Applied Soft Computing Journal 85 (2019) 105827

Contents lists available at ScienceDirect

Applied Soft Computing Journal

journal homepage: www.elsevier.com/locate/asoc

A clustering-based ensemble approach with improved pigeon-inspired optimization and extreme learning machine for air quality prediction

Feng Jiang [a,*], Jiaqi He [a], Tianhai Tian [b]

[a] School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, 430073, China
[b] School of Mathematical Science, Monash University, Melbourne Clayton 3800, Australia

ARTICLE INFO

Article history:
Received 10 April 2019
Received in revised form 11 August 2019
Accepted 30 September 2019
Available online 9 October 2019

Keywords:
Wavelet packet decomposition
Pigeon-inspired optimization
Extreme learning machine
Multidimensional scaling
K-means clustering

ABSTRACT

In this paper, a novel hybrid learning method is carried out to forecast urban air quality index (AQI). Wavelet packet decomposition (WPD) is firstly performed to decompose the original AQI data into lower-frequency subseries. Then, we improve the pigeon-inspired optimization through using the particle swarm optimization algorithm. The improved pigeon-inspired optimization (IPIO) approach is applied to optimize the initial weights and thresholds of extreme learning machine (ELM) and then the modified ELM (MELM) is employed to forecast the subseries respectively. Moreover, multidimensional scaling and K-means (MSK) clustering methods are utilized to cluster the forecasting outcomes into high frequency, medium–high frequency, medium–low frequency and low frequency subseries. Finally, MELM, as an ensemble approach, is applied to ensemble the subseries together and achieve the final results. To test the predictive precision of the proposed hybrid WPD-MELM-MSK-MELM learning method, AQI of Harbin in China is adopted to make short-term, middle-term and long-term predictions separately. Different decomposition approaches are utilized to compare with WPD, and the non-clustering hybrid model is also compared with the proposed method. The forecasting outcomes indicate that WPD is more suitable for predicting AQI and the proposed WPD-MELM-MSK-MELM learning method has better predictive performance on horizontal precision, directional precision and robustness than some existing methods and benchmark models in this paper.

1. Introduction

Recently, air pollution has become a significant issue in earth, environmental and life sciences. Air quality index (AQI) is an indispensable index to calculate the quality of the air, which has a close connection with human life and health. Accurate prediction of AQI can help the government and environmental protection agency to effectively control and reduce air pollution and promote the sustainable development of the nature and human society.

AQI data are non-stationary, nonlinear and complicated time series. Classical forecasting approaches like Markov chain [1], support vector regression [2], boosting tree [3], artificial neural networks [4–9], convolutional neural network [10] have been widely adopted to forecast the AQI series. Nowadays, intelligent meta heuristic learning algorithms have attracted the eyes of scholars. Wang et al. [11] applied a hybrid approach combining wavelet neural network and genetic algorithm (GA-WNN) to forecast the nonlinear series of fine particulate matter and carbon monoxide. It was shown that the hybrid GA-WNN method has higher prediction accuracy than the back propagation neural network (BPNN). Yang et al. [12] adopted the parallel chaos (PC) searching algorithm to optimize the initial parameters of extreme learning machine (ELM). The simulation outcomes illustrate that the PC-ELM approach has better generalization and optimization performance than the single ELM method. He et al. [13] adopted chaotic particle swarm optimization (CPSO) algorithm to optimize the artificial neural networks for predicting the time series of particulate matter at urban intersection. The hybrid model has better forecasting performance on air pollution concentration. Heidari et al. [14] utilized Lévy flight to modify the grey wolf optimization for enhancing the optimizing efficiency. The pigeon-inspired optimization (PIO) algorithm is a kind of novel metaheuristic learning algorithm inspired by the homing behaviors of pigeons. Through utilizing the map and compass operator and the landmark operator in the algorithm, it can find the global optimal values of the problem [15,16]. Jiang et al. [17] adopted the PIO algorithm to optimize the original parameters of ELM. The hybrid model can effectively enhance the predictive precision of the bulk commodity futures prices. Dou et al. [18] utilized the PIO algorithm to optimize the unmanned air vehicles controller. The PIO algorithm is used in model prediction control, which has

* Corresponding author.
E-mail addresses: fjiang@zuel.edu.cn (F. Jiang), jiaqi.he@zuel.edu.cn (J.Q. He), tianhai.tian@monash.edu (T.H. Tian).

https://doi.org/10.1016/j.asoc.2019.105827
1568-4946/

better performance than the particle swarm optimization (PSO) algorithm.

Due to the wide variety of sources and complicated formative factors, the AQI series is a nonlinear time series with strong vibration [19]. Because of the high requirement of forecasting accuracy, decomposition and ensemble approaches has been utilized to forecast air pollution concentration [20]. Recently, decomposition approaches, such as ensemble empirical mode decomposition (EEMD) [21,22], wavelet packet decomposition (WPD) [23], fast ensemble empirical mode decomposition (FEEMD) [24] and singular spectrum analysis (SSA) [25] have been introduced into the prediction of nonlinear time series. Wang et al. [26] utilized the variational mode decomposition (VMD) method to decompose the commodity futures price and then utilized the PSO algorithm to optimize the initial weights and thresholds of the BPNN. Decomposition approaches can significantly enhance the predictive accuracy of single forecasting models. Wang et al. [27] applied the complementary ensemble empirical mode decomposition (CEEMD) and VMD to decompose the AQI data and the ELM optimized by differential evolution (DE) algorithm is employed to forecast the subseries respectively. The hybrid CEEMD-VMD-DE-ELM-ADD model is used to forecast AQI in Beijing and Shanghai. The hybrid model can effectually enhance the accuracy of DE-ELM. Hence, the hybrid decomposition-optimization-ensemble approaches has been widely utilized in the forecasting of different air pollution concentration.

In this paper, we present a novel hybrid learning method based on WPD, multidimensional scaling and K-means (MSK) clustering approaches and the modified ELM (MELM) for urban air quality index prediction for enhancing the horizontal and directional forecasting accuracy of the nonlinear AQI series. In order to reduce the prediction difficulty and errors, we adopt WPD to decompose the initial AQI data into several subseries. Then, we utilize the PSO algorithm to modify the PIO algorithm. The improved PIO (IPIO) algorithm is used to optimize the original parameters of ELM. MELM is performed to forecast the subseries separately. In order to enhance the predictive precision and increase the interpretability of the subseries, MSK is applied to cluster the subseries into high frequency, medium–high frequency, medium–low frequency and low frequency subseries. The proposed MELM model, as an ensemble approach, is utilized to ensemble the four subseries together and achieves the final results of AQI.

The rest of this paper is organized as below. Related methods of WPD, IPIO, MELM, MSK and the proposed hybrid learning framework are detailed in Section 2. The empirical outcomes, robustness outcomes and comparison results of the proposed approach and benchmark models are presented in Section 3. Section 4 discusses the conclusions and the considerations on future work.

2. Methodologies

In this section, we introduce WPD, the IPIO algorithm, MELM, the MSK clustering approach respectively and then present a novel hybrid WPD-MELM-MSK-MELM learning method for urban AQI prediction.

2.1. Wavelet packet decomposition

WPD is a kind of adaptive and efficient decomposition approach [28]. WPD can be utilized to decompose the original series into several lower-frequency subseries, which is an effective extension of wavelet decomposition (WD). It can make multi-level divisions of the original signal frequency band and improve the resolution ratio of signal in time domain and frequency domain. WD can only divide the original signal into lower-frequency subseries, however, the WPD algorithm can both partition the higher-frequency and lower-frequency subseries of the initial signal [29]. WPD can be adopted to decompose and analyze the complicated and non-stationary time series signals. WPD has been widely used in the fields of industry, environmental analysis, economics and finance [17].

2.2. Improved pigeon-inspired optimization

PIO is a novel intelligent learning method, which has been widely utilized in industrial engineering and financial analysis [30]. The PIO algorithm imitates the homing behavior of pigeons. Through using the map and compass operator and the landmark operator, PIO algorithm can find the global optimal solutions. PSO is also a classical and efficient meta heuristic learning algorithm. Inspired by the hybrid dragonfly algorithm [31], we introduce the PSO algorithm into the PIO algorithm and propose a novel IPIO learning algorithm for enhancing the optimization ability and global searching ability. The process of the IPIO learning algorithm is described as follows. The pseudo code of the IPIO algorithm is displayed in Algorithm 1.

Step 1: Initialize parameters of PSO, utilize PSO for parameter optimization and introduce the adaptive mutation operation into PSO. The particles in PSO are initialized with a certain probability during each interaction. The adaptive mutation operation can enhance the diversity of the population of PSO and expand the searching space of particles. The updating formula of velocity and position in PSO are shown as below.

$$V_i(t+1) = \omega \cdot V_i(t) + l_1 \cdot rand_1 \cdot (P_i(t) - X_i(t)) + l_2 \cdot rand_2 \cdot (P_{gb}(t) - X_i(t)), \tag{1}$$

$$X_i(t+1) = X_i(t) + V_i(t). \tag{2}$$

Here $V_i(t)$ and $X_i(t)$ represent the velocity and position of the ith particle in the tth iteration. ω is the inertia weight, l_1 and l_2 represent acceleration factors in PSO, and the values of l_1 and l_2 are always greater than 0. P_i is the individual best position and P_{gb} represents the global best position. $rand_1$ and $rand_2$ are two random numbers between 0 and 1.

