DOI:10.28974/idojaras.2023.1.4

IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 127, No. 1, January – March, 2023, pp. 55–76

A novel ensemble wind speed forecasting method using the differential weighting scheme and principal component analysis

Laleh Parviz

Faculty of Agriculture
Azarbaijan Shahid Madani University
35 Km Tabriz- Maragheh Road,
Tabriz, Iran P.O.B: 53714-161

Corresponding author E-mail: laleh_parviz@yahoo.com

(Manuscript received in final form April 6, 2022)

Abstract— Wind speed forecasting has found economic significance as it can increase operational efficiency. In this regard, an accurate forecast of wind speed is crucial in the application of wind resources. This study is intended to incorporate independent and output variables as the input of support vector regression (SVR) to forecast wind speed of Zanjan and Ahvaz stations in Iran. The independent variables were minimum, maximum, and mean temperatures, relative humidity, precipitation, average visibility, and dew point temperature. The incorporation of independent and output variables were conducted with principal component analysis (PCA) and differential weighting scheme (DWS), respectively. DWS combined the forecasts of linear regression, SVR, and group method of data handling (GMDH) in which the SVR showed the best The forecast of DWS outperformed the other three mentioned models. The incorporation of DWS and PCA (DWS-PCA) improved the forecasts and the capability of DWS-PCA as a novel method was significant in terms of forecast stability. The novel method can be a robust approach for wind speed forecasting in some subjects such as renewable energy, and meteorological decisions.

Key-words: wind speed forecasting, incorporation, differential weighting scheme, principal component analysis

1. Introduction

Weather forecasting plays an essential role in our daily life as increasing losses are reported every day due to extreme weather events. Timely weather warnings such as high wind warning can protect many lives (*Pan et al*., 2021). Wind forecasting is an important issue in hydrological and meteorological decisions. Accurate wind speed forecasting is necessary for the stable function of wind turbines to generate wind power. Wind power can be effectively managed through wind speed monitoring. Furthermore, renewable and non-polluting energy sources have recently found increasing popularity due to global warming effects. Wind energy can be one of the fundamental components of green energy sources. The appropriate operation of wind turbines and the optimal power generation can be achieved considering wind speed as an essential element. The difficulty in wind forecasting can be assigned to the periodic nature of wind speed (*Jaseena* and *Kovoor*, 2021) making the development of an accurate wind forecasting method a challenging task. Physical, statistical, and hybrid approaches can be exploited in wind speed forecasting. Physical models use mathematical concepts with historical data to forecast wind speed. The historical time series can help to forecast future data by statistical models. Hybrid models are composed of two or more forecasting models with their performance which can outperform single models (*Jaseena* and *Kovoor,* 2021).

Wind speed was predicted with generalized regression neural network (GRNN) and multi-layer precipitation (MLP) in some regions of India, considering longitude, latitude, daily horizontal solar irradiance, relative humidity, air temperature, elevation, earth temperature, cooling degree-days, heating degree-days, and atmospheric pressure as the input variables. The accuracy of GRNN was higher than that of MLP (*Kumar* and *Malik,* 2016). Four artificial intelligence methods including artificial neural networks (ANN) with radial basis function, adaptive neuro-fuzzy inference system, ANN with a genetic algorithm, and ANN with particle swarm optimization (PSO) were used for wind speed forecasting. The minimum root mean square error (RMSE) was related to ANN-PSO for Tehran, Iran (*Fazelpour et al.,* 2016). The multilayer feed-forward neural network (MLFFNN), support vector regression (SVR) with radial basis function, and adaptive neuro-fuzzy inference system (ANFIS) were employed to predict wind speed and direction in Bushehr. The input variables were temperature, pressure, local time, and relative humidity. ANFIS was optimized with the PSO method (ANFIS-PSO). The evaluation indices showed that the SVR model outperformed the MLFFNN and ANFIS-PSO models (*Khosravi et al*., 2018). Another study in association with wind speed prediction is a comparative analysis of ANN and chaotic time series forecasting. The results indicated that the neural network approaches outperformed the chaotic model (*Jamil* and *Zeeshan,* 2019). A multi-variate long short-term memory (MV-LSTM) network was also proposed for short-term forecasting of wind speed considering meteorological variables of temperature, humidity, and air pressure. The superiority of MV-LSTM to ARMA was proved for short-term forecasting of wind speed (*Xie et al.,* 2021). In addition, hybrid models were also proposed to forecast wind speed. Two hybrid models consisting of neural network and neuro-fuzzy models combined with wavelet were introduced for monthly wind forecasting in Yazd (Iran) considering variables such as wind speed, maximum temperature, mean temperature, evaporation, and relative humidity. The results showed the high performance of the wavelet neural-fuzzy method (*Afkhami et al.,* 2015).

The decomposition-based hybrid deep BiDLSTM (Bidirectional Long Short Term Memory) models with skip connection were proposed by *Jaseena* and *Kovoor* (2021) which showed superior performance in terms of wind speed forecasting. The hybrid machine intelligence using variants of SVR (ε-SVR, LS-SVR, ε-twin support vector regression (ε-TSVR), twin support vector regression (TSVR)) was employed for wind forecasting in four wind farm sites. The ε-TSVR outperformed the other models (*Dhiman et al*., 2019). Another hybrid model was based on data division and a deep learning network for efficient short-term wind speed prediction. This system could improve the accuracy of forecasts relative to the other conventional methods (*Liu et al*., 2021). Generally, the combined forecasting approaches outperform the single models and can be a good choice for wind speed forecasting. In the combination approaches, the information can be derived from the single models through a precise method with high forecasting performance in various fields. For instance, a multi-granularity heterogeneous combination method was employed for forecasting crude oil (*Wang et al*., 2020). Moreover, different combination methods such as mean, linear regression, nonlinear regression with machine learning algorithm were adopted to forecast day-ahead solar power (*Dewangan et al.,* 2020), or different weight combination methods such as inverse variance method and simple weight average method were used for air quality forecasting (*Song* and *Fu*, 2020). Along with the combination of dependent variables (named as forecast combination), there is another method with emphasis on the combination of independent variables. The method with increased forecasting accuracy uses principal component analysis (PCA) as model inputs rather than original variables (*UI-Saufie et al*., 2011).

The present study is thus aimed at the development of a novel approach for monthly wind speed forecasting. In this regard, an ensemble of independent and output variables was carried out with the PCA and the forecast combination method (DWS), respectively. The forecast combination method combined the results of linear regression, SVR, and group method of data handling (GMDH) models. The input of single models was meteorological variables such as temperature.

2. Material and Methods

2.1. Case study

Data observed at Ahvaz and Zanjan meteorological stations in Iran were used in this study. The studied provinces are Khuzestan and Zanjan (*Fig 1.a)* and the location of stations is shown in *Figs. 1d* and *e*. The studied period was from 2008 to 2020 during which the selected calibration period was from 2008 to 2016. The monthly times series of wind speed are presented in *Figs. 1b* and *f*. The average wind speeds for the study period in Ahvaz and Zanjan stations are 8.12 and 11.8 km/h, respectively. The climate of Ahvaz and Zanjan stations can be classified as arid and semi-arid based on the De Martonne's climate classification (*De Martonne*, 1925). To investigate the monthly temperature and precipitation variations, the embrothermic diagram of each station was drawn in *Figs. 1c* and *g* as reported by *Emberger et al*. (1963). In the embrothermic diagram of Ahvaz station, except for two months (November and December), the temperature was higher than the precipitation, suggesting the arid climate of Ahvaz. In Zanjan station, the points with higher precipitation compared to temperature were equal to those whose temperature was higher than the precipitation.

