i for each IMFs;

doi: 10.17221/268/2015-AGRICECON

Step 2: Using the t-test to identify for which the mean significantly departs from zero.

Step 3: From the first tested mean, once a mean is identified as a significant change point, a partial reconstruction with IMFs from the first one to this one is identified as the high-frequency fluctuation, the rest of IMFs are identified as the low-frequency fluctuation. Finally, the residue is defined as the trend term.

Extreme learning machines

The extreme learning machines (ELM) proposed by Huang et al. (2006) are a novel and powerful approach for the data classification and regression. Detailed discussions of the ELM can be found in (Zhu et al. 2005; Huang et al. 2006), but a brief description on the formulation of the ELM for hog price forecasting is provided here.

The ELM is an improved learning algorithm for the single feed-forward neural networks (SLFNs) architecture. The ELM is different from the traditional neural networks in the sense that all the input weights, biases, and hidden nodes are randomly chosen in advance and the output weights are analytically tuned using the Moore-Penrose (MP) generalized inverse. Theoretically, the SLFNs with randomly chosen input weights, hidden layer biases and a nonzero activation function can approximate any continuous functions on any input set (Huang et al. 2006). The ELM not only outperforms the traditional gradient-based learning algorithm in terms of a faster learning speed with a higher generalization performance, but it also avoids many difficulties faced by the traditional gradient-based learning algorithm such as the stopping criteria and local minima (Huang and Babri 1998; Huang 2003; Huang et al. 2006; Shrivastava and Panigrahi 2014).

Given a data set with N arbitrary distinct sample (x_i, t_i) where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R^m$ and, the standard SLFNs with $\tilde{N}$ hidden neurons and activation function $g(\cdot)$ are mathematically modelled as

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i\left(w_i \cdot x_i + b_i\right) = y_i, \quad j = 1, 2, \ldots, N \tag{2}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ denotes the weight of the connection from the input neurons to the ith hidden neuron, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ denotes the weight vector which connects the output nodes with the ith hidden node, b_i is the threshold of the ith hidden node. The inner product of w_i and x_i is denoted by the operation $w_i \cdot x_i$. The above N equations can be written compactly as

$$H\beta = T \tag{3}$$

Where H is the hidden layer output matrix.

$$H = \begin{bmatrix} g\left(w_1 \cdot x_1 + b_1\right) & \cdots & g\left(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}\right) \\ \vdots & \cdots & \vdots \\ g\left(w_1 \cdot x_N + b_1\right) & \cdots & g\left(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}\right) \end{bmatrix} \tag{4}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix} \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} \tag{5}$$

The ELM theories claim that the input weights $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$ and hidden biases b_i are randomly generated and do not require any tuning. To minimize the cost function $\left\| Y - T \right\|$, the evaluation of the output weights $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$ is equivalent to find the least-square (LS) solution to the given linear system $H\beta = T$. The minimum norm LS solution is

$$\hat{\beta} = H^{\dagger}T \tag{6}$$

where $H^{\dagger}$ is the Moore-Penrose (MP) generalized inverse of matrix H.

Autoregressive integrated moving average

The autoregressive integrated moving average (ARIMA), proposed by Box and Jenkins (1976), is a most commonly-used univariate time series model. In the ARIMA, the future value is defined to be a linear function of several past observations and random errors. The formulation of the ARIMA model is as follows.

$$x_t = \theta_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \tag{7}$$

where x_t denotes the actual observation at time t, ε_t denotes the random error at time t, ϕ_i ($i = 1, 2, \ldots, p$) and θ_j ($j = 1, 2, \ldots, q$) are coefficients, p and q are the autoregressive and moving average polynomials, respectively.

THE IMPROVED EEMD-BASED HYBRID APPROACH

In this section, the proposed improved EEMD-based hybrid approach is formulated and the corresponding steps involved in this implementation are presented

doi: 10.17221/268/2015-AGRICECON

in details. Given the hog price series $x(t)$ for t = 1, 2, ..., n, we would like to make H-step-ahead forecasting, i.e., $x(t + H)$. A four-step modelling framework integrating the EEMD, the fine-to-coarse reconstruction algorithm, the ARIMA, the polynomial function, and the ELM can be formulated for the short-term hog price forecasting, see Figure 1.

As shown in Figure 1, the proposed improved EEMD-based ELM approach is generally made up of the following four main steps:

Step 1: The original hog price series is decomposed into several IMFs and a residue using the EEMD technique.

Step 2: The fine-to-coarse reconstruction algorithm is applied to compose the obtained IMFs and residue into the high-frequency fluctuation, the low-frequency fluctuation, and trend terms.

Step 3: The ELM is used to solve the low-frequency fluctuation forecasting task, while the ARIMA and the polynomial function are employed to deal with the high-frequency fluctuation and trend item forecasting task, respectively.