Step 2: Initialize parameters of the PIO algorithm and introduce the global best values of PSO into the PIO algorithm. If the fitness of the global best values of PSO is better than the initial values of PIO, the global best values of PSO will be brought into the PIO algorithm for optimization.

Step 3: PIO is applied for parameter optimization. The inertia weight is added into the PIO algorithm. The inertia weight can make a balance between global search and local search in PIO. The formula of the inertia weight is defined as follows.

$$w(t) = w_s - (w_s - w_e) \times t/T. \tag{3}$$

Here $w(t)$ is the inertia weight in the tth iteration. w_s represents the initial inertia weight and w_e is the inertia weight when the iteration stops. T represents the maximum number of iterations.

The PIO algorithm consists two significant operators-the map and compass operator and the landmark operator. In the map and compass operator, the updating formula of velocity and location are displayed as below.

$$V_j(t+1) = w(t) \cdot V_j(t) \cdot e^{-Rt} + rand_3 \cdot (X_g - X_j(t)), \tag{4}$$

$$X_j(t+1) = X_j(t) + V_j(t+1). \tag{5}$$

Here $V_j(t)$ and $X_i(t)$ represent the velocity and location of the jth pigeon in the tth iteration. $rand_3$ is a random number between 0

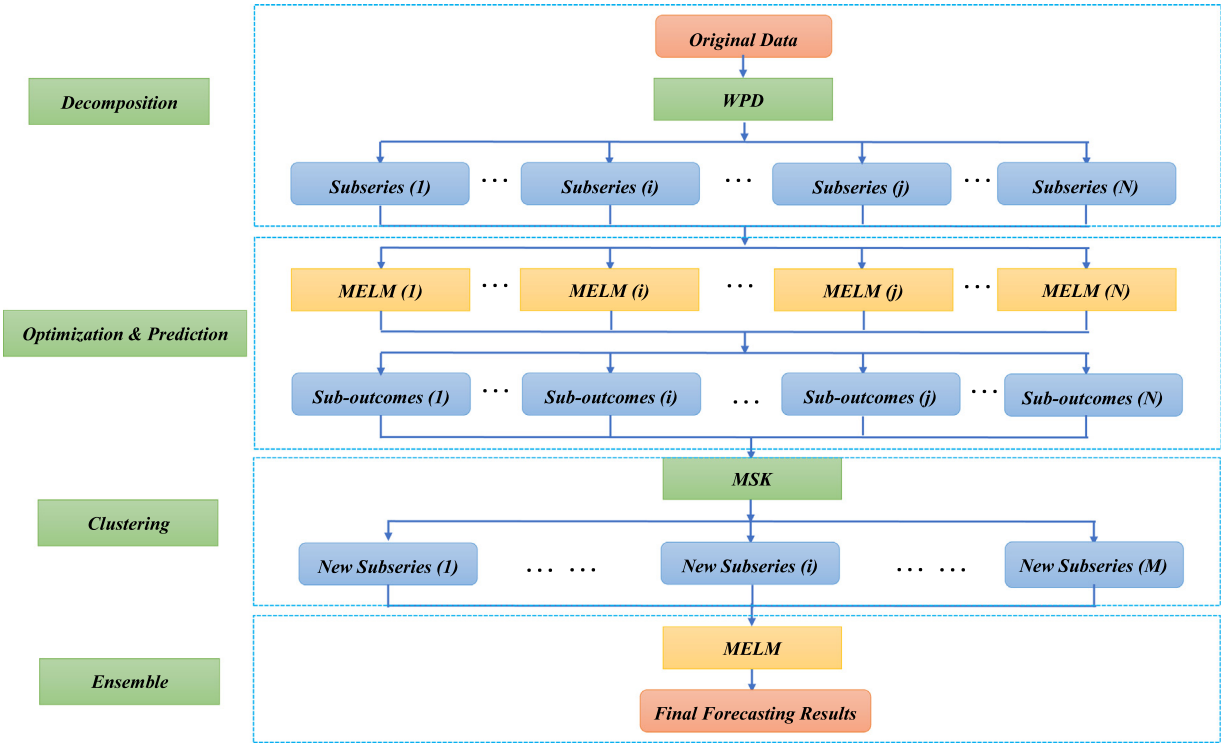

Fig. 1. Learning framework of the hybrid WPD-MELM-MSK-MELM approach.

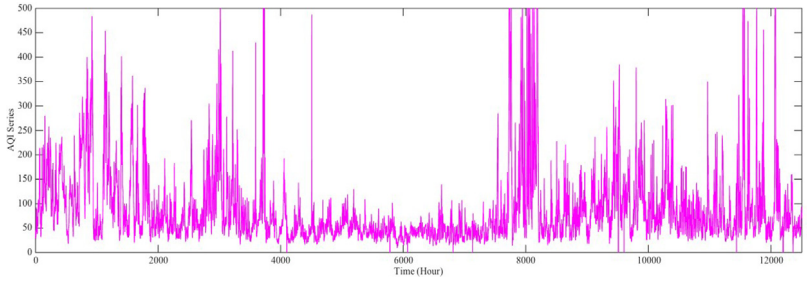

Fig. 2. AQI series of Harbin.

and 1. R is the map and compass factor, whose value is between 0 and 1.

When the map and compass operator reached the maximum number (T_1), the landmark operator starts to find the optimum solution. The updating formula of pigeons' location in the landmark operator is described as follows.

$$N_{landmark}(t+1) = \frac{N_{landmark}(t)}{2}, \tag{6}$$

$$X_{cen}(t) = \frac{\sum X_j(t) \cdot g(X_j(t))}{N \cdot \sum g(X_j(t))}, \tag{7}$$

$$X_j(t+1) = X_j(t) + rand_4 \cdot (X_{cen}(t+1) - X_j(t)). \tag{8}$$

In the process of the landmark operator, the amount of pigeons decreases by half in each iteration. $N_{landmark}(t)$ represents the number of pigeons in the tth iteration. X_{cen} is the center location of the pigeons and $g(\cdot)$ represents the fitness function. $rand_4$ is a random number between 0 and 1.

Step 4: When the PIO algorithm achieves the maximum number of the landmark operator (T_2), the PIO algorithm stops and the location of the global optimal solution (X_{gb}) is the final outcomes of parameter optimization.

2.3. Modified extreme learning machine

ELM is an improved machine learning method of single hidden-layer feed forward neural networks [32–35]. ELM is an efficient learning algorithm, which can overcome the challenges of slow training speed and local optimum issues. Initial weights and thresholds are two indispensable factors in ELM. In the MELM approach, the IPIO algorithm is adopted to optimize the initial weights and thresholds of ELM for improving the prediction performance of ELM. In MELM, the root mean square error (RMSE) of ELM is set as the fitness function of the IPIO algorithm. The equation of RMSE is displayed as below. And the final global best results of the IPIO algorithm are assigned as the initial weights and thresholds of ELM.

$$RMSE = \sqrt{\frac{1}{L}\sum_{j=1}^{L}(\hat{h}(i) - h(i))^2}. \tag{9}$$

Here L is the amount of sample points, $h(i)$ and $\hat{h}(i)$ represent the real values and the predicting values respectively.

Algorithm 1 IPIO optimized by PSO

1. Utilize PSO with the adaptive mutation operation to find the global optimum solution (P_{gb}) by updating the velocities and locations according to Eqs. (1) and (2);
2. Initialize the parameters of PIO with a random velocity (V_j) and position (X_j) and set the forecasting error as the fitness function;
3. Compare the fitness values of P_{gb} and the initial X_j , and update the initial position with a better fitness value;
4. **while** ($t \le T_1$) **do**
5. **for** each individual
6. Update the velocity and location according to Eqs. (4) and (5);
7. **end for**
8. Update the position of the global best solution X_{gb} ;
9. **end while**
10. **while** ($T_1 < t \le T_2$) **do**
11. **for** each individual
12. Update the location based on Eqs. (6) - (8);
13. **end for**
14. Update the position of the global best solution X_{gb} ;
15. **end while**
16. return X_{gb} ;

Table 1
Calculation formula of statistical indicators.

Statistical indicator	Calculation formula
MAE	$MAE = \frac{1}{M}\sum_{i=1}^{M}\left\|\hat{z}(i) - z(i)\right\|$
NRMSE	$NRMSE = \frac{100}{\bar{z}}\sqrt{\frac{1}{M}\sum_{i=1}^{M}(\hat{z}(i) - z(i))^2} \times 100\%$
RMSE	$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(\hat{z}(i) - z(i))^2}$
Ds	$Ds = \frac{1}{M}\sum_{i=2}^{M}\mu(i) \times 100\%$

2.4. Multidimensional scaling and K-means clustering

Multidimensional scaling (MS) is a kind of analysis approach to visualize the distribution of high-dimensional data [36]. By utilizing Euclidean distance and similarities among samples, MS can build a similar low-dimensional space, in which the distance among samples can remain as consistent as possible. If the low-dimensional space is two-dimensional or three-dimensional, the visual results of high-dimensional data can be plotted.

K-means is a classical prototype clustering method [37]. For a given sample set $Y = \{y_1, y_2, \ldots, y_k\}$, the main idea of K-means is to use a greedy algorithm to calculate an approximate minimum value of clustering error sum of squares:

$$Q = \sum_{m=1}^{n}\sum_{y \in E_m} \|y - e_m\|_2^2, \tag{10}$$

where $E = \{E_1, E_2, \ldots, E_n\}$ is a division of clusters and $e_m = \frac{1}{|E_m|}\sum_{y \in E_m} y$ is the mean vector of the division E_m. The number of clusters is a significant factor in K-means.

