An ensemble approach based on transformation functions for natural gas price forecasting considering optimal time delays

Faramarz Saghi [1], **Mustafa Jahangoshai Rezaee** [Corresp. 1]

[1] Faculty of Industrial Engineering, Urmia University of Technology, Urmia, Iran

Corresponding Author: Mustafa Jahangoshai Rezaee
Email address: m.jahangoshai@uut.ac.ir

Natural gas, known as the cleanest fossil fuel, plays a vital role in the economies of producing and consuming countries. Understanding and tracking the drivers of natural gas prices are of significant interest to the many economic sectors. Hence, accurately forecasting the price is very important not only for providing an effective factor for implementing energy policy but also for playing an extremely significant role in government strategic planning. The purpose of this study is to provide an approach to forecast the natural gas price. First, optimal time delays are identified by a new approach based on the Euclidean Distance between input and target vectors. Then, wavelet decomposition has been implemented to reduce noise. Moreover, fuzzy transform with different membership functions has been used for modeling uncertainty in time series. The wavelet decomposition and fuzzy transform have been integrated into the preprocessing stage. An ensemble method is used for integrating the outputs of various neural networks. The results depict that the proposed preprocessing methods used in this paper cause to improve the accuracy of natural gas price forecasting and consider uncertainty in time series.

1 # An ensemble approach based on transformation functions for natural gas
2 # price forecasting considering optimal time delays

3 **Faramarz Saghi**
4 Faculty of Industrial Engineering, Urmia University of Technology, Urmia, Iran
5 (Email: f.saghi@ine.uut.ac.ir)
6
7 **Mustafa Jahangoshai Rezaee**
8 Faculty of Industrial Engineering, Urmia University of Technology, Urmia, Iran (Corresponding
9 Author, Email: m.jahangoshai@uut.ac.ir), Tel/Fax: +98-44-31980265, orcid: 0000-0002-3340-
10 666X

13 **Abstract**

14 Natural gas, known as the cleanest fossil fuel, plays a vital role in the economies of producing and
15 consuming countries. Understanding and tracking the drivers of natural gas prices are of significant
16 interest to the many economic sectors. Hence, accurately forecasting the price is very important
17 not only for providing an effective factor for implementing energy policy but also for playing an
18 extremely significant role in government strategic planning. The purpose of this study is to provide
19 an approach to forecast the natural gas price. First, optimal time delays are identified by a new
20 approach based on the Euclidean Distance between input and target vectors. Then, wavelet
21 decomposition has been implemented to reduce noise. Moreover, fuzzy transform with different
22 membership functions has been used for modeling uncertainty in time series. The wavelet
23 decomposition and fuzzy transform have been integrated into the preprocessing stage. An
24 ensemble method is used for integrating the outputs of various neural networks. The results depict
25 that the proposed preprocessing methods used in this paper cause to improve the accuracy of
26 natural gas price forecasting and consider uncertainty in time series.
27 **Keywords:** Time Series Forecasting; Ensemble Artificial Neural Networks, Machine Learning;
28 Natural Gas Price; Optimal Time Delays

31 # 1. Introduction

32 In most financial matters, the current data is usually affected by past data modeled in time series,
33 such as time series of natural gas price. These past data must be reliable to obtain acceptable results.
34 For example, in a sample of time series, the observation value of $t\text{-}1^{st}$ must be specified to predict
35 the value of t. Also, the observation value of $t\text{-}2^{st}$ must be specified to predict the value of t-1st and
36 so on. The observation value of $t\text{-}2$ is called a time delay for t-1st period with 1 step.
37 The contributions of this study are divided into three parts. First, a new approach is presented to
38 identify optimal time delays. In previous studies, the delays were identified based on the

39 experience of experts. On the other hand, gas price is influenced by political, economic, social,
40 etc., factors that include uncertainty in values. In the second contribution, the fuzzy transform with
41 different membership functions is proposed to model the uncertainties in time series. Finally, in
42 the third contribution, the wavelet decomposition and fuzzy transform are combined for improving
43 the accuracy of forecasting and uncertainty modeling that this type of integration has not been
44 provided in the previous studies. Integrating the mentioned proposed algorithms with various
45 artificial neural networks (ANNs) are used to forecast time series of natural gas price.
46 The rest of the paper is organized as follows: A literature review of forecasting in different areas
47 is provided in Section 2. The methodologies used in this paper are presented in Section 3. Section
48 4 provides the proposed approach and related subsystems. Section 5 presents a case study, data,
49 results, and related analyses. Section 6 highlights the limitations of the proposed approach in real-
50 world forecasting problems. Finally, the summary and conclusion of the research are given in
51 Section 7.
52

53 2. Literature Review

54 Different kinds of financial time series have recently become a popular field for researchers. In the
55 following, some studies in financial areas are presented. A new approach for integrating RW and
56 ANN was proposed that employs the advantages of RW and ANN. The RW model was performed
57 for the linear section, and the rest of the data or nonlinear section was modeled by feed-forward
58 ANN and Elman ANN. The results show which models are better for forecasting in terms of
59 accuracy and performance [1]. In another investigation, the various machine learning methods
60 have been used for modeling and forecasting the BUX stock index. The results showed that the
61 generalized autoregressive conditional heteroscedasticity (GARCH) model is better than the other
62 ones [2]. An exponent back propagation neural network has been improved to forecast the cross-
63 correlation between two financial data [3].
64 Lin et al. [4] proposed a new method for forecasting of Nasdaq Composite Index, which is based
65 on integrating phase space reconstruction (PSR), empirical mode decomposition (EMD), and
66 neural networks optimized by particle swarm optimization (PSO). The empirical results have
67 represented which the introduced PSR-EMD-NNPSO model has provided the best forecasting
68 stock index. Hota et al. [5] provided a robust forecasting model using the RBF and error
69 backpropagation network (EBPN) for BSE 30 and INR/USD Foreign Exchange. The results
70 showed the RBF network is better than EBPN. Wang et al. [6] introduced a new approach to
71 integrating EMD with a stochastic time-strength neural network (STNN). The mentioned method
72 has been used to measure the forecasted results of NYSE, DAY, FTSE, and HSI stock market
73 indexes. Kusy et al. [7] proposed a new algorithm for the structure reduction of the probabilistic
74 neural network (PNN). The calculations have been applied fuzzy C-means (FCMs) for choosing
75 PNN's pattern neurons based on the derived centroids. Sukeno et al. [8] developed a neural
76 network model and applied it for inflation data in Indonesia. They optimized the development of

77 a feed-forward neural network model. In the following, some studies are presented about the
78 integration of wavelet decomposition with ANNs.
79 Bai et al. [9] integrated wavelet decomposition with a backpropagation neural network (BPNN)
80 for improving the accuracy of air pollutants forecasting. Wavelet decomposition was applied to
81 decompose original data to different levels. Then, each level was considered input of BPNN. Yu
82 et al. [10] proposed a hybrid approach for short-term wind speed forecasting by the wavelet packet
83 decomposition (WPD), density-based spatial clustering of applications with noise (DBSCAN), and
84 the Elman neural network (ENN). The results showed the WPD-DBSCAN-ENN model was better
85 than the others. Ji et al. [11] introduced an approach based on the wavelet neural network for short-
86 term forecasting of the parking guidance information systems problem. Also, an integration
87 wavelet decomposition and ANN are presented for daily streamflow forecasting. The results
88 showed WA hybrid models are better than classic ANNs [12]. In another research, Dadkhah et al.
89 [13] provided an approach integrating clustering and classification algorithms for predicting short-
90 term power output forecasting of hourly operation based on climate factors. Also, they examined
91 the effects of wind direction and wind speed on the accuracy of prediction results. In the following,
92 some studies about using fuzzy models for forecasting time series are presented.
93 An adaptive fuzzy inference system was proposed for predicting the power plant output by
94 considering various climate factors [14]. The authors have been applied metaheuristic algorithms
95 for increasing the accuracy and performance of the designed system. In another study, a hybrid
96 model has been used for chaotic time series prediction. An approach has been proposed for training
97 second-order Takagi- Sugeno- Kang fuzzy systems using an adaptive neuro-fuzzy inference
98 system (ANFIS) [15]. A similar study for forecasting time series using a robust fuzzy time series
99 model may be found in Yolcu et al. [16]. In another research, for forecasting air pollution in
100 Turkey, a novel fuzzy time series model has been used based on the Fuzzy K-Medoid clustering
101 [17]. The proposed method handled outliers in time series. Wang et al. [18] applied fuzzy logic to
102 forecast air pollution in China. They combined the fuzzy time series forecasting method and data
103 reprocessing approaches for forecasting main air pollutants.
104 Using new artificial neural networks are another category of a forecasting method. Wang et al.
105 [19] introduced a semi-heterogeneous approach to forecast crude oil prices. The mentioned
106 approach applied decomposed techniques, including wavelet decomposition, singular spectral
107 analysis, empirical mode decomposition, and variation mode decomposition in the preprocessing
108 stage. The autoregressive, ARIMA, support vector regression (SVR), and ANN have been used to
109 forecast. Finally, the results of forecasting have been integrated by a new combination approach.
110 Agrawal et al. [20] provided a new ensemble approach based on relevance vector machines (RVM)
111 and boosted trees. In the first stage, the RVM has been used based on Gaussian RBF and
112 polynomial kernels. The performance of the model has been improved by Extreme Gradient
113 Boosting. The outputs of each model have been integrated using Elastic net regression. Rezaee et
114 al. [21] provided an approach to predict the stock market using dynamic fuzzy C-means, Data
115 Envelopment Analysis, and multilayer perceptron (MLP). In another study, a novel ensemble

116 method has been introduced for time series forecasting integrating various machine learning
117 models [22]. Finally, the MLP has been integrated with inverse neural network and mathematical
118 programming for determining optimal daily amounts of fuel consumption used by a power plant
119 [23].

120 ## 3. Methodology

121 ### 3.1 Wavelet Transform

122 In computational and functional analysis, a discrete wavelet transform (DWT) is a general WT for
123 which the wavelets are discretely sampled. In this paper, a type of DWT family known as
124 Daubechies with order 18 (Daubechies-18) has been used in the preprocessing stage [24].