The forecast combination approach is focused on the integration of competing forecasts considering the superiority of individual forecasts of the models (*Wang et al*., 2020). Regarding the strong effect of forecast combination methods on forecasting issues, in this study, the incorporation of input or independent variables and output or dependent variables were conducted. PCA was used for input variables combination. To combine output variables or forecasted data from some models, DWS was also utilized using wind speed forecasting models such as linear regression (LR), GMDH, and SVR. The incorporation of PCA and DWS was finally conducted by SVR (PCA-DWS). *Fig. 2* shows the steps related to the performance of the novel method.

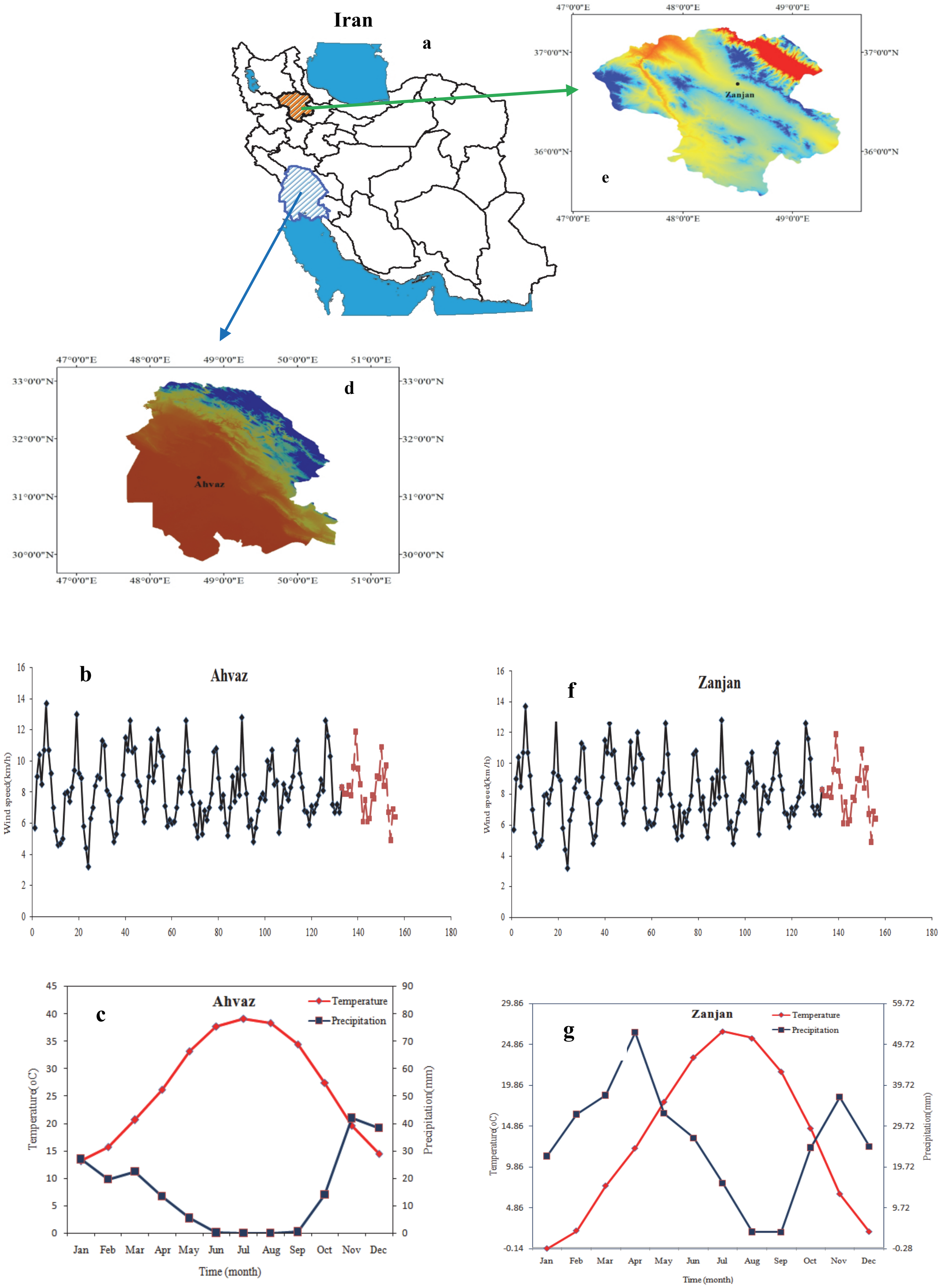

Fig. 1. Map of Iran (a), monthly time series of wind speed for Ahvaz (b) and Zanjan (f) stations during 2008–2020, embrothermic diagram of Ahvaz (c) and Zanjan (g) stations, location of Ahvaz (d) and Zanjan (e) stations in Khuzestan and Zanjan provinces.

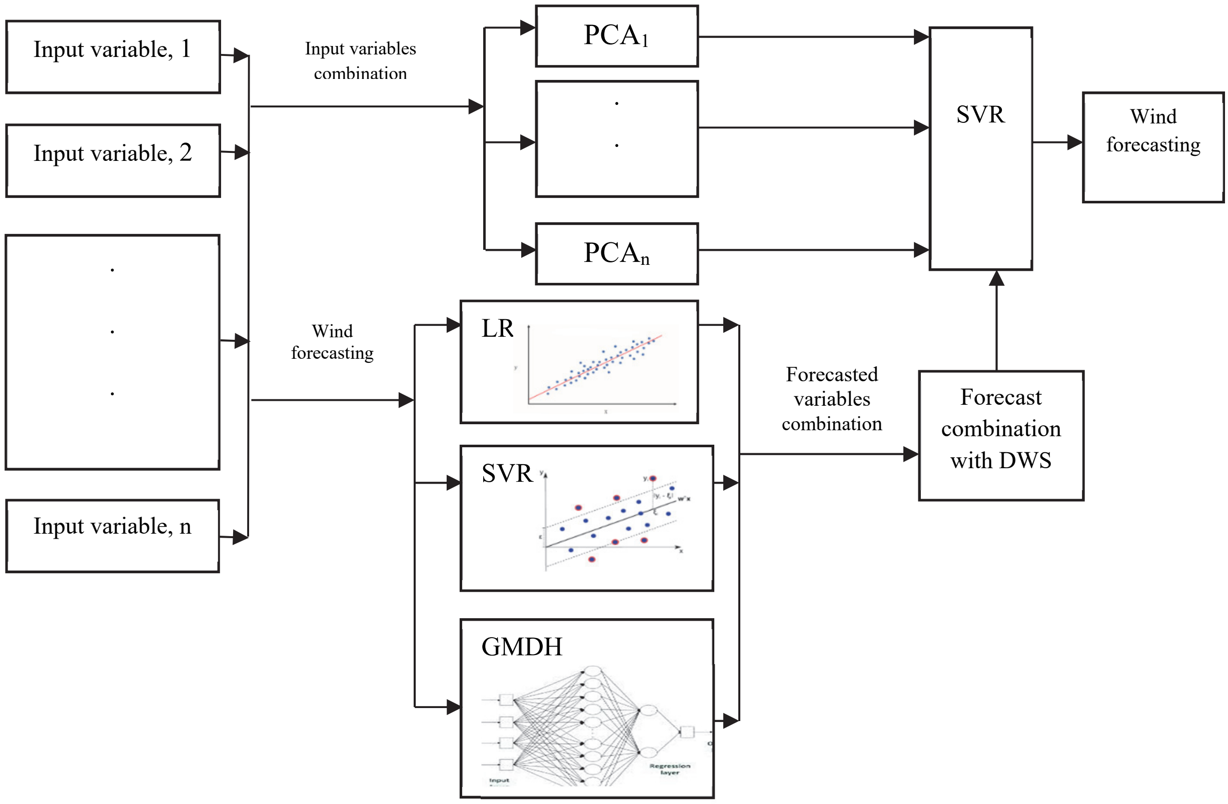

Fig. 2. Different stages of the proposed method performance.