Due to high dimensions of the subseries, it is a difficult issue to decide the number of clusters. In this paper, MS is used to decide the number of clusters intuitively and K-means is applied to cluster the subseries for effectually reducing the complexity of prediction in this paper.

2.5. Learning framework of the hybrid WPD-MELM-MSK-MELM approach

We present a novel learning approach of the hybrid WPD-MELM-MSK-MELM approach to improve the forecasting performance of AQI series. Firstly, WPD is applied to decompose the original AQI series into several lower-frequency subseries, which can effectively reduce the forecasting difficulty and errors of the original AQI series. Secondly, the IPIO algorithm is employed to optimize the initial weights and thresholds of ELM. Thirdly, the subseries of AQI are predicted by MELM separately. The hybrid MELM model can have better prediction performance than the traditional ELM approach. Then, the MSK clustering are adopted

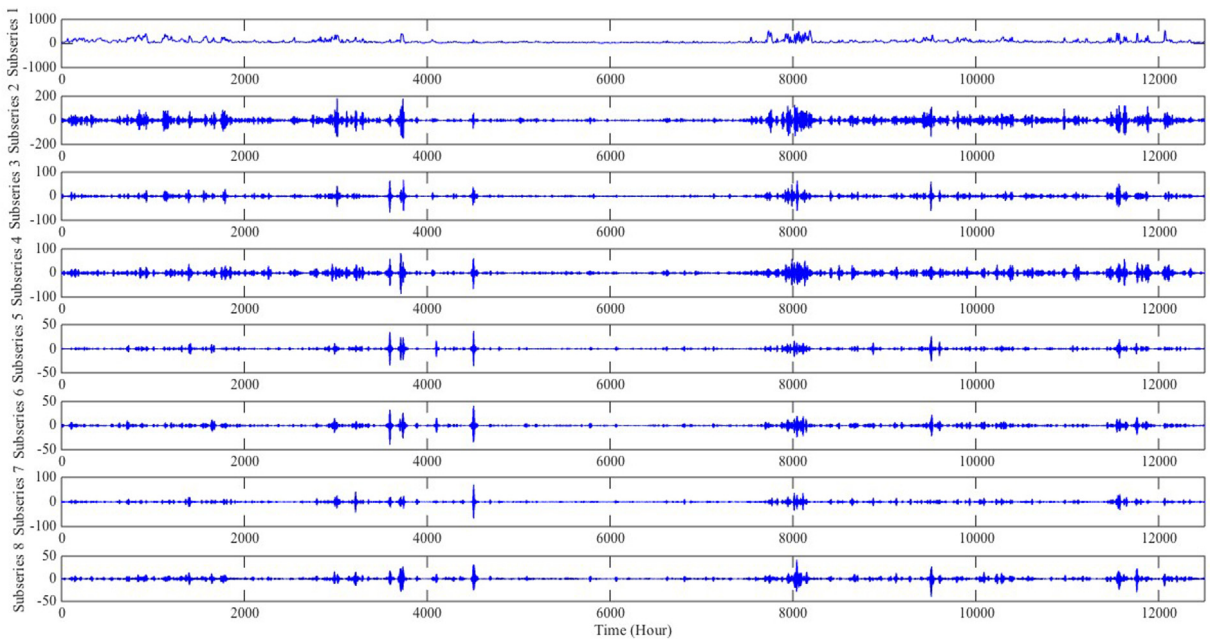

Fig. 3. Decomposition of original AQI data.

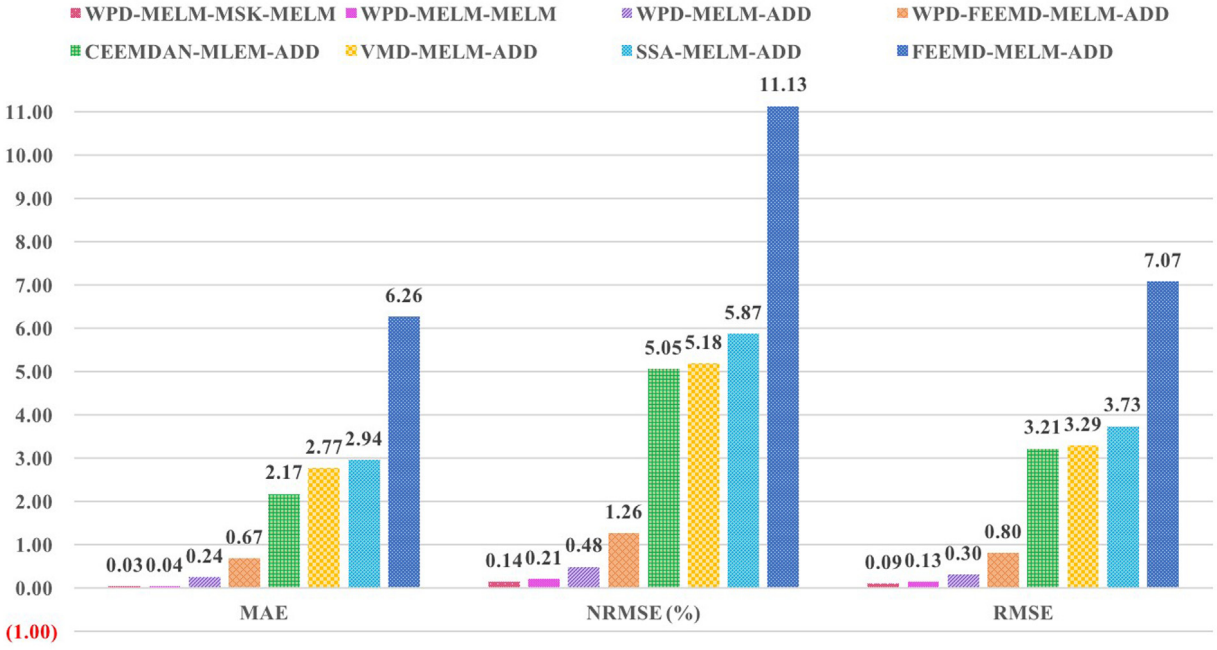

Fig. 4. Horizontal errors of hybrid decomposition models in short-term prediction.

to cluster the forecasting results into high frequency, medium–high frequency, medium–low frequency and low frequency subseries. The clustering approach can increase the interpretability of the subseries and enhance the forecasting accuracy. Lastly, MELM, as the ensemble approach, is utilized to calculate the final forecasting outcomes of the initial AQI data. Fig. 1 displays the framework of the proposed hybrid WPD-MELM-MSK-MELM learning method.

3. Empirical analysis

In this section, data description and predictive outcome analysis are discussed respectively to verify the forecasting accuracy of the proposed learning method.

3.1. Data description

Hourly AQI data of Harbin in China are applied to test the accuracy and applicability of the proposed approach. Harbin is a city in the northeast of China. The developments of heavy industry and central heating in winter make the air quality in Harbin become worse and worse. Hourly AQI data of Harbin with 12504 samples are gained from the website of Ministry of ecological environment of People's Republic of China. AQI data of Harbin are from 0:00 in 1th December 2016 to 23:00 in 5th May 2018. Original AQI data of Harbin are displayed in Fig. 2.

AQI data in Harbin are irregular, non-stationary and nonlinear. For verifying the proposed hybrid approach in this paper, we predict AQI series of Harbin of the next one day, next one week

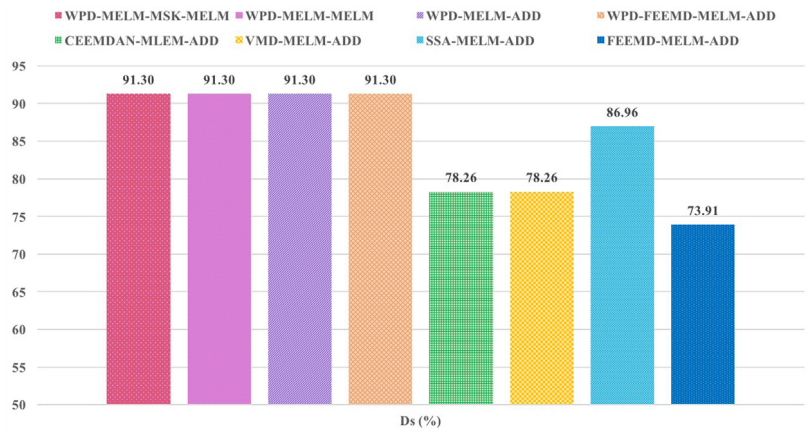

Fig. 5. Directional errors of hybrid decomposition models in short-term prediction.

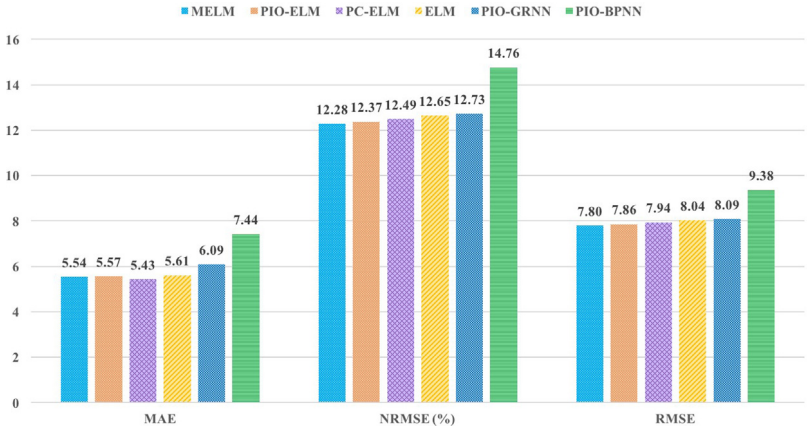

Fig. 6. Horizontal errors of non-decomposition learning models in short-term prediction.