125 $$a = 2^j \quad j \in Z \quad and \quad b = ka \quad k \in Z \qquad\qquad (1)$$

126 Where $a = 2^j$ is the scale Parameter and k is the time-translation parameter. Therefore, the discrete
127 wavelet is expressed as follows:

128 $$\varphi_{j,k}(t) = 2^{\frac{-j}{2}} \varphi(2^{-j}t - k) \qquad\qquad (2)$$

129 It gives the range of j and k as $0 < k < 2^{J-j} - 1$ and $1 < j < J$.

130 ### 3.2 Fuzzy Transform

131 The basic idea of fuzzy transformation is to transform an original space into a special space for
132 discovering information. It can use a continuous function on a fixed distance, $[a,b] \in R$ (transforming
133 the fuzzy space of functions on an open space of real numbers by a direct transformation into the real
134 n-dimensional vectors). Their inverse conversion returns the derived n-vector to the initial or
135 approximate function of it.
136 Let $x_1 < x_2 < \cdots < x_n$ be fixed nodes in [a, b], such that $x_1 = a, x_n = b$ and fuzzy sets $A_1, A_2, \cdots, A_n$
137 have their membership functions $A_1(x), A_2(x), \cdots, A_n(x)$ defined on [a, b], form a fuzzy partition if
138 they satisfy the following conditions ($k = 1, 2, \cdots, n$):

139 1. $A_k : [a, b] \rightarrow [0,1], A_k(x_k) = 1,$

140 2. $A_k(x_k) = 0 \; if \; x \notin (x_{k-1}, x_{k+1})$ where $x_1 = a$ and $x_n = b$

141 3. $A_k(x)$ is continuous,

142 4. $A_k(x), k = 2,3, \cdots n$ are strictly increasing on $[x_{k-1}, x_k]$ and $A_k(x), k = 1,2, \cdots, n-1$

143 5. $\forall x \quad x \in [a, b] \quad \sum_{k=1}^{n} A_k(x) = 1 \qquad\qquad (3)$

144 6. $A_k(x_k - x) = A_k(x_k + x)$, for all $x \in [0, h], k = 2, \cdots, n-1$ where all $x_1, x_2, \cdots, x_n$ are
145 equidistant, i.e., $x_k = a + h(k-1), k = 1, \cdots, n, \; h = (b-a)/(n-1)$.

146 7. $A_k(x) = A_{k-1}(x-h)$, For all $k = 2,3,\cdots, n-1$ and $\in [x_k, x_{k+1}]$, and $A_{k+1}(x) = A_k(x-h)$,

147 for all $k = 2, \cdots, n-1$ and $x \in [x_k, x_{k+1}]$ where $h = (b-a)/(n-1)$.

148 Where h is the length of support of $A_1 \ldots A_n$ and $2h$ is the support of other basic functions A_k,

149 $k = 2,3, \cdots, n-1$ in case of uniform partition. The membership functions $A_1, A_2, \cdots, A_n$ are called

150 basic functions.

151 Let $C([a, b])$ be the set of continuous functions on the domain [a, b], $A_1, A_2, \cdots, A_n$ are basic

152 functions on $[a, b]$ and let f be any function in $C([a, b])$. Then n-tuple vector of real numbers of $[F_1$

153 , $F_2, \cdots, F_n]$ is defined by Equation (4):

154
$$F_k = \frac{\int_a^b f(x)A_k(x)dx}{\int_a^b A_k(x)dx}, \quad k = 1,2, \cdots, n. \tag{4}$$

155 Equation (4) is called the F-transform f concerning $A_1, A_2, \cdots, A_n$. In case of discrete F-transform,

156 f be given at nodes $P_1, P_2, \cdots, P_l$ and $A_1, A_2, \cdots, A_n$, $n < l$ are basic functions which form a fuzzy

157 partition on $[a, b]$, then, $[F_1, F_2, \cdots, F_n]$ are given by the following formula:

158
$$F_k = \frac{\left(\sum_{j=1}^l f(P_j)A_k(P_j)\right)}{\left(\sum_{j=1}^l A_k(P_j)\right)} \tag{5}$$

159 One of the most important issues in fuzzy systems is membership functions. These functions

160 specify the membership value of variables. They can be defined as either discrete or continuous.

161 If a defined variable to be is a discrete variable, the discrete membership function is used. If the

162 variable being defined is continuous, the continuous membership function is defined. For

163 continuous variables, membership functions may be either mathematical or real analytic functions.

164 The fuzzy transform functions used in this research are provided in Table 1.

165   Table 1. The membership functions used for considering uncertainty in time series

166

167 **3.3 Multilayer Perceptron**

168 MLP is the most famous ANN applied in time series forecasting problems. A multilayer

169 perceptron is essentially a feed-forward structure of the entrance, one or more hidden layers, and

170 one output layer. There are some nodes on each layer that are linked through irreversible links to

171   a forwarding layer. The relationship between inputs and outputs in a multilayer perceptron
172   network is given by the following formula:

173
$$y_t = G\left(a_0 + \sum_{j=1}^{h} F\left(\alpha_j \beta_{0j} + \Sigma_{i=1}^{p} \beta_{ij} \times y_{t-i}\right)\right) \qquad (10)$$

174   Where α_j and β_{ij} ($i = 1, 2, ..., p; j = 1, 2, ..., h$) are the alliance weights, α_0 and β_{0j} are the bias,

175   and F, G are the hidden and output layer activation functions, respectively. In the MLP, each input

176   data (y_1, y_2, y_{t-p}) is first multiplied in relative weights ($\beta_{11}, \beta_{21}, \beta_{p1}$), then it is added by β_{01}, multiple

177   in α_j and enters into the transfer function (F). The output of the transformation function is found

178   by multiplying the weights with their corresponding inputs, adding the bias (α_0), and entering into

179   the activation function (G). This calculation will keep on until $j = h$. h is the number of hidden

180   layer neurons. The most common activation function in the MLP is the sigmoid function.

181

182   **3.4 Radial Basis Function**

183   RBF operates radial-base function as the activation function. The output of this network is a linear

184   mixture of radial-base functions for input parameters and neurons. RBF networks are used in

185   function approximation, time series forecasting, classification, and control of the system. RBF

186   networks have a three-layer structure. The initial layer is described as the input layer, the next

187   layer is described as the hidden layer, and the latest layer is described as the output layer. The RBF

188   describes the potential of jth neuron in the layer on the ground of varying of the Euclidean distance

189   obtained by the vector:

190
$$u^j = \left\| x - w^j \right\|^2 \text{, for } j = 1, 2, \cdots, s \qquad (11)$$

191   Where x is the k-dimensional input vector, and w^j represents the weights in the hidden layer. The

192   RBF neural network uses varying species of activation functions famous on the ground of

193   Gaussian or RBF. The activation function for jth hidden neuron is described as:

194
$$\psi_2(u^j) = e^{\frac{-u^j}{2\sigma_j^2}} = e^{\frac{\|x - w^j\|^2}{2\sigma_j^2}} \text{, for } j = 1, 2, \cdots, s \qquad (12)$$

195   Where σ_j^2 is the variance of the inward time series. $\left\| x - w^j \right\|$ is the Euclidean norm, w^j and σ stand

196   for the center and variance of Gaussian function, respectively. $x = \left\{ x_1^p, x_2^p, ..., x_m^p \right\}$ is the pth input

197   vector. If the time series data in the entrance vectors are not erect, hence the activation function

198   makes Equation (13).

199
$$\psi_2(u^j) = e^{\frac{-(x-w^j)^T \Sigma^{-1}(x-w^j)}{2}} \qquad (13)$$

200   Where Σ^{-1} is the reverse of the variance-covariance matrix for the entrance vectors of time series.

201   **3.5 Group method of data handling**

202 Group method of data handling (GMDH) has some of the layers, and each layer has a number of
203 neurons. All neurons utilize contiguous structures in which have two entrances and one output.
204 Without loss of generality, suppose for each neuron, five weights, and one bias to bring about
205 processing function between input and output time series. Therefore, we have:

206 $$y_{ik}^* = N(x_{i\alpha}, x_{i\beta}) = b^k + w_1^k x_{i\alpha} + w_2^k x_{i\beta} + w_3^k x_{i\alpha}^2 + w_4^k x_{i\beta}^2 + w_5^k x_{i\alpha} x_{i\beta} \qquad (14)$$

207 Where $(i = 1, 2, 3, \cdots, N)$, N is the number of input and output data, $(K = 1, 2, 3, \cdots, C_m^2)$, $\alpha, \beta \in$
208 $\{1, 2, 3, \cdots, m\}$, m is the number of the preceding layer weights computed by MSE and therefore
209 obvious and stable values are substituted in each neuron. The number of candidates of new neuron
210 foundation is calculated by $C_m^2 = \frac{m(m-1)}{2}$.
211 The GMDH algorithm is capable of constructing models for complex systems with a high degree
212 of regression. The criterion for selecting the optimal neural network model is to minimize the
213 relationship (15). A model with less prediction error is considered a desirable model. In the
214 relationship (15), n is the number of observations.

$$E = \frac{\sum_{i=1}^{n} (y_i^{forecast} - y_i^{actual})^2}{n} \qquad (15)$$

215 ### 3.6 Adaptive Neuro-Fuzzy Inference System

216 ANFIS was developed in 1990. Fuzzy Inference System (FIS) is based on the nonlinear approach
217 that delineates the input/output relationships of an actual system using a set of if-then rules. In this
218 neural network, the Takagi-Sugeno species FIS has been used. The output of each rule is a linear
219 mixture of entrance variables. The ultimate output is the average weighted output of each rule.
220 Theoretically, ANFIS models include five primary sections: Input and output database, a fuzzy
221 model creator, an FIS, and a consistent ANN. ANFIS eliminates the initial problem in the design
222 of the fuzzy system, the parameters of the membership function, and the design of the if-then rules.
223 The ability to train ANNs are used for creating automated fuzzy rules and parameters optimization.

224 ### 3.7 Support Vector Regression

225 Support vector machines (SVMs) were first used to solve classification problems. Then, it was
226 developed for regression problems in 1992. Support vector machines associated with modeling
227 and prediction are called support vector regression. In the SVR, the activation functions are first
228 mapped on the input vectors, and the kernel functions perform (.) product between the outputs of
229 these mapped vectors. The original SVR formulation is:
230 $$f(x) = w\varphi(x) + b \qquad (16)$$

231 Where $f(x)$ is the evaluated model output, w is a weight vector, and b is a bias period. The vector
232 w is a component of the characteristic space of the issue. Hence the issue is to select the straight
233 nonlinear maps $\varphi(x)$, and the fitting for excellence RBF is one of the grossest used nonlinear

234   kernels for SVR models and is performed in this paper. In SVR, first, the input support vectors (
235   x_1, x_2, x_N) and x_{test} are considered inputs of the activation functions. Then, activation functions are
236   mapped on the input vectors. The output of the activation function for each input vector is
237   multiplied by the output value of the x_{test} activation function, and each result is multiplied by the
238   corresponding weight ($\alpha_1, \alpha_2, \ldots, \alpha_N$) and enters into the transformation function. The output of
239   the transformation function in the form of a linear combination is aggregated with a bias (b), and
240   the output is derived from the SVR model.