2.2. *Principal component analysis (PCA)*

PCA can be defined as a linear combination of original variables and the weights assigned to the linear combination of the original variables also known as eigenvectors (*Wang* and *Wang,* 2015). In *Fig. 2,* all components were not used; only components with eigenvalues more than 1 were selected. The performance of PCA as a reduction method can be also described: *M* can be considered a *t*-dimensional data set. The retained variance is maximal orthonormal onto principal axes of $G_1, G_2,\ldots, G_N$ in the projected space. $G_1, G_2,\ldots, G_N$ are obtained with *n* leading eigenvectors of sample covariance

$$matrixC = \left(\frac{1}{L}\right) \sum_{k=1}^{L} (x_k - m)^T (x_k - m) \quad , \tag{1}$$

where *m* is the samples average and *L* is the samples number (*Avci* and *Turkoglu,* 2009).

PCA can discover and reduce the dimensionality of data by their clustering (*Wang* and *Wang*, 2015), and it can also identify and observe the source of variation (*UI-Saufie et al*., 2013).

2.3. *Support vector regression (SVR)*

Relying on statistical learning or Vapnik-Chervonenkis theory, support vector machines (SVMs) have found applications in yet-to-be-seen data. The generalization of the classification problem can be regarded as a regression problem. In this case, the output of the model is a continuous value. Therefore, the performance of the regression model relies on continuous-valued multivariate function estimation. The classification problems can be solved by the convex optimization problems using SVMs (*Vapnik*, 1998). The optimization problem tries to find the maximum margin separation of the hyperplane. In SVM, the optimal hyperplane can be represented by support vectors. Generalization of SVM to SVR can be achieved by introducing the ε-insensitive region around the function also known as the ε-tube. The construction of SVR with the ε-insensitive loss function was proposed by *Vapnik* (1998).

The form of a linear function *f(x)* can be written as

$$f(x) = \langle \omega, x \rangle + b, \tag{2}$$

where *b* is the bias.

The problem could be considered a convex optimization problem. The optimization structure encompasses a regularization parameter, which affects the tradeoff between the approximation error and the weight vector norm. The optimization problem could be changed into a dual problem using Lagrange multipliers with a kernel function. Kernel functions have diverse types including linear, polynomial, Gaussian radial basis function (RBF), and sigmoid as presented (*Acosta et al*., 2022) in the following:

$$k(x_n, x) = x_n^T x \ , \tag{3}$$

$$k(x_n, x) = (\gamma x_n^T x + u)^d \ , \tag{4}$$

$$k(x_n, x) = exp(-\frac{1}{2r^2}\|x_n - x\|^2 = exp(-\gamma\|x_n - x\|^2) \ , \tag{5}$$

$$k(x_n, x) = tanh(\gamma x_n^T x + u), \tag{6}$$

where d is the degree of the polynomial and r defines the width of kernel, $\gamma = \frac{1}{2} r^2 andr > 0$.

2.4. Group method of data handling (GMDH)

The structure of GMDH consists of neurons that could be linked by the quadratic polynomial, giving rise to new neurons in the next layer. This model is aimed to minimize the squared of the differences between the forecasted and observed data:

$$\sum_{i=1}^{M}[\hat{f}(x_{i1}, x_{i2}, \ldots, x_{in}) - y_i]^2 \rightarrow min \quad (7)$$

The complex discrete form of the Volterra functional series can be used to state all connections between the input and output data defined as the Kolmogorov-Gabor polynomial:

$$y = a_0 + \sum_1^n a_i x_i + \sum_1^n \sum_1^n a_{ij} x_i x_j + \sum_1^n \sum_1^n \sum_1^n a_{ijk} x_i x_j x_k + \quad (8)$$

The coefficient of the polynomial can be found with the regression method which minimizes the difference between observed and estimated values (*Razzaghi et al.,* 2021).

2.5. Forecasts combination

In the linear mode of the combination methods, the combination of forecasts can be calculated with the linear function of the contributing individual forecasts from individual models. The importance of individual models can be determined by assigning nonnegative and unbiased weights:

$$\hat{y}_k = w_1 \hat{y}_k^{(1)} + w_2 \hat{y}_k^{(2)} + \ldots + w_n \hat{y}_k^{(n)}\ , \quad (9)$$

where w is the weights of single models and n is the number of single models (*Adhikari* and *Agrawal,* 2014).

The error-based methods (*Armstrong,* 2001), the least square regression (*Frietas* and *Rodrigues,* 2006), and the differential weighting scheme (DWS) of *Newbold* and *Granger* (1974) are among the forecast combination methods the differential weighting method is selected from in this study. The minimization of the combined forecast error variance is one of the approaches in the determination of the weights of forecast combinations. *Newbold* and *Granger* (1974) proposed five differential weighting schemes. Two of them have exhibited excellent performance (*Winkler* and *Maridais,* 1983) as presented in the following formulas:

$$w_i = (\sum_{s=t-v}^{t-1}(e_s^{(i)})^2)^{-1} / \sum_{j=1}^{n}(\sum_{s=t-v}^{t-1}(e_s^{(j)})^2)^{-1}, DWS - I \quad (10)$$

$$w_{i,t} = \beta w_{i,t} + (1-\beta)[\sum_{s=t-v}^{t-1}(e_s^{(i)})^2)^{-1} / \sum_{j=1}^{n}(\sum_{s=t-v}^{t-1}(e_s^{(j)})^2)^{-1}], DWS - II, (11)$$

where n is the number of single models, t is the forecasted time period, $w_{i,t\text{-}1}$ is the weight of the ith model using the data of preceding period, v and β are constant parameters, where β is between 0 and 1, and e_t is the percentage forecast error (*Winkler* and *Makridakis*, 1983).

2.6. Evaluation metrics

Two classes of metrics were used to evaluate the wind speed forecasting performance of the novel method (PCA-DWS). Metrics which investigate the accuracy of forecasts are listed in *Table 1*, where O_i and F_i (i=1,…,N) are the observed and forecasted time series, respectively, while O_{max} and O_{min} are the maximum and minimum values of the observed time series. The following metric explores the stability of forecasting:

$$DIS = \sqrt{\frac{\sum_{i=1}^{N}(e_i - (\frac{\sum_{i=1}^{N} e_i}{N}))^2}{N}}, \tag{12}$$

where e_i is the difference between observed and forecasted data.

Table 1. The evaluation metrics for investigation the accuracy of forecast.