Step 4:The forecasts of the high-frequency fluctuation, low-frequency fluctuation, and trend terms are aggregated using another independent ELM model, which models the relationship among the three parts,

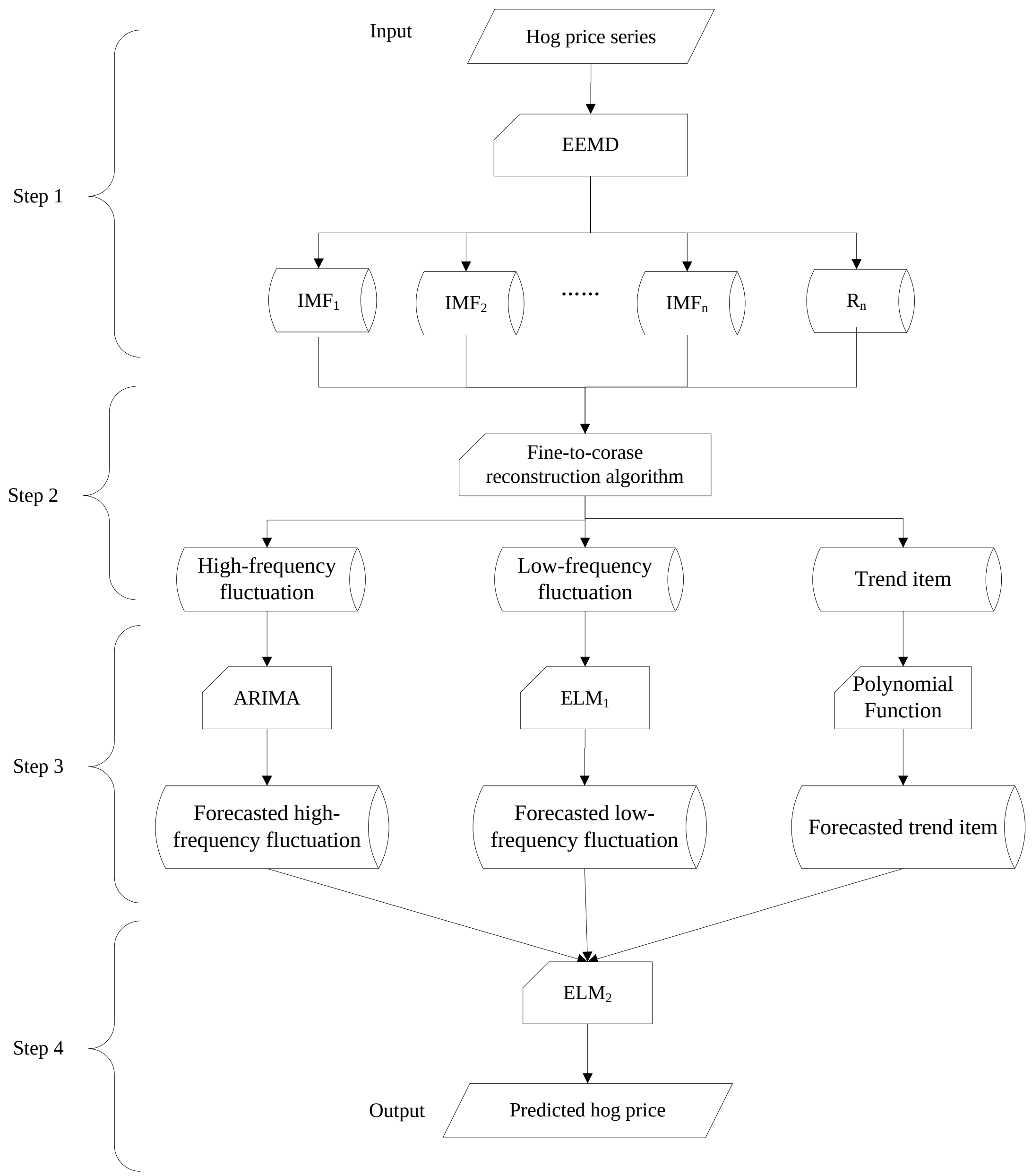

Figure 1. The proposed EEMD-based hybrid approach

doi: 10.17221/268/2015-AGRICECON

to generate and ensemble forecasts for the original hog price series.

For the comparison purposes, some popular individual forecasting approaches, which have been used for the hog price forecasting, should be adopted as the benchmarks. According to the extant literature investigation aforementioned, the main methods are divided into two categories, i.e., the traditional statistical techniques and the AI tools. Concerning the former, the ARIMA is selected as a competitor here. Amongst the AI tools, three individual modelling techniques, i.e., the feed-forward neural networks (FNN), the support vector regressive (SVR), and the ELM, can be seen as excellent benchmarks. Furthermore, some variations of the hybrid ensemble learning approaches without the fine-to-coarse reconstruction algorithm, i.e., the EEMD-ARIMA, the EEMD-FNN, the EEMD-SVR, and the EEMD-ELM are also employed as the hybrid ensemble benchmarks. In summary, four individual approaches, i.e., the ARIMA, F the NN, the SVR, and the ELM, and four hybrid ensemble learning approaches, i.e., the EEMD-ARIMA, the EEMD-FNN, the EEMD-SVR, and the EEMD-ELM are performed as counterparts to compare with the proposed EEMD-based hybrid model in terms of the prediction performance.

In order to verify effectiveness of the proposed EEMD-based hybrid approach, the hog price series in China is used as the testing target, as illustrated in the following section.