3.8 Ensemble forecasting model

242   Preceding studies have shown which combination models commonly lead to distinct forecasts. An
243   ensemble method is referred to as multiple forecasts and handles linear or nonlinear integrations
244   of the forecasting results, direct to a coherence forecast. This research checks an ensemble method
245   to combination forecasting of natural gas price, where the forecasts are established from five
246   models, including decomposition and transformation techniques. If there are *m* types of
247   forecasting models for solving a special problem, the outcomes of them are added together.
248   Suppose that:

$$Combination_{model}(Model = MLP, RBF, GMDH, ANFIS, SVR)$$

250   Where $Combination_{model}$ is the forecast combination of each model, the output of the forecast
251   combination model based on five individual models can be displayed as:

252

$$\text{P} = f(y^{forecast}_1, y^{forecast}_2, \cdots, y^{forecast}_m) \tag{17}$$

254   Chu et al. [25], according to a study at Stanford University, computed and evaluated the
255   complexity of various Machine Learning algorithms on Multicore CPUs. Table 2 shows the
256   complexity of the algorithms used in this paper based on the mentioned study (updated in some
257   cases). The complexities may be changed according to the number of layers, the number of
258   neurons, epochs/iterations, or other hyperparameters of each algorithm.

259   Table 2: Complexity analysis of algorithms used in this paper

260   Where *m ,n, p,* and *l* are the number of data/batches, the number of neurons in each layer, number
261   of features, and the number of rules, respectively.

4. Proposed approach

263   Time-series data of natural gas price are gathered, which have 3720 records, with daily data from
264   2004/09/09 to 2019/06/10, which were obtained from http://finance.yahoo.com. The descriptive
265   statistics including the measures for natural gas daily prices: Minimum=1.49, Mean=4.6451,

266 Maximum=15.39, Median=3.86, SD=2.3406, Skewness=1.5064, Kurtosis=5.3989. Time delays
267 play an important role in increasing forecasting accuracy. Determining the optimal time delay has
268 a great impact on the training process of models in time series forecasting. If optimal time delays
269 are not selected, the training process is hampered, and neural networks perform the learning
270 process with low accuracy. On the other hand, overfitting occurs when the network does the
271 learning process more than necessary and the network is not able to forecast appropriately new
272 data. For handling this problem, the models should adjust hyper-parameters. The values of all
273 hyper-parameters such as learning rate, epochs, the number of inputs membership functions in
274 ANFIS, the number of clusters in ANFIS, max layer neurons and alpha parameters in GMDH,
275 hidden layer size in MLP, spread and max neurons in RBF, and epsilon and sigma in SVR are
276 determined by various methods proposed in articles about this issue. Various values are examined
277 to obtain the best conclusions in terms of the accuracy and performance of ANNs for test data
278 (unseen data).

279 Therefore, in the preprocessing stage, a new method has been proposed to identify optimal time
280 delays for time series of natural gas price. The steps of the new approach to identify optimal time
281 delays of time series are expressed as follows:

282   Step 1: Determining input and target vectors for each delay.

283   Step 2: Normalizing input and output vectors.
284   Step 3: Calculating the Euclidean distance between input and output vectors for each
285   delay.
286   Step 4: Calculating the ratio between the consecutive distances in pairs.
287   Step 5: Determining larger ratios as optimal delays.

288 Then, wavelet decomposition is used to study natural gas prices and reduce noise and increase
289 forecasting accuracy. The mentioned time series data were decomposed by wavelet decomposition
290 into four levels as the best number of levels (*w1, w2, w3, w4*).
291 To model the uncertainty, the fuzzy transforms with different membership functions are used.
292 Time-series are inherently uncertain which the fuzzy approach is used for importing uncertainty
293 to the system. In a fuzzy system, membership functions play an important role. Therefore, in this
294 study, five different and well-known membership functions are used in a fuzzy transform called
295 Triangular type 1, Gaussian, Trapezoidal, Triangular type 2, and Bell-shaped. F-transform is
296 applied for decomposing the original time series, and therefore, the decomposed data are used as
297 inputs (*F1, F2, F3, F4, F5, F6*). Figure 1 shows the results of the fuzzy transformation with the
298 Triangular type 1 membership function for the time series of natural gas prices.

299   **Figure 1. Levels of fuzzy transform with Triangular type 1 membership function**

300   Afterward, the decomposition techniques and combinations are integrated with MLP, RBF,
301   GMDH, ANFIS, and SVR. The efficiency and accuracy of the neural networks are evaluated by
302   the RMSE and *R* (correlation coefficient). In summary, this study uses wavelet decomposition and
303   fuzzy transform in the preprocessing phase. Also, MLP, RBF, GMDH, ANFIS, and SVR are
304   applied for forecasting the natural gas price. Finally, an ensemble method combines the results.
305   These steps are shown in Figure 2.

306   **Figure 2. Flowchart of preprocessing and forecasting steps**

307   ## 5. Data and results

308   Energy is supplied by various carriers such as oil, gas, etc. According to the effective role in the
309   development and economics of countries and due to the limited resources, price forecasting is an
310   essential tool for the short-term energy market. In this study, for all models, 70% and 30% of data
311   are considered the train and test data, respectively. Also, optimal time delays are identified
312   according to the new method presented in the previous section. According to the calculations,
313   delay 1 has the largest ratio and is considered the optimal delay. Then, other delays with higher
314   ratios are added to the delay 1 respectively, until there is no improvement in the neural network
315   performance.
316   Finally, delays 1, 2, and 5 are considered optimal delays for the natural gas price time series. The
317   identified delays are rational because the current day price is related to one, two, and five working
318   days ago. The same delays are used in all calculations for the gas price time series.
319   First, the original time series is considered the input of neural networks without wavelet
320   decomposition and fuzzy transformation, with time delays of 1, 2, and 5. In this study, the
321   correlation coefficient (*R*) is a criterion that indicates the severity of the relationship as well as the
322   type of relationship (direct or reverse) between the two set variables (targets and outputs).
323   According to the results, the RBF performs the best result for the original natural gas price time
324   series. According to the results in Figure 3, the root-mean-square error (RMSE) of the RBF for
325   time series is 0.1483, and the correlation coefficient is 0.95731.

326   **Figure 3. The results of the RBF neural network for the original time series of gas price**

327   In the next examination, the wavelet decomposition and various neural networks are integrated.

328   **Table 3. Calculating error and correlation coefficient for time series applied wavelet**
329   **decomposition**

330   In Table 3, the original time series of natural gas price is decomposed by wavelet decomposition
331   into four levels. The levels (*w1-w2-w3-w4*) are considered the input of neural networks. Table 3
332   shows the results of implementing neural networks for four levels. The sum of noises that each
333   wavelet level has derived from the original data is calculated. The sum of noises for Levels 1 to 4

334   is equal to 203.6893, 280.7320, 330.8307, and 435.7385, respectively. Also, the sum of the
335   cumulative noise taken up to Level 4 is 1251. In Table 3, 20 combinations are evaluated. The W4-
336   MLP combination has the best performance. However, its performance is very little different from
337   the performance of other networks. According to Figure 4, the root-mean-square error of the W4-
338   MLP is 0.061674, and the correlation coefficient is 1.

339   **Figure 4. The results of the MLP neural network for Level 4 obtained from the wavelet**
340   **decomposition**

341   **Table 4. Calculation of noise between original data and decomposed data by wavelet**
342   **transform**

343   For reducing noise and increasing the accuracy, the natural gas price time series was decomposed
344   by wavelet transform into four levels. According to Table 4, it should be determined which of
345   these levels are more similar to the original data. In this study, to examine the behavior and
346   performance of the presented algorithms, the noise between original and decomposed data has
347   been calculated and presented in Table 4. Calculating noise between original data and decomposed
348   data has been used by two criteria (RMSE and R). Less root-mean-square error and high correlation
349   coefficient mean less difference between the original data and decomposed data. Of course, it does
350   not mean which the decomposed data inside itself has less noise and swing. Table 4 compares the
351   calculated values by RMSE and R. It depicts W1-ANFIS has the most similarity to the original
352   time series. In other words, the W1-ANFIS has the least noise toward the original data and has a
353   closer behavior to it. Also, W1-MLP, W1-GMDH, and W1-RBF have similar behavior to the
354   original time series. In the next stage, the fuzzy neural networks with different membership
355   functions are used (see Figure 2). According to this figure, first, the original time series is
356   decomposed into six levels by fuzzy transform with five membership functions, then the fuzzy
357   levels (*F1-F2-F3-F4-F5-F6*) are considered inputs of each model.

358   **Table 5. Calculating error and correlation coefficient of time series by fuzzy neural**
359   **networks with various membership functions**

360   In Table 5 (second and third columns), the Triangular type 1 membership function is used. They
361   show the results of implementing neural networks for six fuzzy levels with Triangular type 1
362   membership function. In these columns, 30 combinations are considered. According to the
363   mentioned columns, the MLP has the best performance for Level 6 of the fuzzy transform. The
364   root-mean-square error of the MLP for Level 6 is 0.00010555, and the correlation coefficient is
365   0.99949.

366   **Figure 5. The results of MLP for Level 6 obtained from fuzzy transform with Triangular**
367   **type 1 membership function**

368   In the fourth and fifth columns of Table 5, the Gaussian membership function is used in the fuzzy
369   transform structure. They represent the results of implementing neural networks for six fuzzy

370 levels with the Gaussian membership function. In these columns, 30 combinations are evaluated.
371 According to the results, the SVR has the best performance for Level 1. According to Figure 6, the
372 root-mean-square error of the SVR for Level 1 is 0.0002834, and the correlation coefficient is
373 0.95448.

374 **Figure 6. The results of SVR for Level 1 obtained from fuzzy transform with Gaussian**
375 **membership function**

376 In the sixth and seventh columns of Table 5, the Trapezoidal membership function is used in the
377 fuzzy transform structure. They depict the performance of neural networks for six fuzzy levels
378 with the Trapezoidal membership function. In these columns, the time series is forecasted based
379 on 30 different combinations. According to the results, the ANFIS has the best performance for
380 Level 6. Figure 7 presents the root-mean-square error and the correlation coefficient for Level 6
381 with values 0.00063896, and 0.9625 respectively.

382 **Figure 7. The results of ANFIS for Level 6 obtained from fuzzy transform with**
383 **Trapezoidal membership function**

384 According to the eighth and ninth columns of Table 5, the Triangular type 2 membership function
385 is used in the fuzzy transform structure. They show the results of implementing neural networks
386 for six fuzzy levels with Triangular type 2 membership function. In these columns, 30 different
387 combinations are evaluated. SVR has the best performance for Level 3. According to Figure 8, the
388 root-mean-square error of the SVR for Level 3 is 0.00016989, and the correlation coefficient is
389 0.95957.