AMAPE	Adopted mean absolute percent error	$AMAPE = \frac{1}{N}\sum_{i=1}^{N}(\frac{\lvert F_i - O_i \rvert}{\frac{1}{N}\sum_{i=1}^{N} O_i}) \times 100\ \%$
RMSE	Root mean square error	$RMSE = \sqrt{\frac{1}{N}\sum_{i-1}^{N}(F_i - O_i)^2}$
RRMSE	Relative root mean square error	$RRMSE = \frac{RMSE}{\overline{O}}$
NRMSE	Normalized root mean square error	$NRMSE = \frac{RMSE}{O_{\max} - O_{\min}}$
MAE	Mean absolute error	$MAE = \frac{1}{N}\sum_{i-1}^{N}\lvert O_i - F_i \rvert$
AE	Average error	$AE = \frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)$
TIC	Thiel inequity coefficient	$TIC = \frac{\sqrt{\frac{1}{N}\sum_{i-1}^{N}(F_i - O_i)^2}}{\sqrt{\frac{1}{N}\sum_{i-1}^{N}(O_i)^2} + \sqrt{\frac{1}{N}\sum_{i-1}^{N}(F_i)^2}}$
VAF	Values of account for	$VAF = \left[1 - \frac{\mathrm{var}(O_i - F_i)}{\mathrm{var}(O)}\right]$
NSE	Nash –Sutcliffe coefficient	$NSE = 1 - \frac{\sum_{i=1}^{N}(F_i - O_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}$

Better forecasting performance can be obtained with lower values of RMSE, DIS, AMAPE, MAE, TIC, RRMSE, and NRMSE. The values of NSE close to 1 indicate the perfect fit. The higher values of VAF are indicative of forecast improvement (*Wang et al.,* 2021a*; Luo et al*., 2021*; Temeng et al.,* 2020; *Dhiman et al.,* 2019; *Yang et al.,* 2020; *Minaeian* and *Ahangari,* 2013).

In addition to the mentioned criteria, some improvement percentage indicators such as PRMSE were employed to compare the performance of different models. PRMS is defined in as

$$PRMSE = \left|\frac{RMSE_1 - RMSE_2}{RMSE_2}\right| \times 100\% \,, \tag{13}$$

where $RMSE_1$is the RMSE of the first model, $RMSE_2$is the RMSE of the second model (*Liu et al.,* 2021).

3. Results

Minimum, maximum, and mean temperatures, relative humidity, precipitation, average visibility, and dew point temperature were regarded as independent variables for monthly wind forecasting through linear regression, GMDH, and SVR models. Evaluation of the assimilation of forecast combination scheme and PCA was conducted from 2019 to 2020. The correlation coefficients of wind speed with the mentioned variables are listed in *Table 2*. According to *Table 2*, the maximum correlation coefficient of meteorological variables and wind speed in Zanjan and Ahvaz was that of mean temperature and relative humidity. The significant correlation coefficients in the Ahvaz station are mean, maximum, and minimum temperatures, relative humidity, and precipitation. In Zanjan station, these significant correlation coefficients include mean, maximum, and minimum temperatures, relative humidity, and dew point.

Table 2. Correlation coefficients of meteorological variables and wind speed.

meteorological variables	**stations**	
	Zanjan	**Ahvaz**
Mean temperature	0.427**	0.623**
Maximum temperature	0.398**	0.597**
Minimum temperature	0.425**	0.619**
Relative humidity	-0.459**	-0.176*
Precipitation	-0.112	-0.386**
Average visibility	0.034	-0.031
Dew point temperature	0.338**	0.017

*. Correlation is significant at the 0.05 level (2-tailed)
**. Correlation is significant at the 0.01 level (2-tailed)

Therefore, temperature and relative humidity can be regarded as effective variables on wind speed. The average correlation coefficient of temperature (maximum, minimum, and mean) with wind speed in Ahvaz is higher than in Zanjan. The first step of the developed method is to use linear regression, ε-SVR, and GMDH (regarded as single models) to model the relationship between meteorological variables and wind speed. The parameters of SVR and GMDH are listed in *Table 3*.

Table 3. Parameters of single models for wind forecasting

Model	**Name of parameter**	**station**	
		Zanjan	**Ahvaz**
GMDH	Maximum number of neurons in a layer	9	5
	Maximum number of layers	2	2
	Selection pressure	0.9	0.9
SVR	Kernel function	Linear function	Gaussian radial basis function
	Regularization parameter, C	0.25	0.25

The sensitivity analysis is one of the most important steps in the modeling process. The MAE decreasing of SVR from sigmoid to linear kernel function (C=0.25) was 55.83% in Zanjan. The variation of maximum number of neurons per layer from 5 to 9 decreased MAE by 35.09% in Ahvaz. After sensitivity analysis of models, the wind forecasting performance of three models was investigated with some evolution metrics as shown in *Fig. 3*.

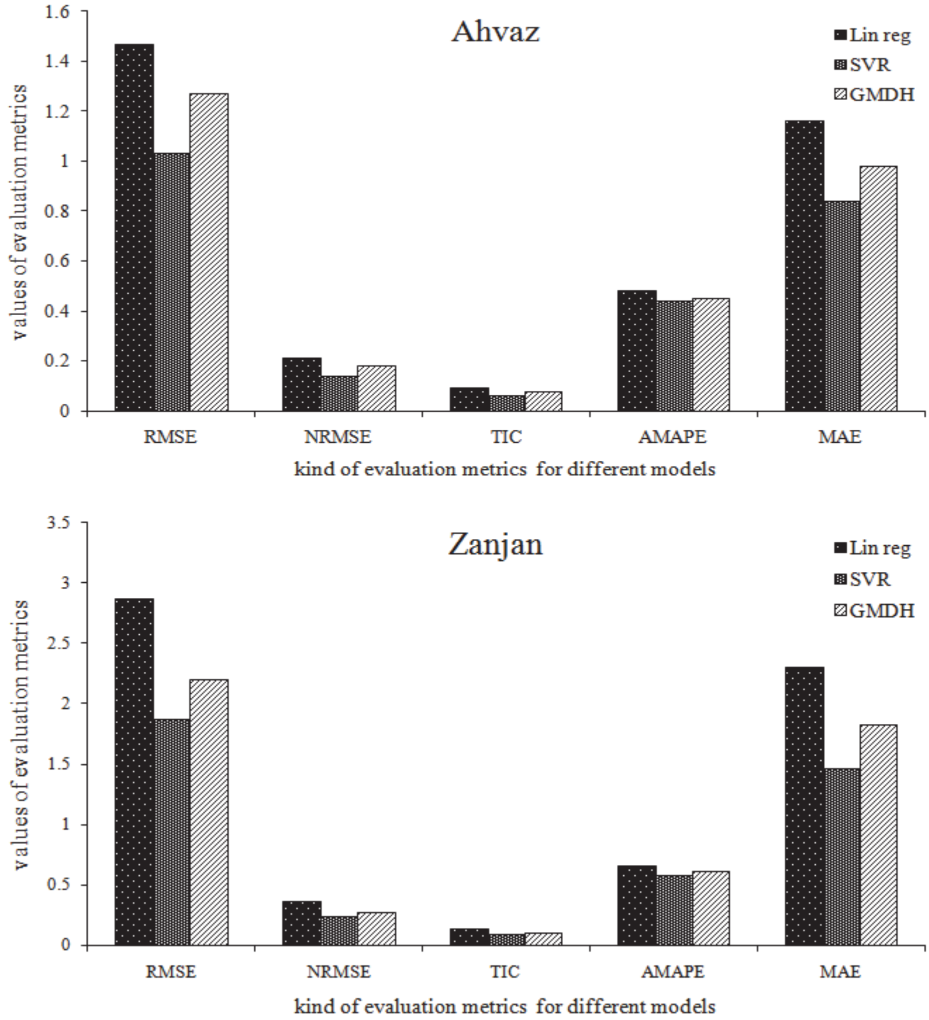

Fig. 3. Evaluation of the performance of three models, linear regression (Lin reg), SVR, and GMDH regarding to wind forecasting with some evaluation metrics in the two stations.