EXPERIMENTAL ANALYSIS

Data description and experiment design

In this study, the hog price in China is chosen as experimental sample. The data are freely obtained from the Ministry of Agriculture of the People's Republic of China[1]. The object of the current study is to develop an improved EEMD-based hybrid approach for the short-term hog price forecasting with various prediction horizons. By doing so, the monthly series of hog price from January 2000 to May 2015, with a total 185 observations, is employed for various prediction horizons H (in our case $H = \{1,3,6\}$). Hog price series is split into an estimation sample and hold-out sample. The data from January 2000 to March 2010 are used for the estimation sample (123 observations), and the last 62 observations from April 2010 to May 2015 is used as the hold-out sample. The original data are showed in Figure 2. Each examined model is trained on the estimation sample, and the forecasts are produced for the whole hold-out sample. The forecasts are then compared to the hold-out sample to evaluate the performance of each model.

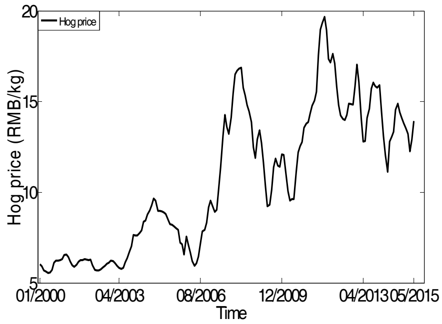

Figure 2. Monthly hog price series in China

Table 1 reports the summary statistics of the hog price. The estimated measure of skewness suggests that the hog prices have non-symmetric distributions. Moreover, the large positive value of kurtosis statistic, 1.806, suggests that the hog price is fat-tailed and sharply peaked around the mean in comparison to the standard normal distribution. The large value of the Jarque-Bera statistic further suggests that the null hypothesis of following the normal distribution is rejected for the hog price at 5% level.

In this paper, we adopt the liner transference to adjust the original hog price series scaled into the range of [0, 1] as shown in Equation 8. The two main advantages of scaling are to avoid inputs in greater numeric ranges from dominating those in smaller numeric ranges, and to prevent numerical difficulties during the calculation.

$$y(t) = \frac{x(t) - \min x(t)}{\max x(t) - \min x(t)} \quad (8)$$

where $x(t)$ is the original hog price, $y(t)$ is the scaled hog price, max $x(t)$ and min $x(t)$ are the maximum

Table 1. The statistical description of the hog price

Statistic	Value	Statistic	Value
Sample size	182	Median	9.630
Mean	10.610	Skewness	0.310
Standard deviation	3.944	Kurtosis	1.806
Minimum	5.560	Jarque-Bera test	13.726
Maximum	19.680		

[1]The datasets are available at http://www.moa.gov.cn/

doi: 10.17221/268/2015-AGRICECON

and minimum values of hog prices, respectively. After the pre-processing step, we applied the forecasting model on this time series, once the forecasts have been obtained, we unwind the pre-processing.

Additionally, the performances of the proposed approach with the selected competitors on the short-term forecasting (prediction horizon $H = \{1,3,6\}$) are investigated here. It should be noted that the commonly-used iterated strategy for the short-term forecasting is employed in this study due to its simplicity and popularity in literatures (Guo et al. 2012; Taieb et al. 2012; Xiong et al. 2014; Xiong et al. 2015).

To investigate the performance of different models, no single accuracy measure can capture the distributional features of the errors. Here, we consider two alternative forecast accuracy measures: the root mean squared error (RMSE) and the symmetric mean absolute percentage error (SMAPE). The RMSE and SMAPE are the absolute and relative error measures of the actual and predicted values, respectively. The definitions of them are shown as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}\left(x(t)-\hat{x}(t)\right)^2}{N}} \qquad (9)$$

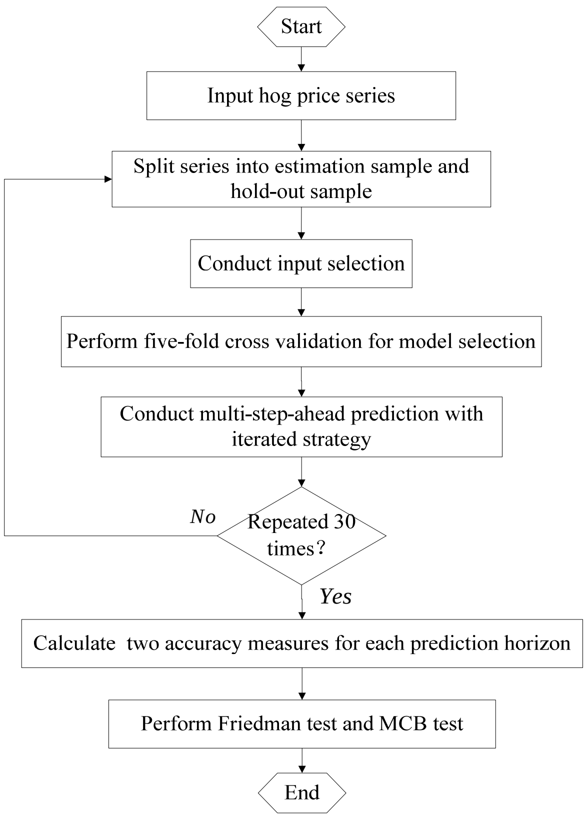

Figure 3. Experiment procedures

$$SMAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x(t)-\hat{x}(t)}{\left(\left|x(t)\right|+\left|\hat{x}(t)\right|\right)/2}\right| \qquad (10)$$

where $x(t)$ denotes the observation at period t, $\hat{x}(t)$ denotes the forecast of $x(t)$, N is the number of forecasting periods. Note that these accuracy measures are computed after rolling back the normalization step performed.