390 **Figure 8. The results of SVR for Level 3 obtained from fuzzy transform with Triangular**
391 **type 2 membership function**

392 According to Table 5 (10th and 11th columns), the Bell-shaped membership function is used. These
393 columns present the results of implementing neural networks for six fuzzy levels with Bell-shaped
394 membership function. According to this table, 30 combinations are considered. The SVR has the
395 best performance for Level 3 of the fuzzy transform. Figure 9 shows the root-mean-square error
396 and the correlation coefficient for Level 3 with values of 0.000046212 and 0.98144.

397 **Figure 9. The results of SVR for Level 3 obtained from fuzzy transform with Bell-shaped**
398 **membership function**

399 In the final stage, the combination of wavelet decomposition and fuzzy transformation is done.
400 First, the original time series is decomposed by wavelet transform into four levels. Then, the levels
401 decomposed by wavelet are considered the input of fuzzy transform with different membership
402 functions and are subdivided into six levels. Continuing, these compounds are considered the input
403 of neural networks.

404 **Table 6. Calculating error and correlation coefficient of time series by wavelet-fuzzy**
405 **transformation with five membership functions**

406 In each wavelet-fuzzy combination with a distinct membership function, 120 different
407 combinations are evaluated. Table 6 presents the best combination in terms of performance and
408 accuracy. In this stage, 600 different combinations are made to forecast the natural gas price. In
409 the combination of wavelet and fuzzy transform with the Triangular type 1 membership function,
410 W1-F5-RBF is the best performing among 120 compounds. Combined with the Gaussian,
411 Trapezoidal, Triangular type 2, and Bell-shaped membership functions, W4-F5-MLP, W2-F1-
412 MLP, W4-F5-ANFIS, and W4-F6-ANFIS have the best performance, respectively. By comparing
413 the results of decomposition techniques, the wavelet decomposition performed better result than
414 others. Then, the fuzzy transform with the Triangular type 1 membership function has the best
415 performance. Similarly, the combination of wavelet decomposition and fuzzy transform
416 separately, the Triangular type 1 and the Bell-shaped membership functions have appropriate
417 performance.

418 **Table 7. Results of models under combined methods**

419 In Table 7, in the single model, the original time series is considered the input of neural networks.
420 In this state, no decomposition technique is used. The results of the single model for each network
421 are presented in Table 7. Then, decomposition techniques are applied. In DWT, the first original
422 data is decomposed into four levels. These four levels are integrated and considered the input of
423 neural networks. The results of the DWT are presented in Table 7. In the fuzzy transform, the first
424 original data is considered the input of fuzzy transform with five membership functions. The
425 original data is decomposed into six levels by the fuzzy transforms with various membership
426 functions. Thirty levels are obtained from fuzzy transforms with five membership functions, and
427 they are integrated and considered the input of neural networks. The results of the FT for each
428 network are presented in this table. In DWT-FT, first, the original data is decomposed into four
429 levels by wavelet decomposition. These decomposed levels are considered the input of fuzzy
430 transforms with five membership functions. The results of DWT-FT for each network are
431 presented in Table 7.

432 **Table 8. Ensemble method results**

433 According to Table 8, in the single model, original data is considered input of neural networks.
434 Then, the output of neural networks is calculated, and these output vectors are integrated.
435 According to the obtained output and target, the root-mean-square error and the correlation
436 coefficient are calculated. In the combined model, different levels are integrated and considered
437 input of neural networks. Then, the outputs of neural networks are calculated and integrated.
438 According to the obtained target and output, the root-mean-square error and the correlation
439 coefficient are calculated (see Table 8). The results and accuracy of forecasting are improved when
440 the decomposition techniques are used. In the single model, five levels are integrated, but in the

441   combination model, 154 levels are merged. The scale of data in the combined model is bigger than
442   the single model. Therefore, to compare the results, the correlation coefficient should be
443   considered.

6. Limitations

445   Despite the good results that the proposed ensemble approach theoretically provides, in practice,
446   price time-series forecasting is faced with challenges that limit their use and, in some cases, make
447   it impossible. Algorithms use methods such as violet, which not only removes noise in the data
448   but also changes the dimensionality of the data. Prices in the real world are accompanied by noise,
449   so removing noise can change the data's nature. Although the use of fuzzy transformations reduces
450   the accuracy of the models, it preserves the data's nature. Also, time-series data are strongly
451   correlated with delay 1 (see related figures in Section 5). It causes the forecasting to be done with
452   one delay phase. To remove this problem, the smoothed price, such as moving average or
453   exponential moving average may be used instead of the original price.

7. Summary and conclusion

455   In this research, a new approach is presented to identify optimal time delays. Also, for removing
456   the noise from data, the wavelet decomposition has been used. On the other hand, for modeling
457   uncertainty of gas price time-series that influenced by political, economic, social, etc., factors, the
458   fuzzy transform with different membership functions is used. Afterward, the wavelet
459   decomposition and fuzzy transformation are combined for improving the accuracy of forecasting
460   and modeling uncertainty. Finally, the best-selected models are integrated as an ensemble method
461   for obtaining the best accuracy and performance. In this paper, for the time-series of natural gas
462   prices, optimal time delays (1, 2, and 5) were identified by a new proposed approach. Then,
463   wavelet decomposition and fuzzy transform with different membership functions are used in the
464   preprocessing stage. In this study, four models were presented to forecast the natural gas price. In
465   the first model with the input of the original time series, RMSE and R of RBF were 0.1483 and
466   0.95731 respectively as the best performance. In the second model or wavelet neural networks, the
467   values of RMSE and R of W4-MLP have been obtained 0.00061674 and 1, respectively, as the
468   best performance. In the third model or fuzzy neural networks, the values of RMSE and R of F6-
469   MLP with Triangular type 1 membership function were 0.00010555 and 0.99949, respectively, as
470   the best performance. In the fourth model or wavelet-fuzzy neural networks, the values of RMSE
471   and R of W1-F5-RBF combination have been obtained 0.00011825 and 0.99875, respectively, as
472   the best performance. Finally, by integrating the forecasting results for the original time series of
473   natural gas price, the values of RMSE and R were 0.75907 and 0.95513, respectively. The
474   combination of the forecasting results for time series decomposed by wavelet and fuzzy transforms
475   has been performed, and the values of RMSE and R have been calculated 0.966634 and 0.99803,
476   respectively. All four models, in terms of RMSE and R, are favorable but comparing the four

477 models, the second model considering the correlation coefficient, performed better than the other
478 three models.

Reference

480 1. Adhikari, R., & Agrawal, R. K. (2014). A combination of artificial neural network and random walk
481 models for financial time series forecasting. *Neural Computing and Applications*, *24*(6), 1441-1449.
482 2. Marcek, D. (2018). Forecasting of financial data: a novel fuzzy logic neural network based on error-
483 correction concept and statistics. *Complex & Intelligent Systems*, *4*(2), 95-104.
484 3. Mo, H., Wang, J., & Niu, H. (2016). Exponent back propagation neural network forecasting for financial
485 cross-correlation relationship. *Expert Systems with applications*, *53*, 106-116.
486 4. Heng-Li, Y. A. N. G., & Han-Chou, L. I. N. (2015). INTEGRATING EMD, CHAOS-BASED NEURAL
487 NETWORK AND PSO FOR FINANCIAL TIME SERIES FORECASTING. *Economic Computation &
488 Economic Cybernetics Studies & Research*, *49*(1).
489 5. Hota, H. S., Handa, R., & Shrivas, A. K. (2018). Neural Network Techniques to Develop a
490 Robust Financial Time Series Forecasting System. *International Journal of Pure and Applied
491 Mathematics*, *118*(19), 125-133.
492 6. Wang, J., & Wang, J. (2017). Forecasting stochastic neural network based on financial empirical mode
493 decomposition. *Neural Networks*, *90*, 8-20.
494 7. Kusy, M. (2018). Fuzzy c-means-based architecture reduction of a probabilistic neural network. *Neural
495 Networks*, *108*, 20-32.
496 8. Hidayat, Y., Sutijo, B., Bon, A. T., & Supian, S. (2016). Indonesian financial data modeling and
497 forecasting by using econometrics time series and neural network. *Global Journal of Pure and Applied
498 Mathematics*, *12*(4), 3745-3757.
499 9. Bai, Y., Li, Y., Wang, X., Xie, J., & Li, C. (2016). Air pollutants concentrations forecasting using back
500 propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric
501 pollution research*, *7*(3), 557-566.
502 10. Yu, C., Li, Y., Xiang, H., & Zhang, M. (2018). Data mining-assisted short-term wind speed forecasting
503 by wavelet packet decomposition and Elman neural network. *Journal of Wind Engineering and Industrial
504 Aerodynamics*, *175*, 136-143.
505 11. Ji, Y., Tang, D., Blythe, P., Guo, W., & Wang, W. (2014). Short-term forecasting of available parking
506 space using wavelet neural network model. *IET Intelligent Transport Systems*, *9*(2), 202-209.
507 12. Guimarães Santos, C. A., & Silva, G. B. L. D. (2014). Daily streamflow forecasting using a wavelet
508 transform and artificial neural network hybrid models. *Hydrological Sciences Journal*, 59(2), 312-324.
509 13. Dadkhah, M., Rezaee, M.J. and Chavoshi, A.Z. (2018). Short-term power output forecasting of hourly
510 operation in power plant based on climate factors and effects of wind direction and wind speed. *Energy,*
511 148, 775-788.
512 14. Rezaee, M.J., Dadkhah, M. and Falahinia, M. (2019). Integrating neuro-fuzzy system and evolutionary
513 optimization algorithms for short-term power generation forecasting. *International Journal of Energy
514 Sector Management*, 13(4), 828-845.
515 15. Heydari, G., Vali, M., & Gharaveisi, A. A. (2016). Chaotic time series prediction via artificial neural
516 square fuzzy inference system. *Expert Systems with Applications*, *55*, 461-468.
517 16. Yolcu, O. C., & Lam, H. K. (2017). A combined robust fuzzy time series method for prediction of time
518 series. *Neurocomputing*, *247*, 87-101.
519 17. Dincer, N. G., & Akkuş, Ö. (2018). A new fuzzy time series model based on robust clustering for
520 forecasting of air pollution. *Ecological Informatics*, *43*, 157-164.