Evaluation metrics in *Fig. 3* are improved upon using SVR compared to the other models. In Ahvaz station, the RMSE, NRMSE, TIC, AMAPE, and MAE decrease from linear regression to SVR was 29.38%, 18.26%, 33.33%, 22.22%, and 30%, respectively. The amount of decline in the same parameters from GMDH to SVR was 18.18%, 8.33%, 2.22%, 27.58%, and 14.28%, respectively. The decrement in RMSE, NRMSE, TIC, AMAPE, and MAE from GMDH to SVR in Zanjan was 15%, 14.81%, 11.57%, 4.91%, and 20.21%, respectively. The average decrease in RMSE, NRMSE, TIC, AMAPE, and MAE in Ahvaz and Zanjan was 15.032% and 13.3%, respectively. The TIC decrement (average for two stations) from linear regression and GMDH to SVR was 32.69% and 14.87%, respectively, reflecting the better performance of SVR and GMDH relative to linear regression. The AE values of the two stations related to SVR were positive, suggesting forecast overestimation. The obtained forecasts from three models were combined with two methods (DWS-I and DWSII), and the combined forecasts methods (DWS-I) were compared with the best model among the three models as illustrated in *Fig. 4*.

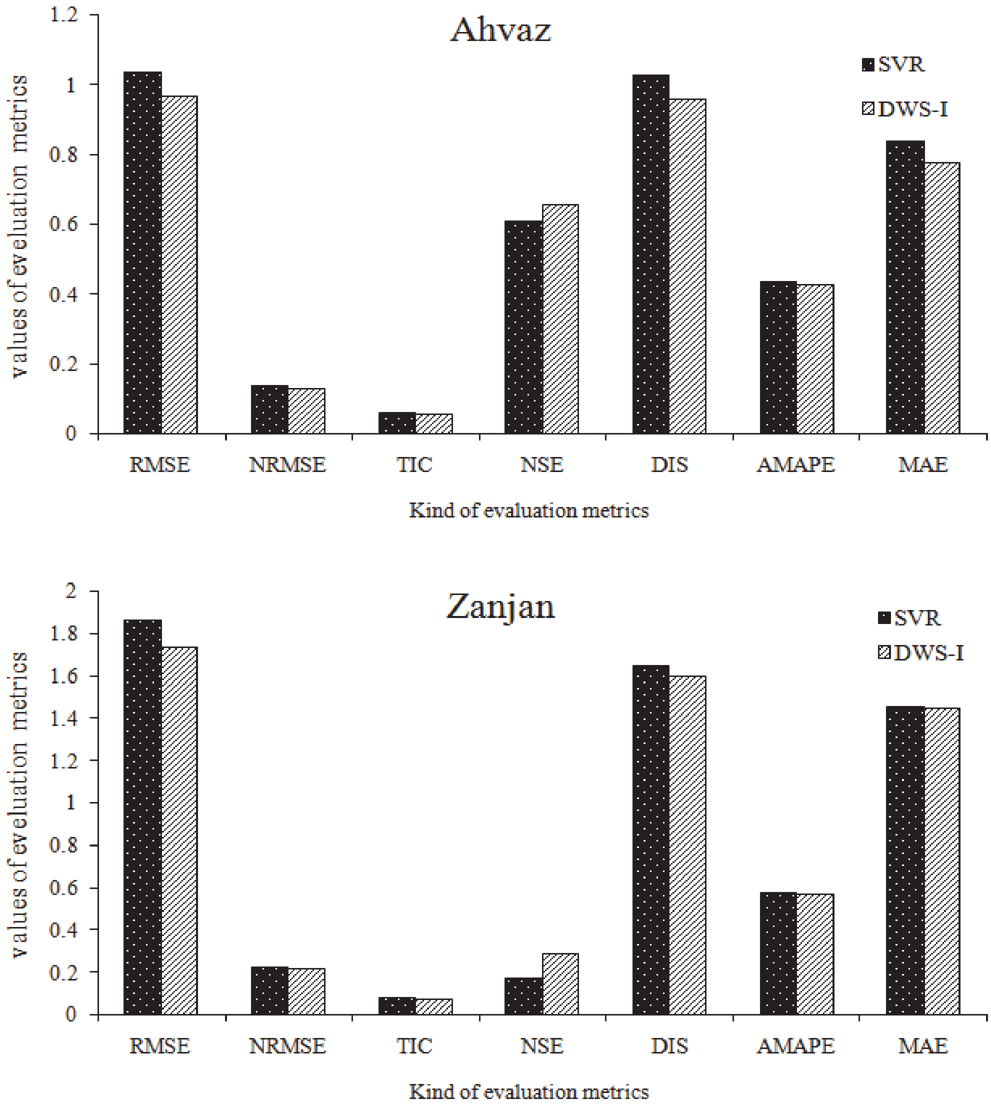

Fig. 4. Evaluation metrics related to SVR and DWS-I performances in the two stations.

DWS method outperformed the SVR, GMDH, and linear regression models, based on *Fig. 4*. The decrease in RMSE, NRMSE, and TIC from SVR to DWS was 6.45%, 7.14%, and 6.34% in Ahvaz and 6.95%, 4.34%, and 5.95% in Zanjan, respectively. The RMSE, NRMSE, TIC, and DIS decrement (average of two stations) from SVR to DWS was 6.7%, 5.74%, 6.14%, and 4.91%, respectively. The average reduction in RMSE, NRMSE, TIC, DIS, AMAPE, and MAE from SVR to DWSI-I for Ahvaz and Zanjan was 6.02% and 3.78%, respectively. Moreover, the values of NSE showed an increase in the two stations. NSE decrease was 8.19% in Ahvaz. AE of DWS-I in Ahvaz and Zanjan was negative and positive, respectively, reflecting the underestimated and overestimated forecasts. DWS-II exhibited better performance relative to the single models, but the difference in the evaluation metrics of DWS-II and DWS-I was low, especially in Ahvaz station. For example, in Ahvaz, the RMSE of DWS-I and DWS-II was 0.971 and 0.973, while its NRMSE was 0.059 and 0.06, respectively. Generally, DWS-I led to more acceptable results. The second combination of the developed method involved the combination of independent variables using PCA. The coefficients of each variable in the first component are displayed in *Fig. 5*.

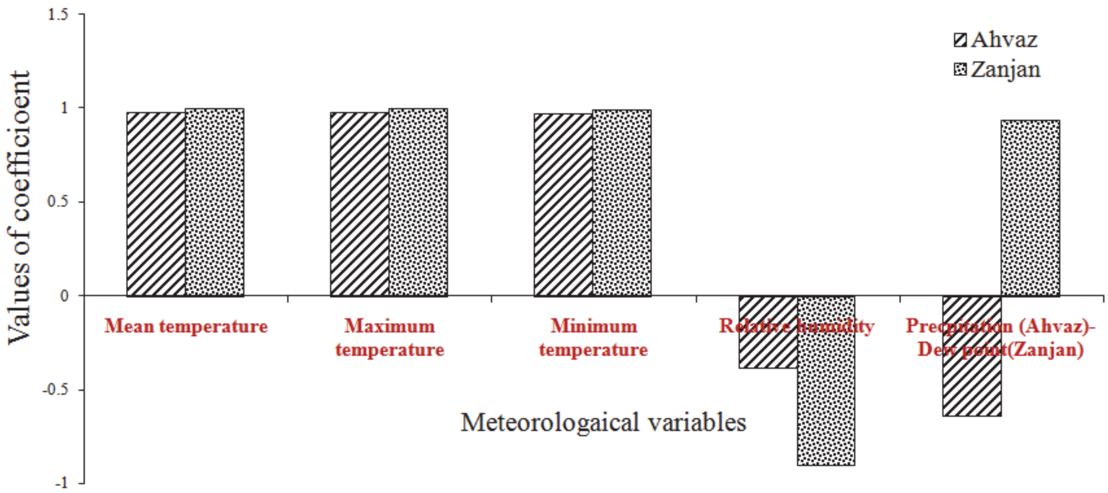

Fig. 5. The coefficients of meteorological variables in the first component of the two stations.