The filter method, which selects the best subset of input variables on the basis of dataset, is adopted for the input selection in the current study. By doing so, a pre-defined criterion, which measures the relationship between each subset of inputs and the outputs (Xiong et al. 2015), should be designed in advance for choosing the best subset of inputs. Specifically, the partial mutual information (PMI) (Sharma 2000) is used as the input selection criterion here. The maximum embedding order d is set to 15.

Figure 3 shows the experimental procedure using the hog price series. Firstly, the original hog price series is split into the estimation sample and hold-out sample. Then, both the input selection and model selection are conducted using the aforementioned filter method and the fivefold cross-validation with the iterated prediction strategies. Finally, the out-of-sample performance of each attained model is justified on hold-out samples, the RMSE and SMAPE are calculated for each prediction horizon H (H = 1, 3, 6). Furthermore, the modelling process for hog price series is repeated thirty times. Afterward, the performance of each model on each prediction horizon is compared on the basis of the mean, averaged by thirty, of the RMSE and SMAPE. In addition, the Diebold-Mariano (DM) test (Diebold and Mariano 1995) is used to test the statistical significance of different forecasting models at the 0.05 significance level.

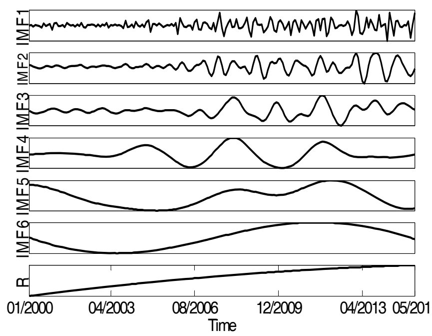

Figure 4. The IMFs and a residue of hog price via the EEMD

doi: 10.17221/268/2015-AGRICECON

Table 2. Measures of the IMFs and the residue for the hog price

Observed	Mean period (month)	Pearson correlation	Kendall correlation	Variance	Variance as % of observed	Variance as % of (Σ IMFs + residue)
IMF1	3.64	0.141*	0.067	0.085	0.545	0.561
IMF2	8.66	0.202***	0.118**	0.290	1.867	1.922
IMF3	22.75	0.392***	0.268***	0.981	6.306	6.492
IMF4	27.36	0.329***	0.193***	2.515	16.173	16.650
IMF5	60.66	0.436***	0.234***	0.864	5.553	5.716
IMF6	91.00	0.513***	0.478***	0.031	0.197	0.203
Residue		0.848***	0.645***	10.342	66.496	68.455
Sum					97.139	100.000

Correlation is significant at ***0.01 level (2-tailed), **0.05 level (2-tailed), *0.1 level (2-tailed)

Experimental results

In this subsection, the forecasting performances of the proposed model are justified with the hog price series against eight selected counterparts.

In the proposed model, the first step is to use the EEMD technique to decompose the original hog price series into several independent components. As shown in Figure 4, all obtained IMF components are listed in the order from the highest frequency to the lowest frequency, and the last one is the residue. It is clear that the original hog price series is decomposed into six IMF components and one residue component. The summary characteristics of the IMFs and residue are reported in Table 2.

It should be noted that the residue accounts for the vast majority of the variance. Indeed, the residue accounts for 66.496% of the total variability of the hog price, which suggests that the hog price is determined mainly by its long-term trends (that is, the residue). In addition, the Pearson correlation coefficient and the Kendal coefficient between the original price and residue are the highest. When looking deeply on the IMFs, we find that the second important mode is the high frequency IMF, i.e., the IMF4, which has a mean period of nearly 27.36 months (nearly 3 years).

In the previous step, the original hog price series is decomposed into 6 IMFs and one residue component. In the current step, the IMFs are grouped into the high-frequency fluctuation and the low-frequency fluctuation. The mean of the fine-to-coarse reconstruction as a function of IMFs index K is shown in Figure 5. The mean of the fine-to-coarse reconstruction departs significantly from zero at the IMF 4. Therefore, the partial reconstruction with the IMF1, IMF2 and IMF3 represents the high frequency component and the partial reconstruction with IMF4, IMF5 and IMF6 represents the low frequency component. The residue is treated separately. Figure 6 shows the three components. The statistical measures of these components and the observed price are given in Table 3. Looking at Table 3, it is clear that the residue is also the dominating one, regardless of the correlation coefficient and the percentage of variance