521 18. Wang, J., Li, H., & Lu, H. (2018). Application of a novel early warning system based on fuzzy time
522 series in urban air quality forecasting in China. *Applied Soft Computing*, *71*, 783-799.
523 19. Wang, J., Li, X., Hong, T., & Wang, S. (2018). A semi-heterogeneous approach to combining crude oil
524 price forecasts. *Information Sciences*, *460*, 279-292.
525 20. Agrawal, R. K., Muchahary, F., & Tripathi, M. M. (2019). Ensemble of relevance vector machines and
526 boosted trees for electricity price forecasting. *Applied Energy*, *250*, 540-548.
527 21. Rezaee, M. J., Jozmaleki, M., & Valipour, M. (2018). Integrating dynamic fuzzy C-means, data
528 envelopment analysis and artificial neural network to online prediction performance of companies in stock
529 exchange. *Physica A: Statistical Mechanics and its Applications*, *489*, 78-93.
530 22. Adhikari, R. (2015). A neural network based linear ensemble framework for time series forecasting.
531 *Neurocomputing*, *157*, 231-242.
532 23. Rezaee, M.J. & Dadkhah, M. (2019). A hybrid approach based on inverse neural network to determine
533 optimal level of energy consumption in electrical power generation. *Computers & Industrial Engineering,*
534 *134,* 52-63.
535 24. Han, J., Kamber, M. and Pei, J. (2012). Data mining concepts and techniques (3rd edition),
536 Chapter 3: Data Preprocessing. *The Morgan Kaufmann Series in Data Management Systems*, 83-
537 124.
538 25. Chu, C.T., Kim, S., Lin, Y.A., Yu, Y., Bradski, G., Olukotun, K. & Ng, A. (2006). Map-reduce for
539 machine learning on multicore. *Advances in neural information processing* systems, 19*, 281-288.
540
541
542
543
544
545
546
547
548
549

Figure 1

Levels of fuzzy transform with Triangular type 1 membership function

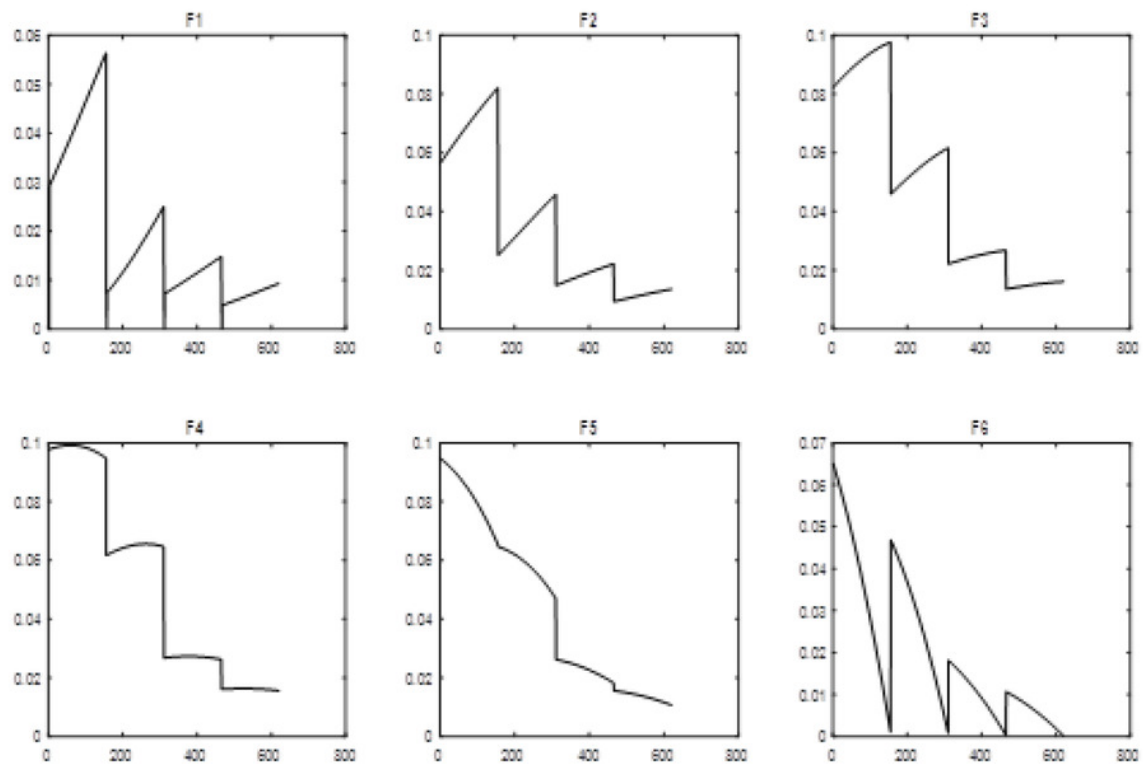



Figure 2

Flowchart of preprocessing and forecasting steps

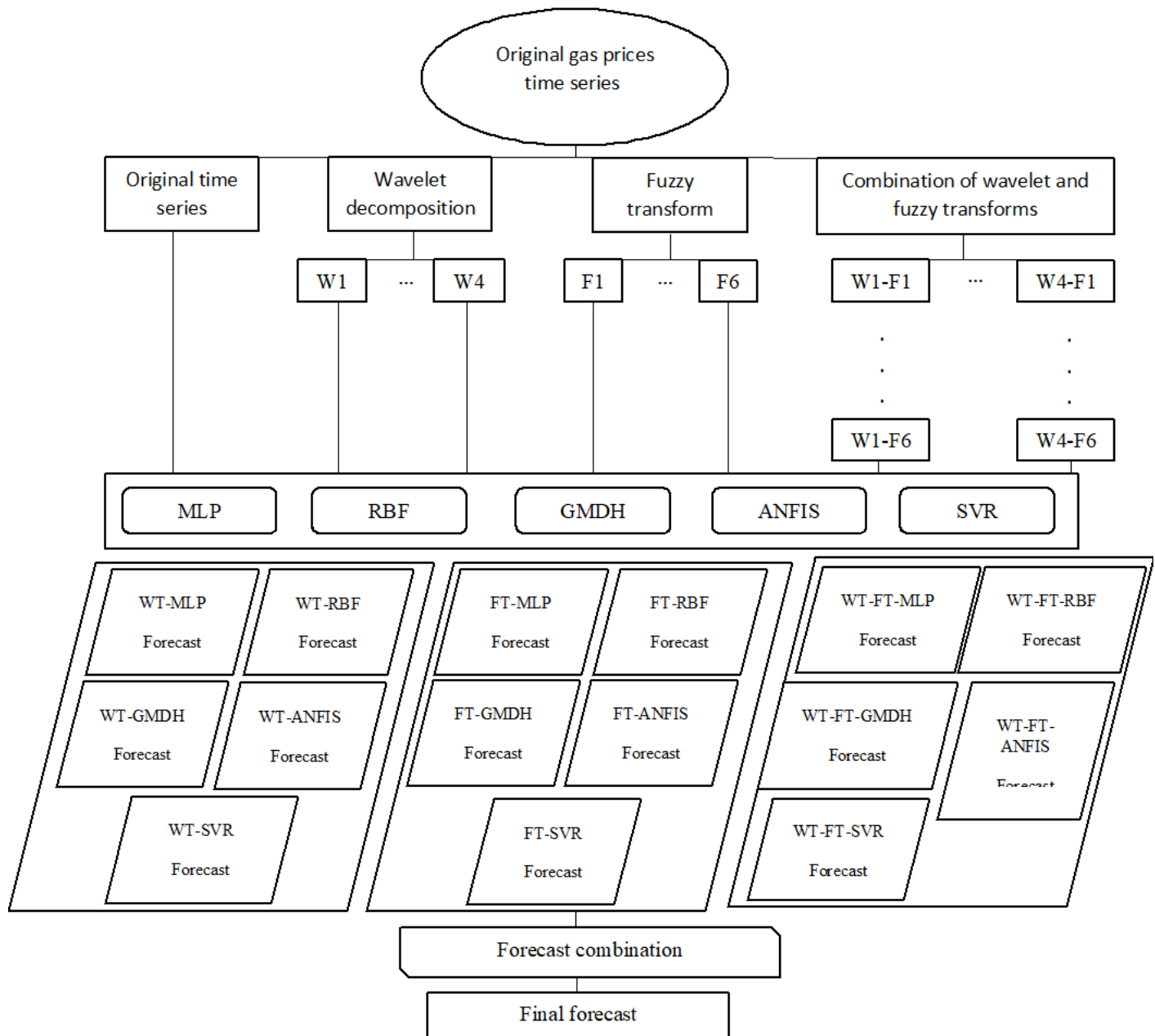



Figure 3

The results of the RBF neural network for the original time series of gas price

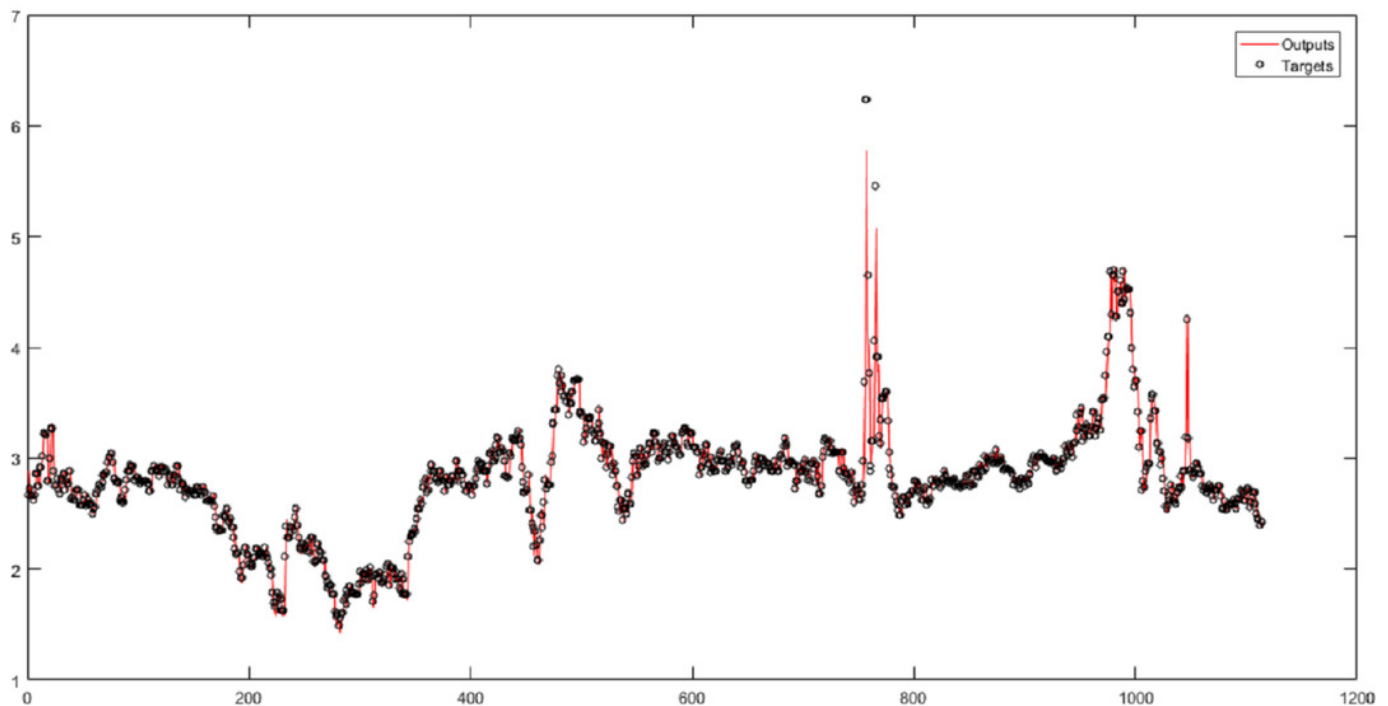



Figure 4

The results of the MLP neural network for Level 4 obtained from the wavelet decomposition