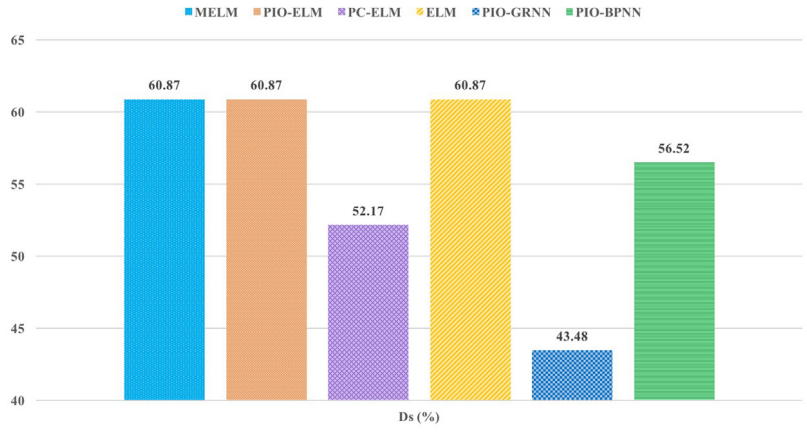

Fig. 7. Directional errors of non-decomposition learning models in short-term prediction.

and next one month respectively. For the one-day prediction, the testing sets are hourly AQI data in 5th May, 2018. For the one-week prediction, the testing sets are hourly AQI data from 29th April to 5th May of 2018. For the one-month prediction, the testing sets are hourly AQI data from 6th April to 5th May of 2018. The rest data are correspondingly assigned as the training sets.

3.2. Performance evaluation criteria and parameter settings

We evaluate the prediction performance from two aspects. We first calculate the forecasting errors and then we make statistical examinations to test the horizontal and directional precision of the forecasting models. Four statistical indicators are employed to calculate the forecasting errors, which are shown in Table 1. Mean absolute error (MAE), normalized root mean square error (NRMSE) and RMSE are utilized to calculate the horizontal forecasting errors, and directional symmetry (Ds) is used to calculate the directional errors.

Here M is the amount of sample points, $z(i)$ and $\hat{z}(i)$ represent the real values and the predicting values of AQI and $\overline{z}$ is the mean value of the real AQI data. If $(\hat{z}(i)-z(i-1))\cdot(z(i)-z(i-1))>0$, $\mu(i)=1$; otherwise, $\mu(i)=0$.

Table 2
DM test outcomes of hybrid decomposition models in short-term prediction.

Benchmark model	Tested model						
	WPD-MELM-MELM	WPD-MELM-ADD	WPD-FEEMD-MELM-ADD	CEEMDAN-MELM-ADD	VMD-MELM-ADD	SSA-MELM-ADD	FEEMD-MELM-ADD
WPD-MELM-MSK-MELM	−1.0805 (0.1456)	−3.7072 (0.0006)	−4.7162 (4.7220e−05)	−2.5186 (0.0096)	−3.1732 (0.0021)	−2.5334 (0.0093)	−6.0470 (1.8120e−06)
WPD-MELM-MELM		−3.0038 (0.0032)	−4.6371 (5.7500e−05)	−2.5159 (0.0097)	−3.1702 (0.0021)	−2.5317 (0.0093)	−6.0444 (1.8240e−06)
WPD-MELM-ADD			−4.1100 (0.0002)	−2.4981 (0.0100)	−3.1439 (0.0023)	−2.5169 (0.0096)	−6.0399 (1.8440e−06)
WPD-FEEMD-MELM-ADD				−2.3971 (0.0125)	−2.9968 (0.0032)	−2.4211 (0.0119)	−5.9693 (2.1830e−06)
CEEMDAN-MELM-ADD					−0.1622 (0.4363)	−1.0541 (0.1514)	−4.1450 (0.0002)
VMD-MELM-ADD						−1.1960 (0.1219)	−4.3656 (0.0001)
SSA-MELM-ADD							−3.5093 (0.0009)

Table 3
DM test outcomes of non-decomposition learning models in short-term prediction.

Benchmark model	Tested model				
	PIO-ELM	PC-ELM	ELM	PIO-GRNN	PIO-BPNN
MELM	−0.3923 (0.3492)	−0.6122 (0.2732)	−0.6122 (0.2732)	−1.1710 (0.1268)	−0.2376 (0.4071)
PIO-ELM		−0.4487 (0.3289)	−1.0320 (0.1564)	−0.1794 (0.4296)	−1.3301 (0.0983)
PC-ELM			−0.4501 (0.3284)	−0.1152 (0.4546)	−1.1901 (0.1231)
ELM				−0.0350 (0.4862)	−1.1375 (0.1335)
PIO-GRNN					−0.8922 (0.1908)

Table 4
PT test outcomes of hybrid decomposition models in short-term prediction.

	WPD-MELM-MSK-MELM	WPD-MELM-MELM	WPD-MELM-ADD	WPD-FEEMD-MELM-ADD	CEEMDAN-MELM-ADD	VMD-MELM-ADD	SSA-MELM-ADD	FEEMD-MELM-ADD
Statistics (*p-value*)	7.8856 (3.1086e−15)	8.2663 (2.2204e−16)	9.0854 (0.0000)	9.0854 (0.0000)	3.4449 (0.0006)	3.3891 (0.0007)	6.5663 (5.1580e−11)	2.7548 (0.0059)

Table 5
PT test outcomes of non-decomposition learning models in short-term prediction.

	MELM	PIO-ELM	PC-ELM	ELM	PIO-GRNN	PIO-BPNN
Statistics (*p-value*)	0.8910 (0.3729)	0.9202 (0.3575)	−0.1808 (0.8565)	0.9837 (0.3253)	−0.7423 (0.4579)	0.4754 (0.6345)

In order to examine the horizontal and directional prediction ability, we also adopt Diebold–Mariano (DM) statistical test [38] and Pesaran–Timmermann (PT) statistical test [39]. DM statistical test is used to test whether the predicting accuracy between test models and benchmark models has significant difference. The square of predicting error is set as the loss function and the null hypothesis is that the benchmark model is more accurate than the test model. PT statistical test is adopted to test whether the models can predict the variation directions accurately and the null hypothesis is the test model cannot predict the variation directions accurately.

To test the forecasting performance of the proposed method synthetically, a series of predicting models are set as comparison models. The comparison models are divided into non-decomposition learning models and hybrid decomposition models. Non-decomposition learning models contain MELM, PIO-ELM, PC-ELM, generalized regression neural network optimized by PIO (PIO-GRNN), PIO-BPNN and ELM. Complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN), is also a kind of classical decomposition approach. Hybrid decomposition models include WPD-MELM-MELM, WPD-MELM-ADD, WPD-FEEMD-MELM-ADD, CEEMDAN-MELM-ADD, VMD-MELM-ADD, SSA-MELM-ADD and FEEMD-MELM-ADD.

In the PIO algorithm, the amount of the pigeons is set as 40, the map and compass factor is assigned as 0.3. The global search algebra is set as 90, and the local search algebra is set as 30. For the comparison models, the number of hidden layer neurons is assigned as 60. And the PIO and PC algorithm is utilizing to optimize the weights and thresholds of ELM. The spread coefficient of GRNN and the weights and thresholds of BPNN are also optimized by the PIO algorithm. And the number of hidden layer neurons is assigned as 10. The hourly AQI series are decomposed into eight subseries. And the subseries are supposed to be clustered into four categories-high frequency subseries, medium–high frequency subseries, medium–low frequency subseries and low frequency subseries.

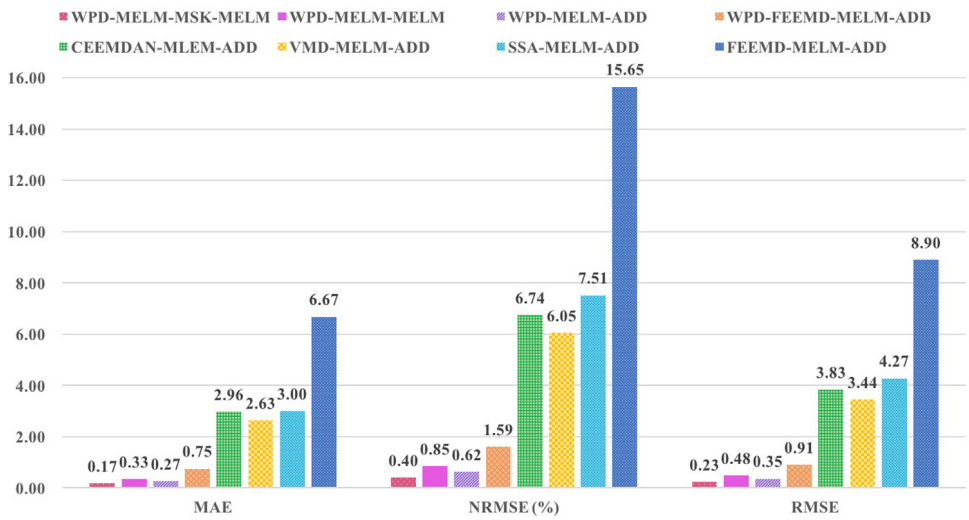

Fig. 8. Horizontal errors of hybrid decomposition models in middle-term prediction.