PCA transforms some independent variables into different components by multiplying the coefficient of each variable to its corresponding variable. It must be said thatas a result of this study, DWS implies DWS-I. According to *Fig. 5*, the maximum coefficient of variables in the first component was related to temperature. It also well matches with the variables by high correlation coefficient of *Table 2*. Finally, the incorporation of independent and output variables was conducted using PCA and DWS with SVR. The comparison of DWS with the novel method is represented in *Fig. 6*.

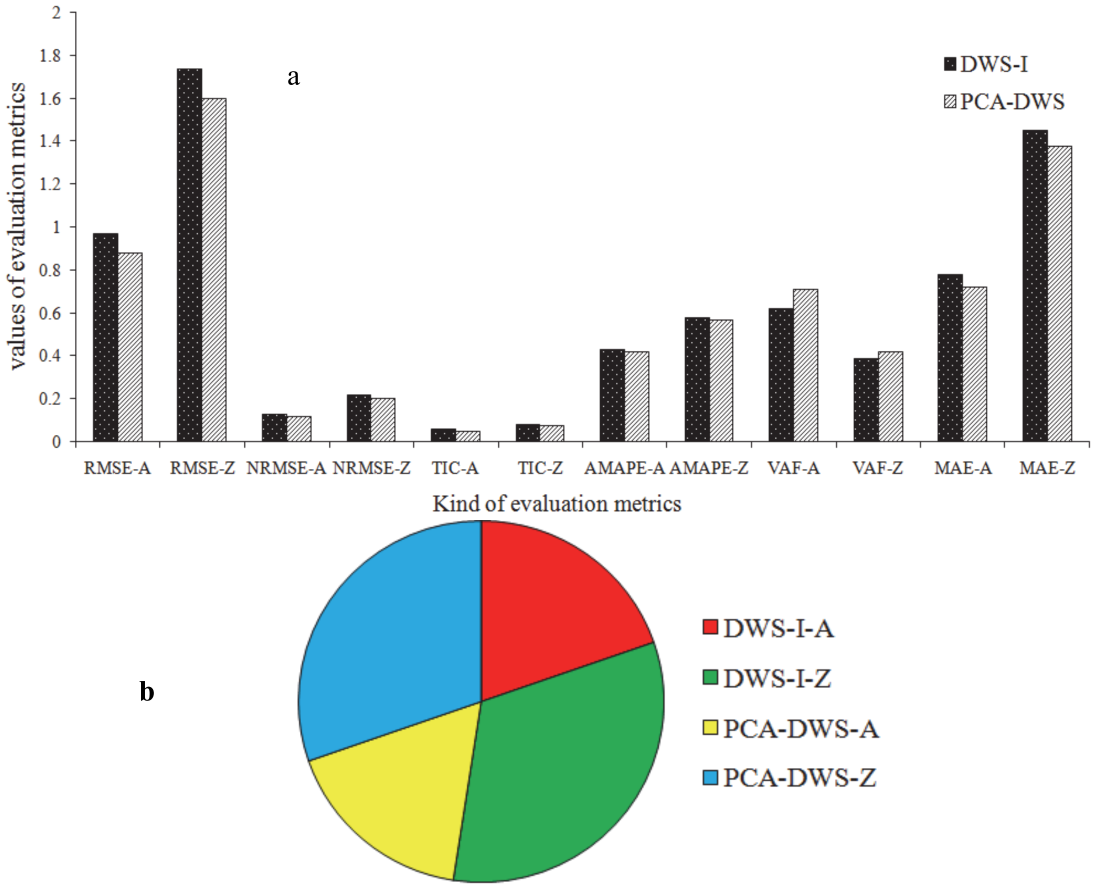

Fig. 6. Comparison of combined and proposed methods with some evolution metrics (a), investigation of the forecast stability of two methods with DIS (b). A and Z indicate Ahvaz and Zanjan.

Evaluation metrics of the proposed method in *Fig. 6a* suggests a better situation for the proposed method than for DWS. The RMSE, NRMSE, TIC, AMAPE, and MAE decrease from DWS to PCA-DWS was 9.27%, 7.69%, 15.25%, 2.32%, and 7.69% for Ahvaz and 8.04%, 9.09%, 7.59%, 1.21%, and 4.82% for Zanjan, respectively. The average decline of RMSE, NRMSE, TIC, AMAPE, and MAE was 8.44% and 6.15% for Ahvaz and Zanjan, respectively, showing a greater decline in calculated average values for Ahvaz than for Zanjan. The rise in VAF from DWS to the proposed method was 14.51% and 7.69% for Ahvaz and Zanjan, respectively. The value of VAF in Ahvaz was higher than in Zanjan. Also, the NRMSE of Ahvaz was lower than of Zajnan. The share of the pie diagram in *Fig. 6b* was decreased by the incorporation of independent and output variables. It indicates the more preservation of forecasts stability in the proposed method. DIS decrease from DWS to the proposed method in Ahvaz and Zanjan was 11.45% and 7.5%, respectively. Positive AE values were seen in the two stations using PCA-DWS, indicating the overestimation of the forecast. The variation of wind speed in the verification period is shown in *Fig. 7*.