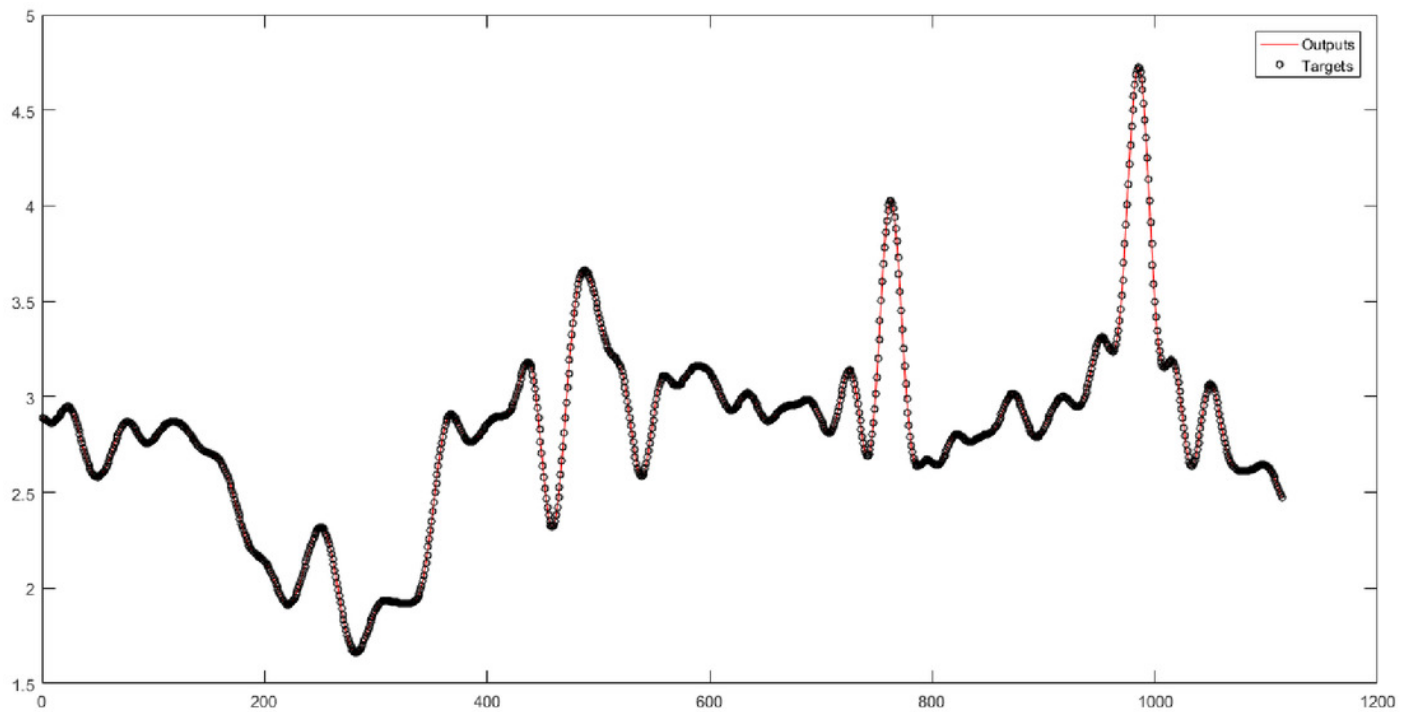



Figure 5

The results of MLP for Level 6 obtained from fuzzy transform with Triangular type 1 membership function

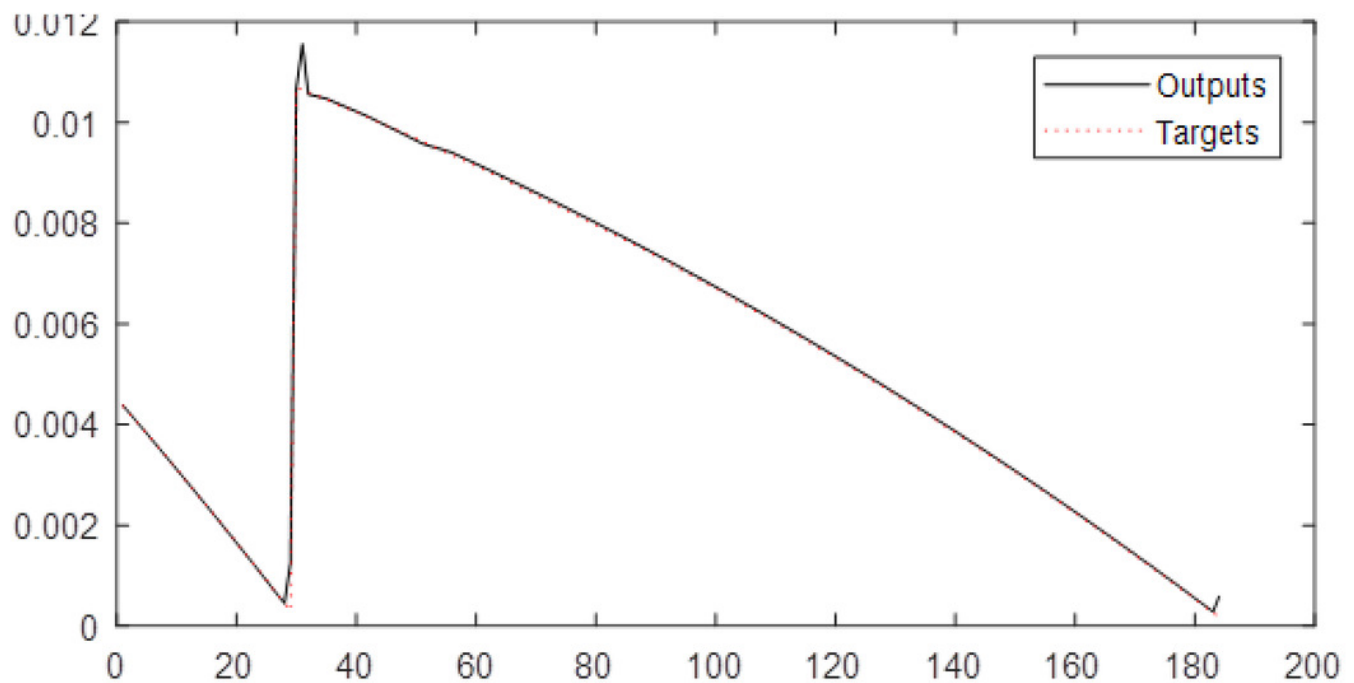



Figure 6

The results of SVR for Level 1 obtained from fuzzy transform with Gaussian membership function

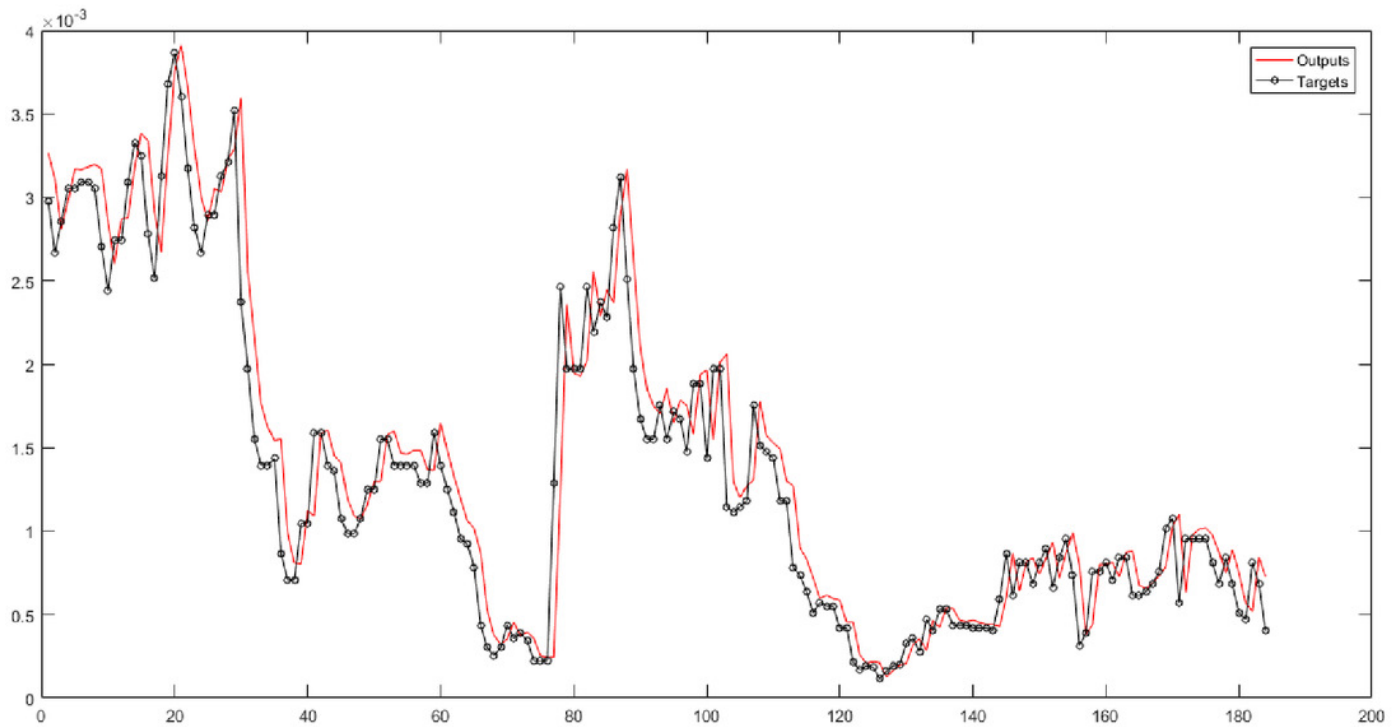



Figure 7

The results of ANFIS for Level 6 obtained from fuzzy transform with Trapezoidal
membership function