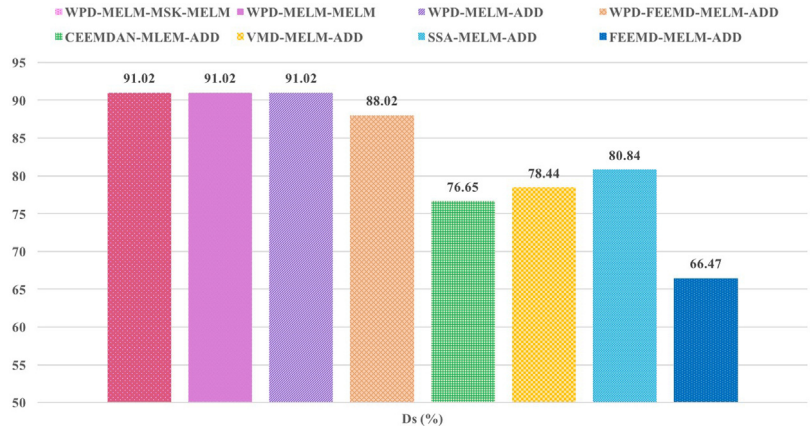

Fig. 9. Directional errors of hybrid decomposition models in middle-term prediction.

3.3. Forecasting result analysis

In this subsection, based on different predictive models, short-term prediction analysis, mid-term prediction analysis and long-term prediction analysis are discussed respectively.

3.3.1. Short-term prediction analysis

Hourly AQI series in Harbin are decomposed into eight lower-frequency subseries by WPD. Subseries of original AQI are displayed in Fig. 3. Subseries 1 represents the overall trend of the original AQI data in Harbin. Through utilizing the MSK clustering method, eight subseries are supposed to be clustered into four categories-high frequency subseries, medium–high frequency subseries, medium–low frequency subseries and low frequency subseries. In order to enhance the predicting precision, we utilize the AQI data of the first five hours to forecast the AQI data in the sixth hour. The forecasting errors of hybrid decomposition models and non-decomposition learning models in short-term prediction are displayed in Figs. 4–7. The outcomes of DM and PT test are shown in Tables 2–5, and the values in brackets are *p-values*. Comparing the prediction errors of short-term prediction, we can come to the following conclusions: (1) The hybrid WPD-MELM-MSK-MELM approach has the best prediction performance on both horizontal and directional accuracy. MAE of the proposed method is 0.03 and the Ds value of the proposed method is 91.30%. The hybrid WPD-MELM-MSK-MELM approach has the best predictable precision among hybrid decomposition models. (2) In non-decomposition learning models, the forecasting accuracy of MELM is higher than other non-decomposition models and PIO-BPNN has the worst forecasting precision in short-term prediction. (3) The horizontal forecasting errors of PIO-ELM are less than PC-ELM, and the Ds values of PIO-ELM and PC-ELM are 60.87% and 52.17% respectively. It demonstrates that the PIO algorithm is more suitable for optimizing the thresholds and weights of ELM than the PC searching algorithm. (4) Among hybrid approaches, FEEMD-MELM-ADD has the largest errors. The MAE, NRMSE and RMSE values of FEEMD-MELM-ADD are 6.26, 11.13% and 7,07 separately. It shows that WPD can effectively decompose the original AQI data. (5) Hybrid models can predict the variation directions accurately; however non-decomposition learning models have bad performance on directional prediction according to the PT test results.

3.3.2. Middle-term prediction analysis

For verifying the forecasting performance of the proposed method, we also make the middle-term prediction analysis of hourly AQI data in Harbin. The prediction errors of hybrid decomposition models and non-decomposition learning models in middle-term predictions are shown in Figs. 8–11, and DM and PT test results are displayed in Tables 6–9. The conclusions are as below: (1) The proposed hybrid method has the highest forecasting

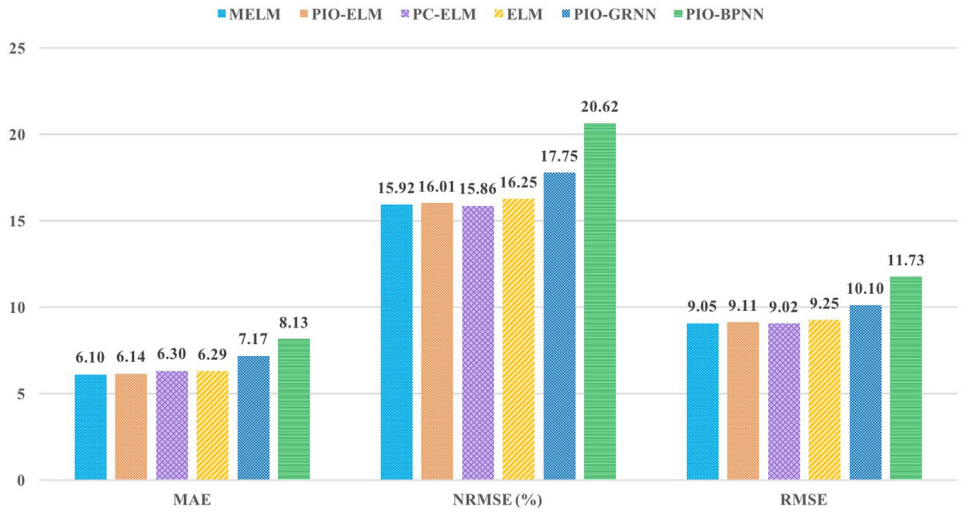

Fig. 10. Horizontal errors of non-decomposition learning models in middle-term prediction.

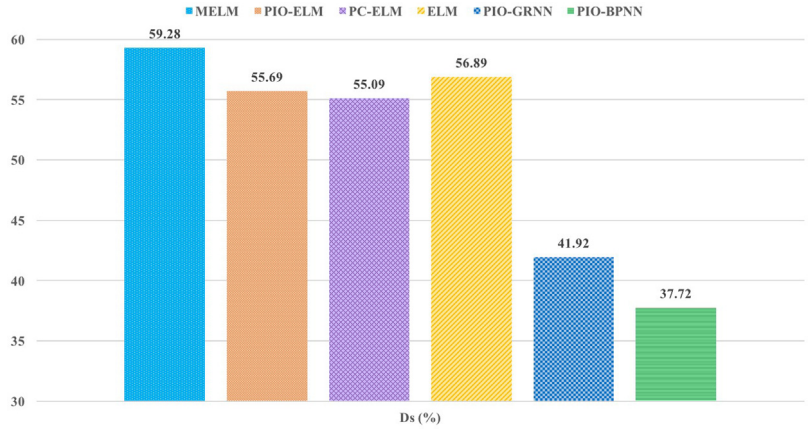

Fig. 11. Directional errors of non-decomposition learning models in middle-term prediction.

accuracy in both short-term prediction and middle-term prediction. The MAE, NRMSE and RMSE errors of WPD-MELM-MSK-MELM are 0.17, 0.40% and 0.21 separately. (2) Comparing the horizontal errors among WPD-MELM-ADD, CEEMDAN-MELM-ADD, VMD-MELM-ADD, SSA-MELM-ADD and FEEMD-MELM-ADD, WPD-MELM-ADD has the best forecasting performance and FEEMD-MELM-ADD has the worst predicting performance. Meanwhile, the predictive capability of VMD-MELM-ADD is better than SSA-MELM-ADD and CEEMDAN-MELM-ADD. (3) The forecasting precision of WPD-MELM-MSK-MELM is better than WPD-MELM-MELM and WPD-MELM-ADD, which shows that the hybrid ensemble approach can enhance the prediction accuracy, and the hybrid ensemble approach is also more suitable than the classical add ensemble approach. (4) The Ds value of WPD-MELM-MSK-MELM is 91.02% and the Ds value of MELM is only 59.28%. It illustrates that the decomposition and ensemble method can improve the forecasting capability of non-decomposition models. (5) The outcomes of DM tests are the same as the forecasting errors. The proposed hybrid WPD-MELM-MSK-MELM method is better than the other learning models. (6) According to the PT test, all hybrid models can predict the changing directions accurately. Among non-decomposition models, PIO-ELM cannot predict the variation directions precisely and MELM can forecast the changing directions accurately. It shows that the IPIO algorithm is better than the PIO algorithm in optimizing the parameters of ELM.