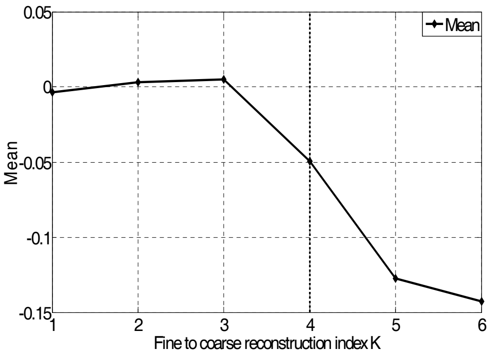

Figure 5. The mean of the fine-to-coarse reconstruction as a function of index K. The vertical dash-line at $K = 4$ indicates that the mean departs significantly from zero ($p < 0.01$)

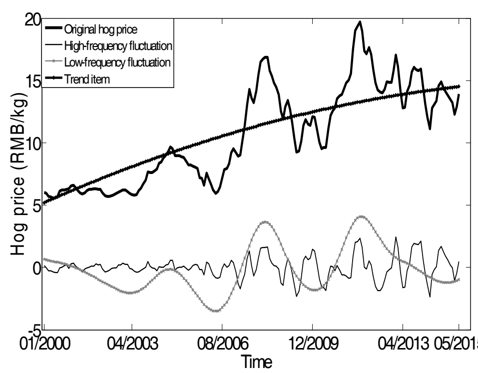

Figure 6. The three components of hog price from January 2000 to May 2015

doi: 10.17221/268/2015-AGRICECON

Table 3. Measures of three components for the hog price

Observed	Mean period (month)	Pearson correlation	Kendall correlation	Variance	Variance as % of observed	Variance as % of (Σ IMFs + residue)
				15.553		
HIGH	5.20	0.281***	0.181***	0.957	6.156	6.268
LOW	45.50	0.633***	0.466***	3.975	25.560	26.026
TREND		0.848***	0.645***	10.342	66.495	67.706
Sum					98.211	100.000

High-frequency component, Low-frequency component, and Trend are given under the rows 'HIGH','LOW', and 'TREND'. Correlation is significant at ***0.01 level (2-tailed), **0.05 level (2-tailed), *0.1 level (2-tailed)

considered. At the same time, noteworthy, the mean period of the high and low frequency components are near 5.2 and 45.5 months, respectively.

The third and fourth steps of the proposed model are the individual forecasting and the ensemble forecasting as mentioned in Figure 1. In this study, the ELM is implemented using the ELM package[2]. The input weights and hidden biases of the ELM are randomly generated in advance; the number of hidden nodes is tuning in a trial-error fashion. By doing so, we construct ten ELM models with different numbers of hidden nodes varying from 5 to 14. Each ELM model is trained repeatedly 20 times and the average MSE of each ELM model is computed. The number of hidden nodes in the best ELM model is chosen.

For the ARIMA estimation, the package 'forecast'[3] in R software developed in (Hyndman and Khandakar 2007) is used in this study. The FNN models used in this experiment are implemented using the MATLAB (Version R2006b) ANN toolbox. The architecture of the FNN model is as follows: the number of hidden nodes varies from 3 to 12 and the optimum number of hidden nodes that minimizes the error rate on the validation set is determined; the number of output nodes is set at one. For the stopping criteria, the number of learning epochs is chosen as 1000 as there is no prior knowledge of this value before the experiment. In the training phase, a gradient descent with momentum algorithms is applied to update the weight and bias values. The learning rate is chosen as 0.9 and the momentum constant is chosen as 0.1. The activation function of the hidden layer is sigmoid and the output node uses the linear transfer function. The LibSVM (version 2.86) (Chang and Lin 2011) is employed for the SVR modelling in this study. The Radial basis function (RBF) is selected as the kernel function through preliminary simulation. To determine the hyper-parameters, namely C, ε, γ, the commonly used grid search is applied in this study. With respect to the polynomial function estimation, the MATLAB functions 'polyfit' and 'polyval' are used here.

Figures 7–8 show the comparisons of performance of different models in all prediction horizons across two indices.

As per the results presented, one can deduce the following observations.

– Overall, the top three models across two indices and three prediction horizons turned out to be the proposed EEMD-based hybrid model, the EEMD-ELM and the EEMD-SVR are almost a tie. It is clear that the proposed EEMD-based hybrid model outperforms all other counterparts.
– It is clear that the proposed EEMD-based hybrid model outperforms the EEMD-ARIMA, EEMD-FNN, EEMD-SVR, and the EEMD-ELM prediction models without a fine-to-coarse reconstruction algorithm. As such, we argue that the superior performance of the proposed prediction model with the fine-to-coarse reconstruction algorithm relative to the EEMD-ARIMA, EEMD-FNN, EEMD-SVR, and EEMD-ELM as a result of composing the IMFs and residue into three components.
– Focusing on the modelling techniques (i.e., the ARIMA, FNN, SVR, and ELM) in hybrid models without the fine-to-coarse reconstruction algorithm, the EEMD-ELM and the EEMD-SVR are almost a tie. In addition, the EEMD-ELM, EEMD-SVR, and EEMD-FNN outperform the EEMD-ARIMA.
– As far as the comparison between the hybrid models (i.e., the proposed model, the EEMD-ELM, EEMD-SVR, EEMD-FNN, and EEMD-ARIMA) and the