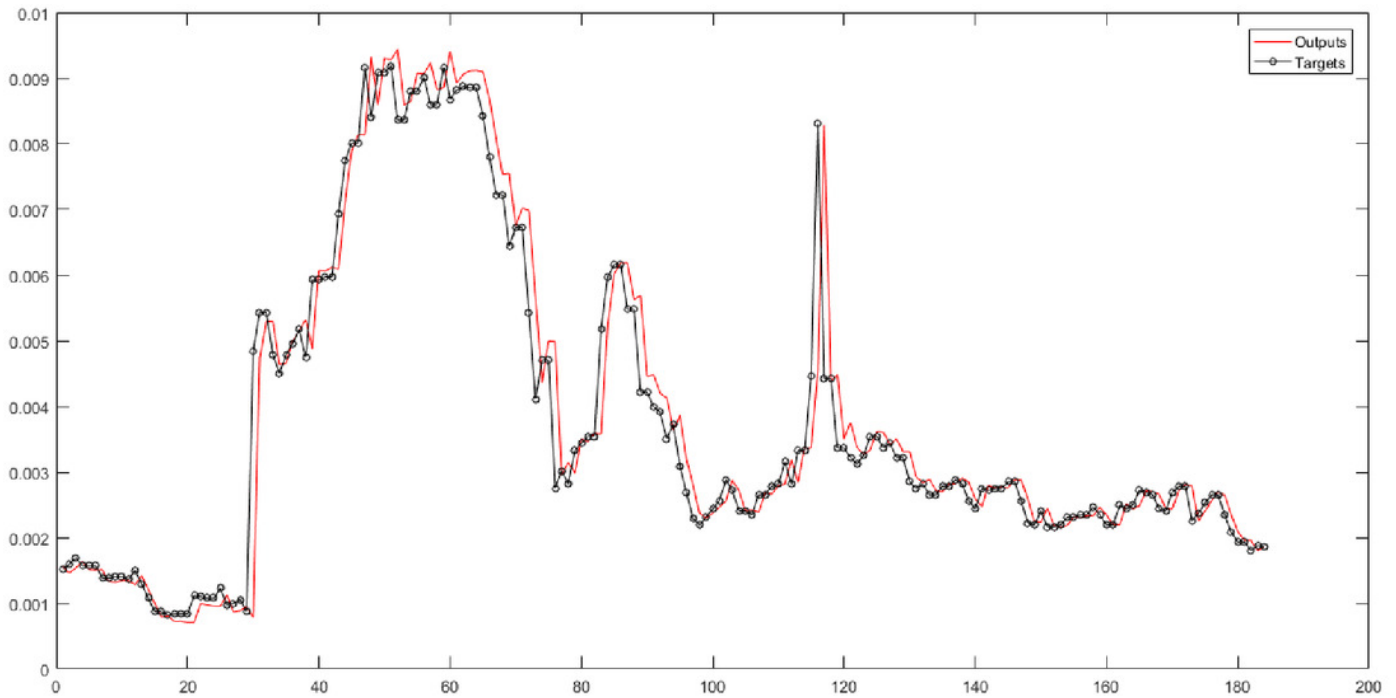



Figure 8

The results of SVR for Level 3 obtained from fuzzy transform with Triangular type 2 membership function

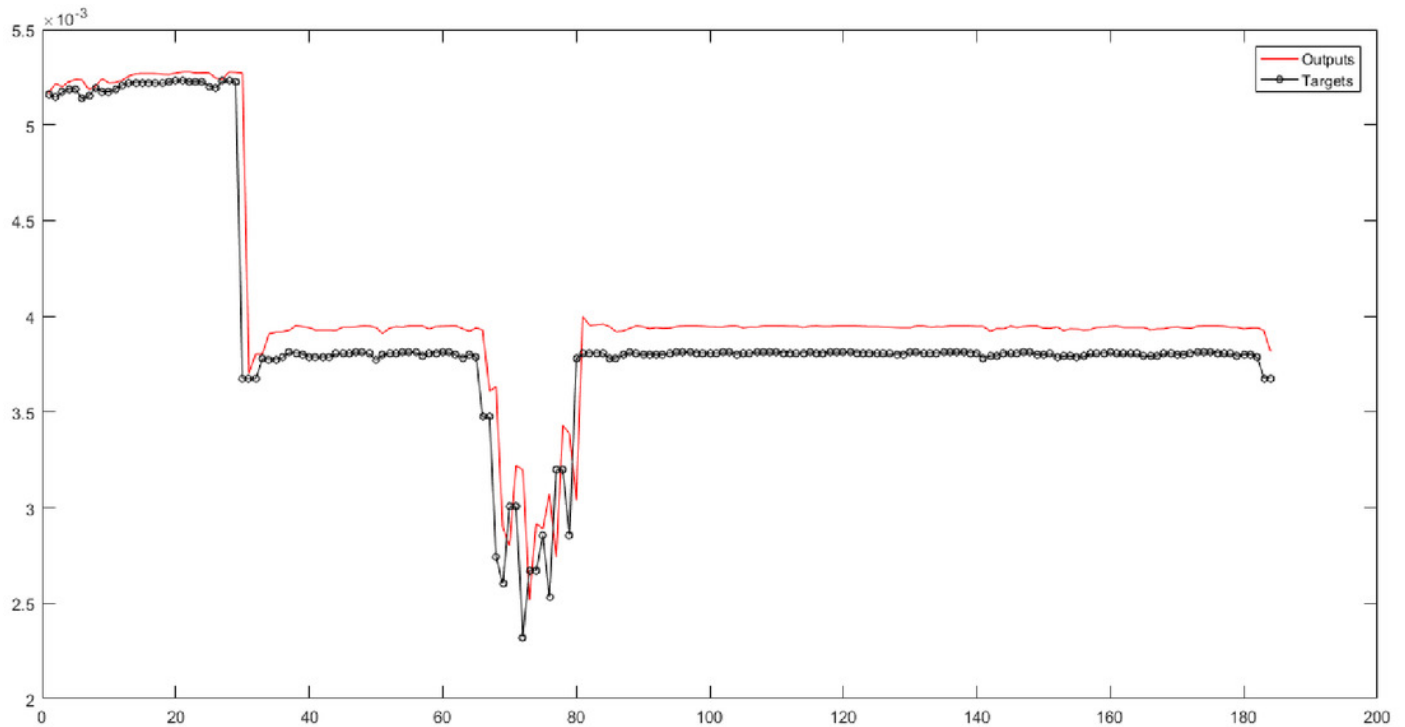
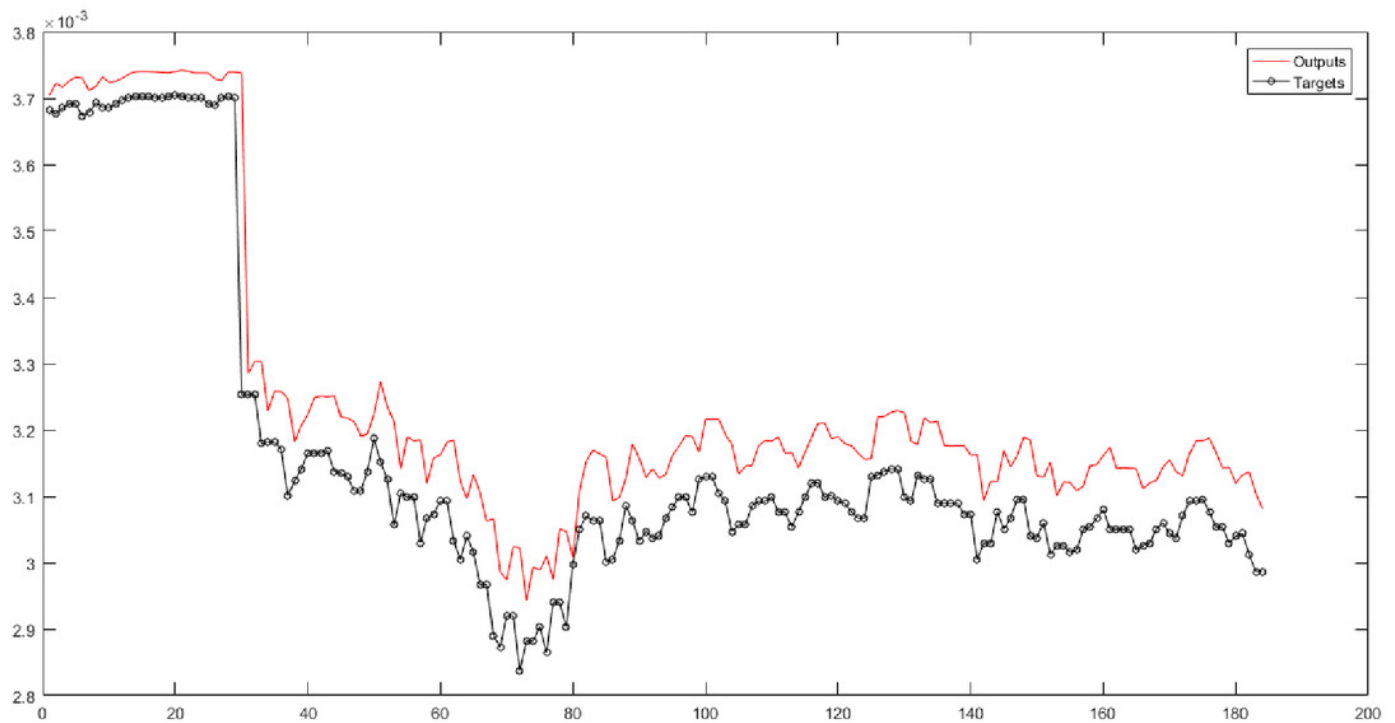



Figure 9

The results of SVR for Level 3 obtained from fuzzy transform with Bell-shaped membership function

The membership functions used for considering uncertainty in time series

3   Table 1. The membership functions used for considering uncertainty in time series

Triangular membership function	$\mu(x) = \max\left(\min\left(\dfrac{x-a}{b-a}, \dfrac{c-x}{c-b}\right), 0\right)$	(6)
Trapezoidal membership function	$\mu(x) = \max\left(\min\left(\dfrac{x-a}{b-a}, 1, \dfrac{d-x}{d-c}\right), 0\right)$	(7)
Gaussian membership function	$\mu(x) = \exp\left[-\dfrac{1}{2}\left(\dfrac{x-m}{\sigma}\right)^2\right]$	(8)
Bell-shaped membership function	$\mu(x) = \dfrac{1}{1 + \left\|\dfrac{x-m}{\sigma}\right\|^{2a}}$	(9)