3.3.3. Long-term prediction analysis

In the long-term prediction, the predictive errors of hybrid decomposition models and non-decomposition learning models are displayed in Figs. 12–15, and the DM and PT test outcomes are shown in Tables 10–13. Based on the prediction results of hybrid decomposition models and non-decomposition learning models in long-term prediction, it can be concluded that: (1) In non-decomposition learning models, MELM has the best forecasting performance on the horizontal and directional accuracy. MELM also has excellent performance among the short-term prediction, middle-term prediction and long-term prediction. The MAE, NRMSE and RMSE values of the MELM approach are 9.89, 20.40% and 16.97 respectively. However, the PIO-BPNN has the worst performance on both horizontal and directional accuracy. The Ds value of PIO-BPNN is 36.16%, which is nearly half of MELM. It shows that the IPIO learning algorithm can effectively improve the PIO algorithm for parameter optimization. (2) In hybrid decomposition models, the proposed WPD-MELM-MSK-MELM approach still has the best forecasting capability in the long-term prediction. The MAE, NRMSE, RMSE and Ds values of WPD-MELM-M-K-MELM are 0.45, 0.74%, 0.61 and 90.69% respectively, which are the lowest among the whole hybrid models. (3) WPD has better decomposition capability than VMD, SSA, CEEMDAN and FEEMD. The forecasting errors of WPD-MELM-MSK-MELM are also lower than WPD-FEEMD-MELM-ADD. It demonstrates that the two-layer decomposition can become a reason of over-fitting

Table 6
DM test outcomes of hybrid decomposition models in middle-term prediction.

Benchmark model	Tested model						
	WPD-MELM-MELM	WPD-MELM-ADD	WPD-FEEMD-MELM-ADD	CEEMDAN-MELM-ADD	VMD-MELM-ADD	SSA-MELM-ADD	FEEMD-MELM-ADD
WPD-MELM-MSK-MELM	−2.4204 (0.0119)	−4.7418 (4.4300e−05)	−4.9843 (2.4240e−05)	−3.2916 (0.0016)	−2.6562 (0.0071)	−2.7268 (0.0060)	−5.1126 (1.7650e−05)
WPD-MELM-MELM		−3.2937 (0.0016)	−4.7131 (4.7580e−05)	−3.2836 (0.0016)	−2.6499 (0.0072)	−2.7231 (0.0061)	−5.1112 (1.7710e−05)
WPD-MELM-ADD			−4.2993 (0.0001)	−3.2739 (0.0017)	−2.6379 (0.0074)	−2.7167 (0.0062)	−5.1057 (1.7950e−05)
WPD-FEEMD-MELM-ADD				−3.1644 (0.0022)	−2.5040 (0.0010)	−2.6502 (0.0072)	−5.0749 (1.9370e−05)
CEEMDAN-MELM-ADD					0.2774 (0.6080)	−1.5072 (0.0727)	−4.4082 (0.0001)
VMD-MELM-ADD						−1.5693 (0.0651)	−4.5093 (7.9080e−05)
SSA-MELM-ADD							−4.0244 (0.0003)

Table 7
DM test outcomes of non-decomposition learning models in middle-term prediction.

Benchmark model	Tested model				
	PIO-ELM	PC-ELM	ELM	PIO-GRNN	PIO-BPNN
MELM	−0.6146 (0.2724)	0.1573 (0.5618)	0.2426 (0.5948)	−0.7424 (0.2327)	−1.6030 (0.0613)
PIO-ELM		0.4282 (0.6638)	1.1294 (0.8648)	−0.7224 (0.2387)	−1.5866 (0.0631)
PC-ELM			−0.0413 (0.4837)	−1.0799 (0.1457)	−1.8815 (0.0363)
ELM				−0.8631 (0.1985)	−1.6920 (0.0521)
PIO-GRNN					−1.5585 (0.0664)

Table 8
PT test outcomes of hybrid decomposition models in middle-term prediction.

	WPD-MELM-MSK-MELM	WPD-MELM-MELM	WPD-MELM-ADD	WPD-FEEMD-MELM-ADD	CEEMDAN-MELM-ADD	VMD-MELM-ADD	SSA-MELM-ADD	FEEMD-MELM-ADD
Statistics (*p-value*)	19.0827 (0.0000)	19.0827 (0.0000)	19.1548 (0.0000)	15.4988 (0.0000)	8.3064 (0.0000)	9.2839 (0.0000)	10.2339 (0.0000)	4.6053 (4.1187e−06)

Table 9
PT test outcomes of non-decomposition learning models in middle-term prediction.

	MELM	PIO-ELM	PC-ELM	ELM	PIO-GRNN	PIO-BPNN
Statistics (*p-value*)	3.2471 (0.0012)	2.3188 (0.0204)	1.9620 (0.0498)	2.7864 (0.0053)	−1.8253 (0.0680)	−2.9823 (0.0029)

and WPD is more suitable for decomposing the AQI data. The outcomes of DM and PT tests are also the same as the error analysis results. It illustrates that the MELM and the WPD-MELM-MSK-MELM methods have the best predictable performance among the non-decomposition learning models and hybrid decomposition models.

3.4. Robustness and discussion

In order to test the compare the predictive capability of the proposed hybrid WPD-MELM-MSK-MELM approach, robustness analysis and comparison analysis are discussed respectively. For testing the robustness of the proposed method, AQI data of Wuhan from 0:00 in 1th December 2016 to 23:00 in 5th May 2018 are utilized to make the short-term prediction, middle-term prediction and long-term prediction separately. We calculate the forecasting errors of the hybrid method, and the prediction errors are displayed in Table 14. According to the forecasting errors in Table 14, it is clearly that the proposed hybrid can predict the AQI data in Wuhan accurately. It illustrates that the proposed WPD-MELM-MSK-MELM model has good performance on robustness.

For verifying the forecasting ability of the proposed method, we also compare the predictive accuracy with the approach presented in [26] and [27]. We utilize the methods to make short-term, middle-term and long-term predictions of AQI in Harbin. The forecasting errors of the two comparison models are shown in Tables 15–16. Compared with the two hybrid models proposed in 2017, we can find that the horizontal prediction errors of WPD-MELM-MSK-MELM is much smaller than VMD-PSO-BPNN-ADD and CEEMD-VMD-DE-ELM-ADD. The Ds values of WPD-MELM-MSK-MELM in short-term prediction, middle-term prediction and long-term predictions are 91.30%, 91.02% and 90.96% respectively. The maximum Ds values of the two compared models is 79.00%. It shows that the proposed WPD-MELM-MSK-MELM method has better performance on both horizontal accuracy and directional accuracy.

4. Conclusions

We present a novel hybrid WPD-MELM-MSK-MELM learning approach for predicting AQI. In the proposed method, WPD is

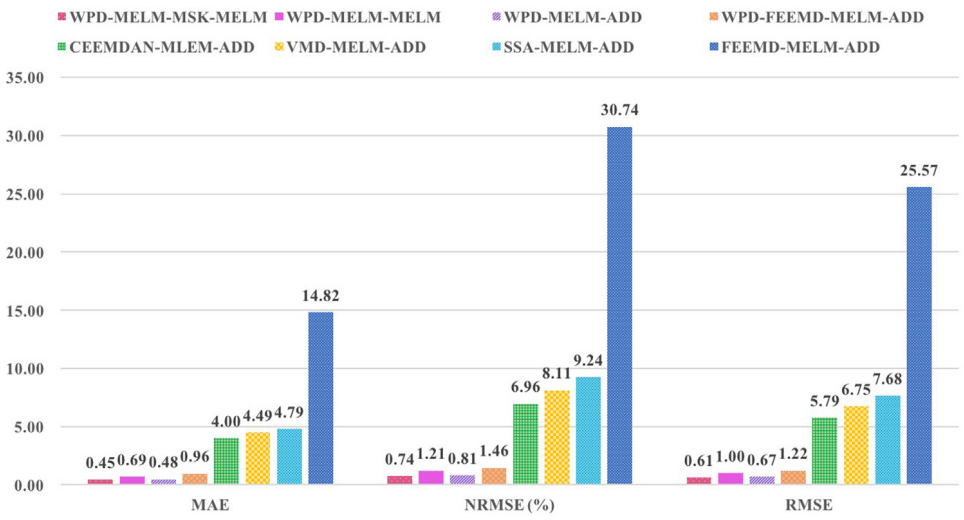

Fig. 12. Horizontal errors of V1 in long-term prediction.

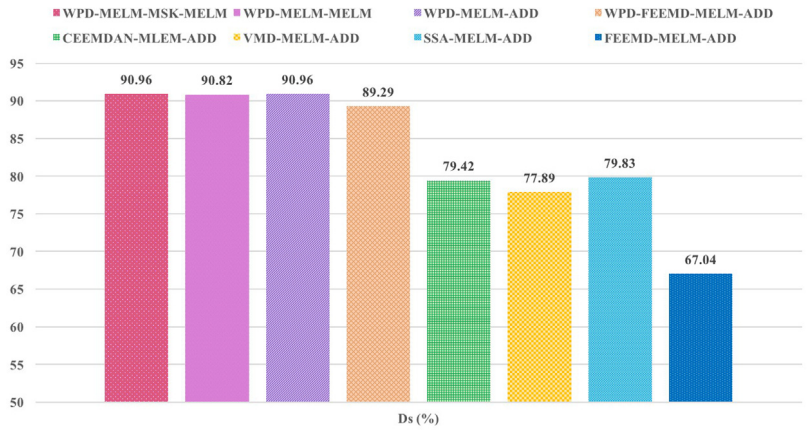

Fig. 13. Directional errors of hybrid decomposition models in long-term prediction.