[2]ELM package is available at http://www.ntu.edu.sg/home/egbhuang/elm_codes.html
[3]R package 'forecast' is available at http://ftp.ctex.org/mirrors/CRAN/

doi: 10.17221/268/2015-AGRICECON

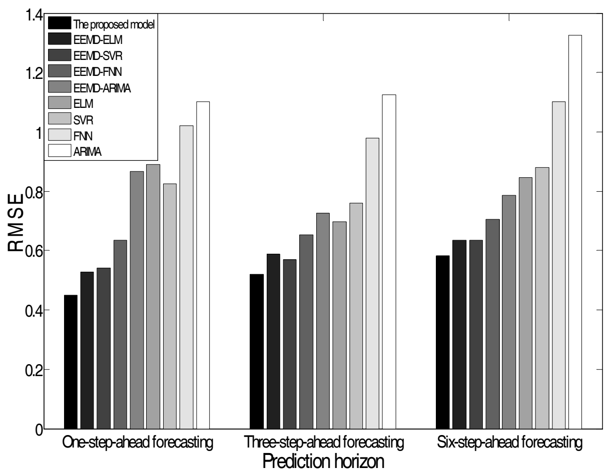

Figure 7. Performance comparison of different models in terms of the RMSE

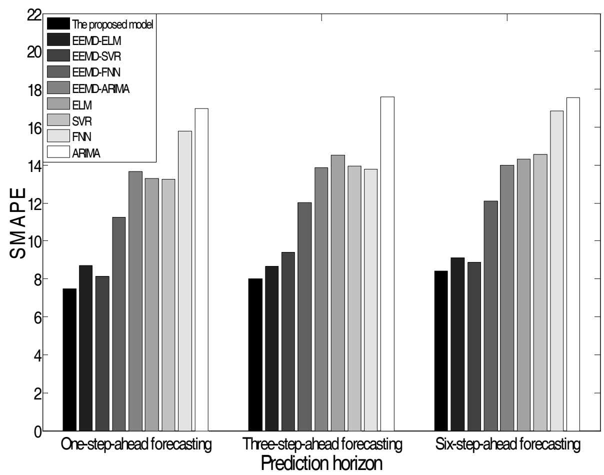

Figure 8. Performance comparison of different models in terms of the SMAPE

corresponding individual models (i.e., the ELM, SVR, FNN, and ARIMA), the hybrid models are better than those of the individual models without exception. The main reason could be that the decomposition technique does effectively improve the forecasting performance.

– When considering the comparison among the four individual techniques, the ELM seems to produce forecasts, which are more accurate than those of the SVR (though only marginally). In addition, the ELM and SVR can yield better prediction accuracy than the FNN and ARIMA. The ARIMA is consistently the worst performing model for prediction.

Following the experimental procedure mentioned early, we perform the DM test to examine the statistical significance of different forecasting models. The results of the DM test results for each prediction horizon are respectively reported in Table 4, where the S statistics are listed. Note that * denotes that the difference between the tested model and reference model is significant at the 0.05 level (2-tailed). According to the obtained results in Table 4, one can deduce the following observations:

– When the proposed model is treated as the testing target, the DM statistics are all negative and marked with the asterisk, indicating that the proposed improved EEMD-based hybrid model performs statistically better than all other models in all cases, under the confidence level of 95%. But some exceptions occur when the EEMD-SVR and EEMD-ELM are selected as the reference models in which the proposed model does not perform significantly better than the EEMD-SVR and EEMD-ELM.

– Concerning the four EEMD-based models, the EEMD-SVR and EEMD-ELM yield better results than the EEMD-FNN and the EEMD-ARIMA with the statistical significance of 95% in all cases.

– As far as the comparison of the EEMD-SVR and EEMD-ELM, the difference in the prediction performance between these two models is not significant at 0.05 level.

– Whatever the forecast horizon and whatever the modelling technique, the hybrid ensemble modelling frameworks (i.e. the EEMD-ELM, EEMD-SVR, EEMD-FNN, and EEMD-ARIMA) outperform the corresponding single models without exception at 95% statistical significance, implying the effectiveness of the "decomposition and ensemble" principle for the time series forecasting.

– Comparing the four single models, we can see that, whatever the prediction horizon, the ELM and SVR provide consistently better forecasts than the