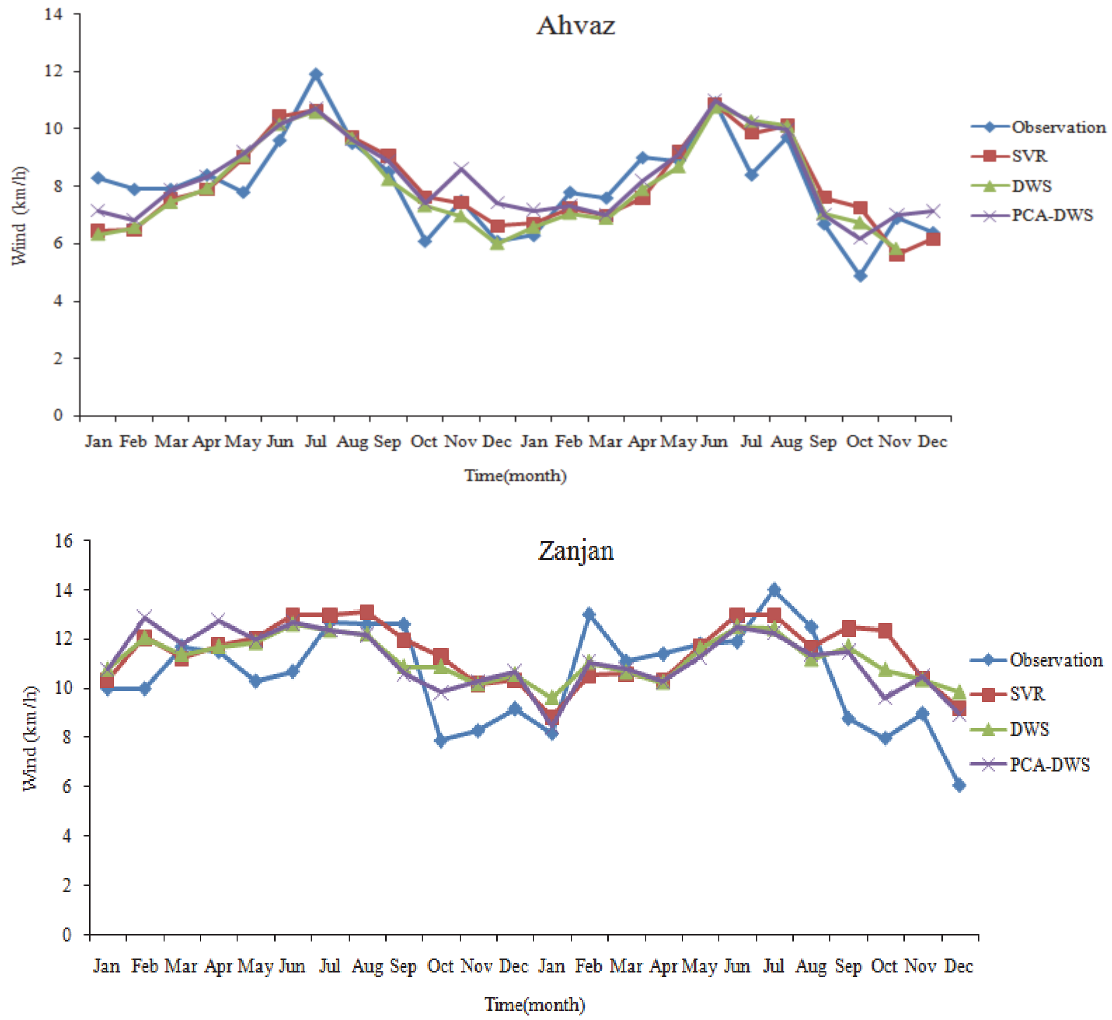

Fig. 7. Observed and forecasted monthly wind speeds with different models in the verification period, 2019-2020, for the two stations.

The average maximum wind speed during the two years of the verification period, 2019 and 2020, occurred in June in the observation data; this time was preserved with SVR, DWS, and PCA-DWS in Ahvaz. The minimum observed wind speed was in October, which was only preserved in the PCA-DWS method. The fitted R-square of the lines presents the maximum value for the PCA-DWS method. The increment in the R-square of the fitted lines from SVR and DWS to PCA-DWS method was 18.4% and 9.45% for Ahvaz and 44.8% and 16.6% for Zanjan, respectively. To compare the performance of the studied models (with regard to linear regression), PRMSE, PNRMSE, PTIC, and PNSE were calculated in *Table 4*.

Table 4. Some improvement percentage indicators for comparison of PCA-DWS with other models.

Station	**Indicators**	**SVR**	**GMDH**	**DWS-I**	**PCA-DWS**
Ahvaz	PRMSE	0.41	0.15	0.51	0.67
	PNRMSE	0.5	0.16	0.52	0.66
	PTIC	0.42	0.17	0.52	0.8
Zanjan	PRMSE	0.53	0.3	0.64	0.79
	PNRMSE	0.56	0.33	0.63	0.8
	PTIC	0.54	0.36	0.64	0.78

The maximum and minimum values of improvement indicators were related to PCA-DWS and GMDH. The maximum value of indicators was for PTC and PNRMSE, while the minimum values of the indicator were seen for PRMSE. In the following, the performance of the proposed method for monthly wind speed forecasting was investigated one by one; the RRMSE values are presented in *Fig. 8*.

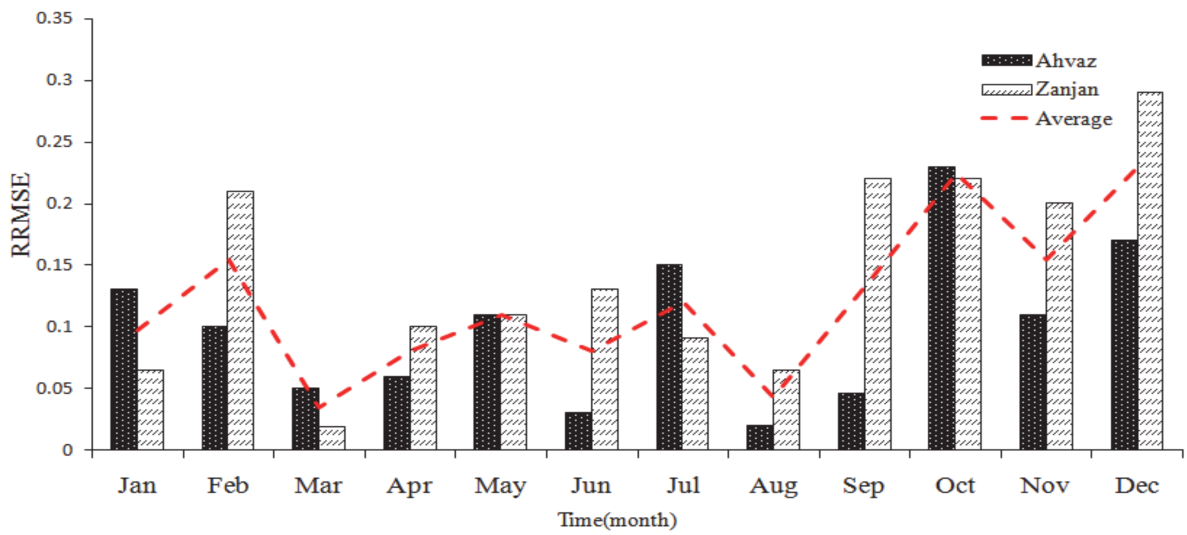

Fig. 8. Performance of PCA-DWS in each month of the verification period for the two stations with RRMSE.

The minimum value of RRMSE in *Fig. 8* for Ahvaz, Zanjan, and their average was in August, March, and March, respectively. The maximum value of RRMSE was in October and December. It can be said that the maximum value of RRMSE occurred in autumn. The maximum value of RRMSE in Ahvaz was lower than in Zanjan, whereas a similar minimum value of RRMSE was detected in the two stations. Adherence of average series from Zanjan series can be assigned to the high error in wind forecasting of Zanjan relative to Ahvaz. To investigate the performance of the novel method, another comparison was made related to the annual scale whose results are depicted in *Fig.9*.

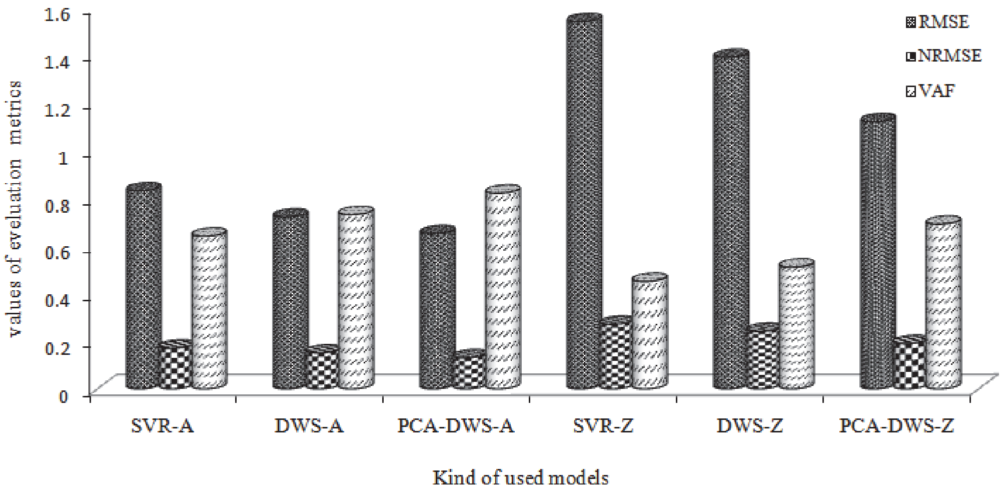

Fig. 9. Some evaluation metrics for PCA-DSW performance evaluation in annual scale. A and Z indicate Ahvaz and Zanjan.