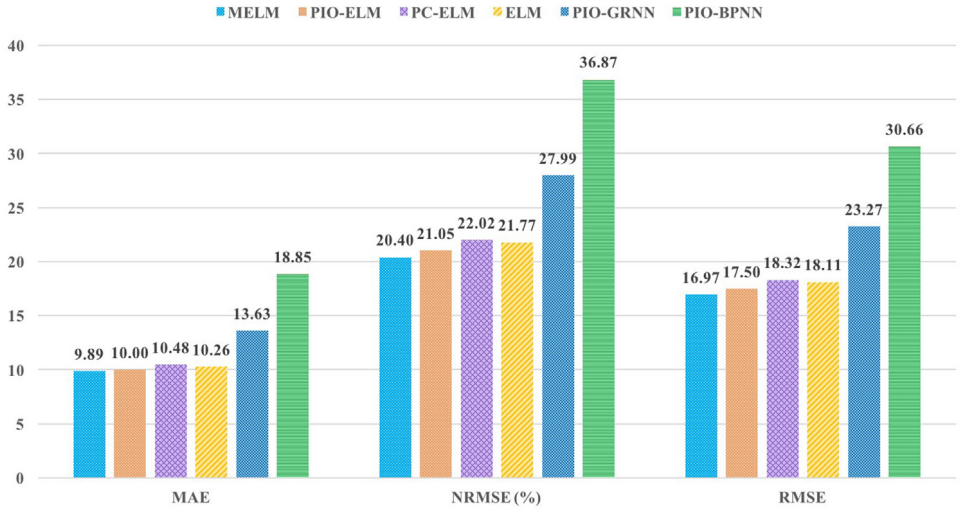

Fig. 14. Horizontal errors of non-decomposition learning models in long-term prediction.

applied to decompose the AQI data, PIO is modified by PSO and then IPIO is utilized to optimize the parameters of ELM, MELM is employed to forecast and ensemble the subseries and MSK are adopted to cluster the forecasting outcomes. For further testing the predictive capability of the proposed WPD-MELM-MSK-MELM method, we make the short-term, middle-term and long-term predictions of the hourly AQI data in Harbin and compare with five kinds of decomposition approaches. The forecasting outcomes illustrate that the hybrid WPD-MELM-MSK-MELM

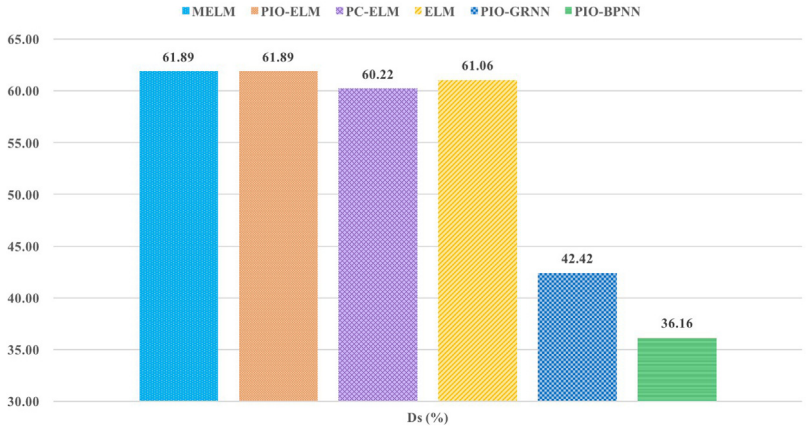

Fig. 15. Directional errors of non-decomposition learning models in long-term prediction.

Table 10
DM test outcomes of hybrid decomposition models in long-term prediction.

Benchmark model	Tested model						
	WPD-MELM-MELM	WPD-MELM-ADD	WPD-FEEMD-MELM-ADD	CEEMDAN-MELM-ADD	VMD-MELM-ADD	SSA-MELM-ADD	FEEMD-MELM-ADD
WPDMELM-MSK-MELM	−2.5098 (0.0098)	−1.1373 (0.1336)	−2.5580 (0.0088)	−3.0155 (0.0031)	−3.0848 (0.0026)	−2.0596 (0.0255)	−2.0258 (0.0273)
WPD-MELM-MELM		3.1544 (0.9978)	−2.3530 (0.0138)	−3.0079 (0.0031)	−3.0565 (0.0028)	−2.0596 (0.0255)	−2.0133 (0.0280)
WPD-MELM-ADD			−2.5389 (0.0092)	−3.0233 (0.0030)	−3.0859 (0.0026)	−2.0606 (0.0254)	−2.0254 (0.0273)
WPD-FEEMD-MELM-ADD				−3.0669 (0.0027)	−3.0766 (0.0027)	−2.0411 (0.0264)	−2.0067 (0.0283)
CEEMDAN-MELM-ADD					−1.0427 (0.1540)	−1.6838 (0.0529)	−1.5896 (0.0628)
VMD-MELM-ADD						−0.6015 (0.2767)	−1.6502 (0.0563)
SSA-MELM-ADD							−1.1304 (0.1350)

Table 11
DM test outcomes of non-decomposition learning models in long-term prediction.

Benchmark model	Tested model				
	PIO-ELM	PC-ELM	ELM	PIO-GRNN	PIO-BPNN
MELM	1.2208 (0.8827)	0.6304 (0.7327)	1.5885 (0.9371)	−0.7475 (0.2312)	−1.6123 (0.0603)
PIO-ELM		0.1358 (0.5534)	0.5727 (0.7138)	−0.9206 (0.1834)	−1.6967 (0.0516)
PC-ELM			0.1499 (0.5589)	−0.9174 (0.1842)	−1.7118 (0.0502)
ELM				−0.9412 (0.1782)	−1.6893 (0.0523)
PIO-GRNN					−1.9637 (0.0309)

Table 12
PT test outcomes of hybrid decomposition models in long-term prediction.

	WPD-MELM-MSK-MELM	WPD-MELM-MELM	WPD-MELM-ADD	WPD-FEEMD-MELM-ADD	CEEMDAN-MELM-ADD	VMD-MELM-ADD	SSA-MELM-ADD	FEEMD-MELM-ADD
Statistics (*p-value*)	38.8219 (0.0000)	38.4430 (0.0000)	38.8693 (0.0000)	34.4523 (0.0000)	19.6879 (0.0000)	18.5137 (0.0000)	20.0011 (0.0000)	9.7511 (0.0000)

Table 13
PT test outcomes of non-decomposition learning models in long-term prediction.

	MELM	PIO-ELM	PC-ELM	ELM	PIO-GRNN	PIO-BPNN
Statistics (*p-value*)	7.3323 (2.2626e−13)	7.3776 (1.6120e−13)	6.1441 (8.0418e−10)	6.7780 (1.2185e−11)	−4.2228 (2.4129e−05)	−7.6978 (1.3843e−14)

method have excellent performance on horizontal accuracy, directional accuracy and robustness and WPD is more suitable than VMD, SSA, CEEMDAN and FEEMD to decompose AQI. The empirical results also show that the proposed approach has good

Table 14
Forecasting errors of AQI in Wuhan.

	MAE	NRMSE (%)	RMSE	Ds (%)
Short-term prediction	0.1531	0.3139	0.1823	100.0000
Middle-term prediction	0.2344	0.5229	0.3308	93.4132
Long-term prediction	0.2888	0.6138	0.5470	90.1252

Table 15
Forecasting errors of VMD-PSO-BPNN-ADD [26].

	MAE	NRMSE (%)	RMSE	Ds (%)
Short-term prediction	2.2237	4.9920	3.1720	73.9130
Middle-term prediction	2.9655	6.7453	3.8368	76.6467
Long-term prediction	3.9896	6.9102	5.7471	78.9986

Table 16
Forecasting errors of CEEMD-VMD-DE-ELM-ADD [27].

	MAE	NRMSE (%)	RMSE	Ds (%)
Short-term prediction	3.3884	7.0888	4.5043	73.9130
Middle-term prediction	5.2650	12.0424	6.8498	62.2754
Long-term prediction	7.9901	14.4952	12.0553	71.6273

predictable precision on short-term prediction, middle-term prediction and long-term prediction. In future studies, we will utilize the deep learning models and multi-objective optimization methods or other optimization approaches to improve the forecasting precision. Meanwhile, we will also take the seasonality, weather, holiday and other factors into account for the prediction of air pollution concentrations. We can also adopt the proposed method to predict other air pollution concentrations, wind speed, electricity price and other nonstationary time series.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.asoc.2019.105827.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61773401, 11601524, 11571368), the Foundation of Hubei Province of China (Grant Nos. 17G024 and 2017132) and the Australian Research Council Future Fellowship (Grant No. FT100100748).