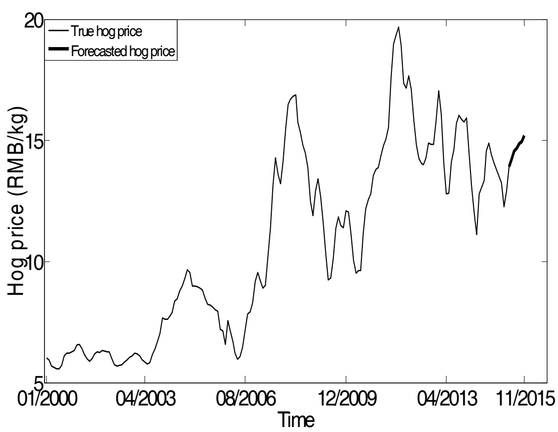

Figure 9. Short-term forecasting (six-step-ahead) for the hog price

doi: 10.17221/268/2015-AGRICECON

Table 4. DM test results for different models on the hold-out sample

Tested model	Reference model							
	ARIMA	FNN	SVR	ELM	EEMD-ARIMA	EEMD-FNN	EEMD-SVR	EEMD-ELM
	Prediction horizon H=1							
FNN	−0.32							
SVR	−2.14*	−1.55						
ELM	−1.97*	−1.38	1.97					
EEMD-ARIMA	−1.82*	−1.43	0.24	0.21				
EEMD-FNN	−3.05*	−2.64*	−0.56	−0.62	−1.08			
EEMD-SVR	−3.65*	−2.87*	−3.15*	−3.22*	−3.64*	−2.41*		
EEMD-ELM	−3.80*	−3.17*	−2.84*	−3.01*	−3.57*	−2.28*	1.02	
The proposed method	−4.21*	−4.02*	−3.42*	−3.63*	−3.89*	−3.07*	−0.57	−0.85
	Prediction horizon H=3							
FNN	−2.54*							
SVR	−2.17*	2.14						
ELM	−1.96*	1.85	1.57					
EEMD-ARIMA	−2.08*	2.08	−0.08	−0.21				
EEMD-FNN	−3.34*	−2.08*	−2.21*	−2.42*	−2.51*			
EEMD-SVR	−3.51*	−2.97*	−3.12	−3.24*	−3.29*	−2.58*		
EEMD-ELM	−3.82*	−3.54*	−3.29*	−3.51*	−3.37*	−2.71*	−0.55	
The proposed method	−4.37*	−3.93*	−4.12*	−3.97*	−3.86*	−2.84*	−0.86	−0.39
	Prediction horizon H=6							
FNN	−0.27							
SVR	−2.58*	−1.85*						
ELM	−2.61*	−1.96*	−0.25					
EEMD-ARIMA	−2.75*	−2.01*	−0.37	−0.19				
EEMD-FNN	−3.21*	−3.26*	−1.83*	−1.81*	−1.68*			
EEMD-SVR	−3.96*	−3.85*	−2.42*	−2.28*	−2.12*	−1.89*		
EEMD-ELM	−3.82*	−3.92*	−2.28*	−2.10*	−2.03*	−1.81*	1.86	
The proposed method	−4.31*	−4.19*	−2.67*	−2.72*	−2.64*	−2.38*	−0.41	−0.58

The entries are the DM statistics. In addition, *denotes that the difference between the tested model and reference model is significant at the 0.05 level (2-tailed)

FNN and the ARIMA at 95% statistical significance in most cases, suggesting that the FNN and the ARIMA are quite inferior in the multi-step-ahead time series modelling and forecasting.

The difference in prediction performance between the SVR and ELM is not significant at the 95% statistical confidence level in most cases, even with a few exceptions.

In addition, using the proposed EEMD-based hybrid approach, an out-of-sample forecasting for the period June 2015 to November 2015 is reported in Figure 9, where the data from January 2000 to May 2015 are also shown for consistency. As can be seen from Figure 9, it is easy to find that the overall tendency of the hog price in China continues to be choppy since 2014.

CONCLUSIONS

In this study, we propose an improved EEMD-based hybrid approach for the short-term hog price forecasting, in which the original hog price series is first decomposed into several IMFs and a residue by the EEMD technique. Then, the fine-to-coarse reconstruction algorithm is used to compose the obtained IMFs and the residue into the high-frequency fluctuation, the low-frequency fluctuation, and trend terms. Afterwards, the ELM is used to model the low-frequency fluctuation, while the ARIMA and the polynomial function are employed to fit the high-frequency fluctuation and the trend term, respectively. Finally, the forecasts of the three components are aggregated using another independent

doi: 10.17221/268/2015-AGRICECON

ELM model. The empirical results on the hog price series in China for short-term forecasting show that the proposed improved EEMD-based hybrid approach outperforms the selected counterparts, indicating that it is a promising alternative for the short-term hog price forecasting.

The limitations of this study lie in two aspects. First, although we have examined the fine-to-coarse reconstruction algorithm for the purpose of composition, there are many other possible algorithms in composing the IMFs and the residue, which may shed a different light on the modelling issue. Second, our experimental study focuses on the iterated strategy for the short-term forecasting. A further research is needed to investigate the performance of the short-term time series prediction with richer strategies.

REFERENCES

Adanacioglu H., Yercan M. (2012): An analysis of tomato prices at wholesale level in Turkey: an application of SARIMA model. Custos e@ gronegócio on line. 8: 52–75.

Box G.E., Jenkins G.M. (1976): Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Chang C., Lin C. (2011): LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2: 27.

Chen C., Lai M., Yeh C. (2012): Forecasting tourism demand based on empirical mode decomposition and neural network. Knowledge-based Systems, 26: 281–287.

Diebold F.X., Mariano R.S. (1995): Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 20: 134–144.

Felipe I.J., Mol A.L., Almeida V. (2012): Application of ARIMA models in soybean series of prices in the north of Paraná. Custos e@ gronegócio Online, 8: 78–91.

Guo Z., Zhao W., Lu H., Wang J. (2012): Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renewable Energy, 37: 241–249.

Hahn W.F. (2004): Beef and pork values and price spreads explained. Electronic Outlook Report from Eonomic Research Service, USDA.