Complexity analysis of algorithms used in this paper

3   **Table 2: Complexity analysis of algorithms used in this paper**

Machine learning algorithm	Complexity
MLP (three layers with backpropagation algorithm for each iteration)	$O(mpn^3)$
RBF	$O(mpn^2)$
GMDH (for maximum three layers)	$O(mpn^2+pn^3) \sim O(n^3)$
ANFIS (using fuzzy c-means algorithm and for each iteration)	$O(mpl)$
SVR	$O(m^2p)$

Calculating error and correlation coefficient for time series applied wavelet decomposition

2   **Table 3. Calculating error and correlation coefficient for time series applied wavelet**
3   **decomposition**

	RMSE(train)	R(train)	RMSE(test)	R(test)
W1-MLP	0.15357	0.99792	0.082804	0.98659
W1-RBF	0.15355	0.99791	0.095043	0.98334
W1-GMDH	0.16353	0.99763	0.10148	0.98105
W1-ANFIS	0.15766	0.9978	0.091917	0.98392
W1-SVR	0.16212	0.99767	0.095822	0.98214
W2-MLP	0.036885	0.99988	0.021522	0.99906
W2-RBF	0.037009	0.99988	0.023842	0.99888
W2-GMDH	0.055633	0.99973	0.038534	0.99719
W2-ANFIS	0.038451	0.99987	0.024723	0.99879
W2-SVR	0.056466	0.99972	0.033676	0.99773
W3-MLP	0.0062521	1	0.0036305	0.99997
W3-RBF	0.0063061	1	0.0036619	0.99997
W3-GMDH	0.017222	0.99997	0.010397	0.99981
W3-ANFIS	0.0065844	1	0.003712	0.99997
W3-SVR	0.03922	0.99986	0.022998	0.99893
W4-MLP	**0.0014284**	**1**	**0.00061674**	**1**
W4-RBF	0.00125	1	0.00064671	1
W4-GMDH	0.0073589	1	0.0040815	0.99997
W4-ANFIS	0.0012957	1	0.00063232	1
W4-SVR	0.022275	0.99996	0.014339	0.99965

Calculation of noise between original data and decomposed data by wavelet transform

1   **Table 4. Calculation of noise between original data and decomposed data by wavelet**
2   **transform**

	RMSE	R
W1-MLP	0.098071	0.98102
W2-MLP	0.14015	0.9609
W3-MLP	0.14577	0.95729
W4-MLP	0.17024	0.94122
W1-RBF	0.11242	0.97656
W2-RBF	0.13914	0.96151
W3-RBF	0.14471	0.95778
W4-RBF	0.16683	0.94341
W1-GMDH	0.10229	0.98104
W2-GMDH	0.13709	0.96286
W3-GMDH	0.14509	0.95724
W4-GMDH	0.16661	0.94297
W1-ANFIS	0.095041	0.98267
W2-ANFIS	0.15258	0.9542
W3-ANFIS	0.15735	0.95065
W4-ANFIS	0.17997	0.93473
W1-SVR	0.16793	0.94787
W2-SVR	0.17651	0.94237
W3-SVR	0.17981	0.93844
W4-SVR	0.18461	0.93287

Calculating error and correlation coefficient of time series by fuzzy neural networks with various membership functions

1   **Table 5. Calculating error and correlation coefficient of time series by fuzzy neural networks with various membership**
2   **functions**

	Triangular Type 1		Gaussian		Trapezoidal		Triangular Type 2		Bell-shaped	
	RMSE(test)	R(test)	RMSE(test)	R(test)	RMSE(test)	R(test)	RMSE(test)	R(test)	RMSE(test)	R(test)
F1-MLP	0.00077095	0.96429	0.00029688	0.95079	0.00018713	0.94312	0.00023004	0.95002	0.000074295	0.97667
F1-RBF	0.00089835	0.95538	0.00047533	0.88392	0.00049873	0.95647	0.00040971	0.91047	0.00010867	0.93474
F1-GMDH	0.00089677	0.95575	0.00030744	0.88272	0.00034627	0.94506	0.00017237	0.95894	0.000049928	0.97914
F1-ANFIS	0.0008857	0.95544	0.00043059	0.87954	0.00097418	0.82092	0.0003805	0.87243	0.000072496	0.96111
F1-SVR	0.011021	0.95239	0.00025853	0.80092	0.00025624	0.94462	0.0002436	0.86729	0.000047595	0.96606
F2-MLP	0.00096726	0.96964	0.00042193	0.94287	0.00063972	0.96233	0.00027185	0.93324	0.00004843	0.9689
F2-RBF	0.0009696	0.96997	0.00029145	0.95781	0.00021151	0.94216	0.00025795	0.94839	0.000075494	0.97789
F2-GMDH	0.0011351	0.95551	0.00045547	0.88869	0.00061929	0.94694	0.00041263	0.91145	0.0001351	0.94137
F2-ANFIS	0.00096457	0.97	0.00036663	0.88684	0.00034359	0.9455	0.00021831	0.95862	0.000096299	0.98399
F2-SVR	0.020301	0.94807	0.00047876	0.85566	0.0010501	0.79268	0.00042402	0.85079	0.00007986	0.95711
F3-MLP	0.00094221	0.975	0.00028629	0.79847	0.00027515	0.93855	0.0002896	0.85124	0.000045736	0.97052
F3-RBF	0.00097379	0.97421	0.00047731	0.92484	0.00066549	0.96232	0.00027782	0.93014	0.000052228	0.9718
F3-GMDH	0.0026492	0.84415	0.00029649	0.95791	0.00021458	0.94488	0.00026718	0.95029	0.000079557	0.97704
F3-ANFIS	0.00098146	0.97421	0.00046039	0.88481	0.0005375	0.95597	0.00041016	0.9092	0.00011344	0.94166
F3-SVR	0.019744	0.9553	0.00036274	0.88229	0.00034244	0.94562	0.00020767	0.95657	0.000095443	0.98397
F4-MLP	0.00076628	0.98065	0.00047923	0.85738	0.0011875	0.77548	0.00042635	0.8444	0.000082222	0.95733
F4-RBF	0.00076543	0.98105	0.00032689	0.79973	0.00027352	0.94065	0.00028689	0.85996	0.000045848	0.97056
F4-GMDH	0.0008288	0.97694	0.00047408	0.93003	0.00067282	0.96138	0.00027152	0.93391	0.000062642	0.9699
F4-ANFIS	0.00076557	0.98104	0.00029412	0.9586	0.00020249	0.94216	0.00026196	0.9497	0.000075291	0.97789
F4-SVR	0.019265	0.96732	0.00046006	0.88919	0.00053991	0.95016	0.00041158	0.91247	0.0001027	0.94256
F5-MLP	0.00031118	0.99495	0.00034228	0.88742	0.00034541	0.94583	0.00021571	0.95863	0.000048891	0.98421
F5-RBF	0.00020388	0.99703	0.00047886	0.85565	0.0010366	0.79575	0.00042866	0.84917	0.000076752	0.95751
F5-GMDH	0.0013618	0.86725	0.00026898	0.79867	0.000272	0.93883	0.00026593	0.85346	0.000044912	0.97053
F5-ANFIS	0.00020426	0.99703	0.00047765	0.92553	**0.00063896**	**0.9625**	0.00027832	0.93011	0.000045137	0.97178
F5-SVR	0.02087	0.99477	**0.0002834**	**0.95448**	0.00019017	0.94023	0.00023964	0.94398	0.000081016	0.97305
F6-MLP	**0.00010555**	**0.99949**	0.00046531	0.88291	0.00051476	0.95395	0.0004165	0.90795	0.00010673	0.93262
F6-RBF	0.00097823	0.96968	0.0003128	0.88239	0.00034145	0.94489	**0.00016989**	**0.95957**	**0.000046212**	**0.98144**
F6-GMDH	0.0012245	0.96956	0.00050671	0.83456	0.0010225	0.80092	0.00042943	0.84146	0.000089798	0.94124
F6-ANFIS	0.00078845	0.96959	0.0002738	0.77788	0.00027034	0.93807	0.00026978	0.84247	0.000061622	0.96019
F6-SVR	0.019798	0.94619	0.00046632	0.92919	0.00063962	0.96284	0.00027712	0.93069	0.000049766	0.96511

Calculating error and correlation coefficient of time series by wavelet-fuzzy transformation with five membership functions

1   **Table 6. Calculating error and correlation coefficient of time series by wavelet-fuzzy**
2   **transformation with five membership functions**

	The best performance among 120 forecasted combinations	RMSE(train)	R(train)	RMSE(test)	R(test)
WT-FT(Triangular type 1 mf)	**W1-F5-RBF**	0.0011086	0.99898	0.00011825	0.99875
WT-FT(Gaussian mf)	**W4-F5-MLP**	0.00015879	0.99717	0.000020236	0.99811
WT-FT(Trapezoidal mf)	**W2-F1-MLP**	0.00041631	0.99816	0.000072099	0.9918
WT-FT(Triangular type 2 mf)	**W4-F5-ANFIS**	0.00016253	0.99769	0.000030826	0.99746
WT-FT(Bell-shaped mf)	**W4-F6-ANFIS**	0.00007106	0.99901	0.000010122	0.9987

Results of models under combined methods

2 **Table 7. Results of models under combined methods**

		RMSE(train)	R(train)	RMSE(test)	R(test)
MLP	**Single model**	0.25641	0.99418	0.15086	0.95575
	DWT	0.20008	0.99978	0.11801	0.99818
	FT	0.0092812	0.99712	0.0027558	0.98336
	DWT-FT	0.084845	0.98665	0.0090304	0.983
RBF	**Single model**	0.2601	0.99401	0.1483	0.95731
	DWT	0.19436	0.99979	0.12927	0.99779
	FT	0.0093718	0.99706	0.0029492	0.98242
	DWT-FT	0.039197	0.9971	0.010384	0.97874
GMDH	**Single model**	0.26551	0.99376	0.15544	0.95292
	DWT	0.20662	0.99976	0.12897	0.99787
	FT	0.0096842	0.99686	0.0057582	0.93457
	DWT-FT	0.039312	0.99708	0.01169	0.97312
ANFIS	**Single model**	0.26696	0.99369	0.15653	0.95291
	DWT	0.20114	0.99977	0.12044	0.99811
	FT	0.0093793	0.99705	0.0028397	0.98247
	DWT-FT	0.03928	0.99709	0.0090655	0.98329
SVR	**Single model**	0.26844	0.99362	0.15785	0.95137
	DWT	0.23098	0.9997	0.13368	0.99767
	FT	0.020093	0.99484	0.012544	0.96885
	DWT-FT	0.1033	0.99532	0.083158	0.97231

Ensemble method results

Table 8. Ensemble method results

	RMSE	R
Single Model	0.75907	0.95513
Ensemble Model	0.966634	0.99803