Appendix. List of abbreviations

Here, all terms mentioned in this paper and their definitions are listed in alphabetical order:
AQI — air quality index
BPNN — back propagation neural network
CEEMD — complementary ensemble empirical mode decomposition
CEEMDAN — complementary ensemble empirical mode decomposition with adaptive noise
CEEMDAN-MELM-ADD — the method utilizing CEEMDAN as the decomposition approach, MELM as the predicting approach and ADD as the ensemble approach
CEEMD-VMD-DE-ELM — the method utilizing CEEMD and VMD as the decomposition approaches, DE as the optimization approach, ELM as the predicting approach and ADD as the ensemble approach
CPSO — chaotic particle swarm optimization
DE — differential evolution
DM — Diebold–Mariano
Ds — directional symmetry
EEMD — ensemble empirical mode decomposition
ELM — extreme learning machine
FEEMD — fast ensemble empirical mode decomposition
FEEMD-MELM-ADD — the method utilizing FEEMD as the decomposition approach, MELM as the predicting approach and ADD as the ensemble approach
GA-WNN — the hybrid approach combining wavelet neural network and genetic algorithm
GRNN — generalized regression neural network
IPIO — improved pigeon-inspired optimization
MAE — Mean absolute error
MELM — modified extreme learning machine
MSK — multidimensional scaling and K-means
MSK — Multidimensional scaling
NRMSE — normalized root mean square error
PC — parallel chaos algorithm
PC-ELM — the forecasting method utilizing the PC algorithm to parameters of ELM
PIO — pigeon-inspired optimization
PIO-BPNN — the forecasting method utilizing the PIO algorithm to parameters of BPNN
PIO-ELM — the forecasting method utilizing the PIO algorithm to parameters of ELM
PIO-GRNN — the forecasting method utilizing the PIO algorithm to parameters of GRNN
PSO — particle swarm optimization
PT — Pesaran–Timmermann
RMSE — root mean square error
SSA — singular spectrum analysis
SSA-MELM-ADD — the method utilizing SSA as the decomposition approach, MELM as the predicting approach and ADD as the ensemble approach
VMD — variational mode decomposition
VMD-MELM-ADD — the method utilizing VMD as the decomposition approach, MELM as the predicting approach and ADD as the ensemble approach
VMD-PSO-BPNN-ADD — the method utilizing VMD as the decomposition approach, PSO as the optimization approach, BPNN as the predicting approach and ADD as the ensemble approach
WD — wavelet decomposition
WPD — wavelet packet decomposition
WPD-FEEMD-MELM-ADD — the method utilizing WPD and FEEMD as the decomposition approaches, MELM as the predicting approach and ADD as the ensemble approach
WPD-MELM-ADD — the method utilizing WPD as the decomposition approach, MELM as the predicting approach and ADD as the ensemble approach
WPD-MELM-MELM — the method utilizing WPD as the decomposition approach, MELM as the predicting approach and MELM as the ensemble approach
WPD-MELM-MSK-MELM — the method utilizing WPD as the decomposition approach, MELM as the predicting approach, MSK as the clustering approach and MELM as the ensemble approach

References

[1] G. Asadollahfardi, H. Zangooei, S.H. Aria, $PM_{2.5}$ concentrations using artificial neural networks and markov chain, a case study karaj city, Asian J. Atmospheric Environ. 10 (2016) 67–79.
[2] Y. Xiang, L. Gou, L. He, S.L. Xia, W.Y. Wang, A SVR-ann combined model based on ensemble EMD for rainfall prediction, Appl. Soft Comput. 73 (2018) 874–883.
[3] N.E. Johnson, B. Bonczak, C.E. Kontokosta, Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment, Atmos. Environ. 184 (2018) 9–16.

[4] U.K. Singh, V. Padmanabhan, A. Agarwal, Dynamic classification of ballistic missiles using neural networks and hidden Markov models, Appl. Soft Comput. 19 (2014) 280–289.
[5] T.W. Huang, C.D. Li, S.K. Duan, A.S. Janusz, Robust exponential stability of uncertain delayed neural networks with stochastic perturbation and impulse effects, IEEE Trans. Neural Netw. Learn. Syst. 23 (2012) 866–875.
[6] M. Oprea, S.F. Mihalache, M. Popescu, Modeling missing data for $PM_{2.5}$ time series forecasting with computational intelligence, Int. J. Comput. Commun. Control 12 (2017) 365–380.
[7] Y.X. Wang, M.Q. Liu, Z.J. Bao, S.L. Zhang, Short-term load forecasting with multi-source data using gated recurrent unit neural networks, Energies 11 (2018) 1–19.
[8] S. Zhu, Q.Q. Yang, S. Yi, Noise further expresses exponential decay for globally exponentially stable time-varying delayed neural networks, Neural Netw. 77 (2016) 7–13.
[9] H.D. He, M. Li, W.L. Wang, Z.Y. Wang, Y. Xue, Prediction of $PM_{2.5}$concentration based on the similarity in air quality monitoring network, Build. Environ. 137 (2018) 11–17.
[10] C.J. Huang, P.H. Kuo, A deep CNN-LSTM model for particulate matter ($PM_{2.5}$) forecasting in smart cities, Sensors 18 (2018) 2220.
[11] Z.Y. Wang, L. Lu, H.D. He, Fine-scale estimation of carbon monoxide and fine particulate matter concentrations in proximity to a road intersection by using wavelet neural network with genetic algorithm, Atmos. Environ. 104 (2015) 264–272.
[12] Y.M. Yang, Y.N. Wang, X.F. Yuan, Parallel chaos search based incremental extreme learning machine, Neural Process. Lett. 37 (2013) 277–301.
[13] H.D. He, W.Z. Lu, Y. Xue, Prediction of particulate matter at urban intersection by using artificial neural networks combined with chaotic particle swarm optimization algorithm, Build. Environ. 78 (2014) 111–117.
[14] A.A. Heidari, P. Pahlavani, An efficient modified grey wolf optimizer with Lévy flight for optimization tasks, Appl. Soft Comput. 60 (2017) 115–134.
[15] H.B. Duan, P.X. Qiao, Pigeon-inspired optimization: a new swarm intelligence optimizer for air robot path planning, Int. J. Intell. Comput. Cybern. 7 (2014) 24–37.
[16] T. Li, C. Zhou, B. Wang, B.X. Xiao, X.D. Zheng, A hybrid algorithm based on artificial bee colony and pigeon inspired optimization for 3D protein structure prediction, J. Bionanosci. 12 (2018) 100–108.
[17] F. Jiang, J.Q. He, Z.G. Zeng, Pigeon-inspired optimization and extreme learning machine via wavelet packet analysis for predicting bulk commodity futures prices, Sci. China Inf. Sci. 62 (2019) 070204.
[18] R. Dou, H.B. Duan, Pigeon inspired optimization approach to model prediction control for unmanned air vehicles, Aircr. Eng. Aerosp. Technol. 88 (2016) 108–116.
[19] Y. Cheng, H. Zhang, Z.H. Liu, Hybrid algorithm for short-term forecasting of $PM_{2.5}$ in China, Atmos. Environ. 200 (2019) 264–279.
[20] H. Luo, D.Y. Wang, C.Q. Yue, Research and application of a novel hybrid decomposition-ensemble learning paradigm with error correction for daily PM_{10}, Atmos. Res. 201 (2018) 34–45.
[21] S.L. Sun, S.Y. Wang, G.W. Zhang, J.L. Zheng, A decomposition-clustering-ensemble learning approach for solar radiation forecasting, Sol. Energy 163 (2018) 189–199.
[22] J.M. Zhu, P. Wu, H.Y. Chen, A hybrid forecasting approach to air quality time series based on endpoint condition and combined forecasting model, Int. J. Environ. Res. Public Health 15 (2018) 1941.
[23] K. Gan, S.L. Sun, S.Y. Wang, Y.J. Wei, A secondary-decomposition-ensemble learning paradigm for forecasting $PM_{2.5}$ concentration, Atmospheric Pollut. Res. 9 (2018) 989–999.
[24] F. Jiang, J.Q. He, Z.G. Zeng, T.H. Tian, A decomposition-optimization-ensemble learning approach for electricity price forecasting (in Chinese), Sci. China (Information Sciences) 48 (2018) 1300–1315.
[25] A. Zhigljavsky, Singular spectrum analysis for time series: introduction to this special issue, Stat. Interface 3 (2010) 1499–1512.
[26] D.Y. Wang, C.Q. Yue, S. Wei, J. Lv, Performance analysis of four decomposition-ensemble models for one-day-ahead agricultural commodity futures price forecasting, Algorithms 10 (2017) 108.
[27] D.Y. Wang, S. Wei, H.Y. Luo, C.Q. Yue, O. Grunder, A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine, Sci. Total Environ. 580 (2017) 719–733.
[28] M.V. Wickerhauser, Acoustic signal compression with wavelet packets, Wavelets 2 (1992) 679–700.
[29] T. Wu, G.Z. Yan, B.H. Yang, H. Sun, EEG feature extraction based on wavelet packet decomposition for brain computer interface, Measurement 41 (2008) 618–625.
[30] D.F. Zhang, H.B. Duan, Social-class pigeon-inspired optimization and time stamp segmentation for multi-UAV cooperative path planning, Neurocomputing 313 (2018) 229–246.
[31] K.S.S. Ranjini, S. Murugan, Memory based hybrid dragonfly algorithm for numerical optimization problems, Expert Syst. Appl. 83 (2017) 63–78.
[32] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[33] T. Wu, M. Yao, J.H. Yang, Dolphin swarm extreme learning machine, Cognitive Computation 9 (2017) 275–284.
[34] D.Q. Yan, Y.H. Chu, H.Y. Zhang, D.S. Liu, Information discriminative extreme learning machine, Soft Comput. 22 (2018) 677–689.
[35] R. Ahila, V. Sadasivam, K. Manimala, An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances, Appl. Soft Comput. 32 (2015) 23–37.
[36] T.F. Cox, M. Cox, Multidimensional scaling, J. R. Stat. Soc. 46 (2001) 1050–1057.
[37] X.L. Hu, J.W. Zhang, P. Qi, B. Zhang, Modeling response properties of V2 neurons using a hierarchical K-means model, Neurocomputing 134 (2014) 198–205.
[38] F.X. Diebold, R.S. Mariano, Comparing predictive accuracy, J. Bus. Econ. Stat. 20 (2002) 134–144.
[39] M.H. Pesaran, A. Timmermann, A simple nonparametric test of predictive performance, J. Bus. Econ. Stat. 10 (1992) 461–465.