Huang G., Babri H. (1998): Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. IEEE Transactions on Neural Networks, 9: 224–229.

Huang G. (2003): Learning capability and storage capacity of two-hidden-layer feedforward networks. IEEE Transactions on Neural Networks, 14: 274–281.

Huang G., Chen L., Siew C.-K. (2006): Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Transactions on Neural Networks, 17: 879–892.

Huang G., Zhu Q., Siew C. (2006): Extreme learning machine: theory and applications. Neurocomputing, 70: 489–501.

Huang N.E., Shen Z., Long S.R., Wu M.C., Shih H.H., Zheng Q., Yen N.-C., Tung C.C., Liu H.H.: (1998): The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. In: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences. The Royal Society: 903–995.

Hyndman R.J., Khandakar Y. (2007): Automatic Time Series for Forecasting: the Forecast Package for R. Department of Econometrics and Business Statistics, Monash University.

Jha G.K., 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: 563–571.

Jumah A., Kunst R.M. (2008): Seasonal prediction of European cereal prices: good forecasts using bad models? Journal of Forecasting, 27: 391–406.

Kisi O., Latifoğlu L., Latifoğlu F. (2014): Investigation of empirical mode decomposition in forecasting of hydrological time series. Water Resources Management, 28: 4045–4057.

Li G., Xu S., Li Z., Sun Y., Dong X. (2012): Using quantile regression approach to analyze price movements of agricultural products in china. Journal of Integrative Agriculture, 11: 674–683.

Li Z., Cui L., Xu S., Weng L., Dong X., Li G., Yu H. (2013): Prediction model of weekly retail price for eggs based on chaotic neural network. Journal of integrative agriculture, 12: 2292–2299.

Martín-Rodríguez G., Cáceres-Hernández J.J. (2012): Forecasting pseudo-periodic seasonal patterns in agricultural prices. Agricultural Economics, 43: 531–544.

Napolitano G., Serinaldi F., See L. (2011): Impact of EMD decomposition and random initialisation of weights in ANN hindcasting of daily stream flow series: An empirical examination. Journal of Hydrology, 406: 199–214.

Paul R.K., Gurung B., Paul A. (2015): Modelling and forecasting of retail price of arhar dal in Karnal, Haryana. The Indian Journal of Agricultural Sciences, 85.

Ramirez O.A., Fadiga M. (2003): Forecasting agricultural commodity prices with asymmetric-error GARCH models. Journal of Agricultural and Resource Economics, 28: 71–85.

doi: 10.17221/268/2015-AGRICECON

Ribeiro C.O., Oliveira S.M. (2011): A hybrid commodity price – forecasting model applied to the sugar – alcohol sector. Australian Journal of Agricultural and Resource Economics. 55: 180–198.

Saengwong S., Jatuporn C., Roan S. (2012): An analysis of Taiwanese livestock prices: empirical time series approaches. Journal of Animal and Veterinary Advances. 11: 4340–4346.

Sharma A. (2000): Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 – A strategy for system predictor identification. Journal of Hydrology, 239: 232–239.

Shih M., Huang B., Chiu N.-H., Chiu C., Hu W. (2009): Farm price prediction using case-based reasoning approach – A case of broiler industry in Taiwan. Computers and electronics in agriculture. 66: 70–75.

Shrivastava N.A., Panigrahi B.K. (2014): A hybrid wavelet-ELM based short term price forecasting for electricity markets. International Journal of Electrical Power & Energy Systems, 55: 41–50.

Su X., Wang Y., Duan S., Ma J. (2014): Detecting chaos from agricultural product price time series. Entropy, 16: 6415–6433.

Taieb S.B., Bontempi G., Atiya A.F., Sorjamaa A. (2012): A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert Systems With Applications, 39: 7067–7083.

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

Xiong T., Bao Y., Hu Z. (2014): Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting. Knowledge-based Systems. 55: 87–100.

Xiong T., Li C., Bao Y., Hu Z., Zhang L. (2015): A combination method for interval forecasting of agricultural commodity futures prices. Knowledge-based Systems, 77: 92–102.

Yu L., Wang S., Lai K.K. (2008): Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Economics, 30: 2623–2635.

Yu L., Wang Z., Tang L. (2015): A decomposition–ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Applied Energy, 156: 251–267.

Yu L., Dai W., Tang L. (2016): A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. Engineering Applications of Artificial Intelligence, 47: 110–121.

Zhang X., Hu T., Brain R., Fu Z. (2005): A forecasting support system for aquatic products price in China. Expert Systems With Applications, 28: 119–126.

Zhang X., Lai K.K., Wang S.-Y. (2008): A new approach for crude oil price analysis based on empirical mode decomposition. Energy Economics, 30: 905–918.

Zhou Q., Jiang H., Wang J., Zhou J. (2014): A hybrid model for PM 2.5 forecasting based on ensemble empirical mode decomposition and a general regression neural network. Science of the Total Environment, 496: 264–274.

Zhu Q., Qin A.K., Suganthan P.N., Huang G. (2005): Evolutionary extreme learning machine. Pattern Recognition, 38: 1759–1763.

Received November 9, 2015
Accepted June 24, 2016
Published online January 12, 2017