In *Fig. 9* PCA-DWS led to better forecasts relative to SVR and DWS. The RMSE decrease from SVR to DWS and DWS to PCA-DWS was 13.25% and 9.77% in Ahvaz and 9.74% and 19.42% in Zanjan, respectively. The NRMSE decline from SVR to DWS and DWS to PCA-DWS was respectively 11.76% and 13.33% in Ahvaz and 11.11% and 20.88% in Zanjan. The increment in VAF from SVR to DWS and DWS to PCA-DWS was 14.06% and 12.32% in Ahvaz and 13.33% and 35.29% in Zanjan, respectively. In Ahvaz, the VAF of PCA-DWS was lower than that of in Zanjan. VAF of each station in the annual series was higher than in the monthly series. The observed annual wind speed in the two stations in 2019 was greater than in 2020, which was preserved in the PCA-DWS method.

4. Discussion

One approach for meteorological data forecasting is to model meteorological variables with the best correlation with each other. *Afkhami et al*. (2015); *Kumar* and *Malik* (2016), and *Khosravi et al*. (2018) employed meteorological data such as temperature to forecast wind speed. In this study, temperature and relative humidity showed the maximum significant correlation with wind speed. The effect of temperature on wind speed was also reported by *Khakzad et al.* (2017) and *Wang et al*. (2021b). Linear regression, the GMDH, and SVR were used to model the relation between correlated meteorological variables and wind speed. In each station, SVR outperformed GMDH and linear regression methods. The superiority of SVM to GMDH was reported in some studies such as the work of *Khosravi et al.* (2018) for wind speed and direction forecasting, *Raza et al*. (2020) for evapotranspiration estimation in four climatic regions, and *Yaghoubi et al.* (2019) for monthly forecasting of streamflow. Comparison of the performance of three models indicated that the NRMSE of SVR in Zanjan was greater than in Ahvaz, while the VAF of SVR in Ahvaz was larger than in Zanjan. Therefore, for increasing the accuracy of forecasts, the mentioned single models were combined using the DWS method. The forecasting performance of DWS was better than the forecasts of three individual models. The rise in VAF from SVR to DWS for Zanjan and Ahvaz was 34.48% and 10.71%, respectively. The combinational forecast is a function involving the sum of weight-assigned single forecast models. The contribution of the single models on the final forecast was determined considering their weights. Therefore, a proper function in the forecast combination process can derive the information of the single models to improve the accuracy of the results. The forecast combination methods reduced the forecasted errors and led to high accuracy (*Adhikari* and *Agrawal*, 2014; *Wang et al*., 2020; *Dewangan et al*., 2020). The evaluation metrics of DWS-I and DWS-II have shown low differences in many cases. For example, the NSE of Ahvaz for DWS-I and DWS-II was 0.66 and 0.662, respectively, whereas the NRMSE of Zanjan for DWS-I and DWS-II was 0.22 and 0.23, respectively. In general, however, DWS-I outperformed DWS-II. One of the reasons explaining the poor performance of DWS-II compared to DWS-I might be the presence of ν and β. *Winkler* and *Makridakis* (1983) recommended $\nu=12$ and $\beta=0.7$. In this study, $\nu=12$ and β value was manually selected in the range of $0<\beta<1$. The values of parameters can indeed affect the forecasting performance of the combinational methods. Restriction of the forecasting to the recent observation can be due to smaller values of ν. Smaller values of β guarantee assigning more weights to recent observations (*Adhikari* and *Agrawal*, 2014). Therefore, differences in weight allocation can affect the forecasts. Regarding the success of the DWS method for wind forecasting, a novel method was proposed to increase the accuracy of DWS by using PCA on independent variables. The accordance of PCA-DWS forecasts with observed wind speed was higher than that of DWS and

SVR. The rise in NSE from DWS to the novel method in Ahvaz and Zanjan stations was 9.09% and 37.39%, respectively. The use of PCA-DWS in Ahvaz led to lower NRMSE values compared to Zanjan. The values of VAF in Ahvaz was higher than in Zanjan. The excellent performance of DWS-PCA relative to other mentioned methods was verified with the R-square of the fitted lines and some improvement percentage indicators. In addition to the monthly time scale, the PCA-DWS exhibited a proper performance on the annual scale. The VAF of the annual scale in the two stations was higher than the VAF of the monthly scale.

5. Conclusions

Wind forecasting with an accurate model can be useful in renewable energy studies, climate sciences, and hydrology studies. To increase the accuracy of wind speed forecasting, the forecast performance of three single models (linear regression, SVR, and GMDH) were combined using the DWS method. DWS outperformed the three models. Another aspect in improving the accuracy of DWS forecasts is the incorporation of DWS and PCA that exhibited good performance on annual and monthly scales. Performance comparison of the two stations indicated the better performance of Ahvaz than the performance of Zanjan, which can be due to climate effects. Some issues can affect the wind forecasting performance of PCA-DWS: 1) Using power model as the single model and its accurate sensitivity analysis. One approach to improve the performance of single models is to determine the model parameters through optimization methods such as genetic algorithm (GA) or PSO. 2) Type of forecast combination method to find the exact weights of the single models for deriving their information. 3) Using the rotation option related to the components in PCA to combine independent variables. 4) Using an efficient model for the incorporation of PCA and DWS. The novel developed method in this study offered proper effectiveness, forecasting accuracy, and stability, further encouraging its application in monthly and annual wind speed forecasts.

References

Acosta, S.M., Amoroso, A.L., Sant'Anna, Â.M.O., and *Junior, O.C.,* 2022: Predictive modeling in a steelmaking process using optimized relevance vector regression and support vector regression. *Ann. Operat. Res.* 316(2), 905–926. https://doi.org/10.1007/s10479-021-04053-9

Adhikari, R., and *Agrawal, R.K.,* 2014: Performance evaluation of weights selection schemes for linear combination of multiple forecasts. *Artific. Intellig. Rev. 42(4)*: 529–548. https://doi.org/10.1007/s10462-012-9361-z

Afkhami, H., Talebi, A., Mohammadi, M., and *Fotouhi, F.,* 2015: Investigation of the feasibility of wind speed prediction using hybrid model of neural networks, neural-fuzzy networks and wavelet (Case Study: Station of Yazd), *Iranian J. Watershed Manag. Sci. Engin. 9(30):* 31–40.

Armstrong, J.S., 2001: Principles of forecasting: a handbook for researchers and practitioners. .Kluwer Academic Publishers, Boston, *USA.* https://doi.org/10.1007/978-0-306-47630-3

Avci, E., and *Turkoglu, I.,* 2009: An intelligent diagnosis system based on principal component analysis and ANFIS for the heart valve diseases. *Exp. Syst. Appl. 36*(2) 2873–2878. https://doi.org/10.1016/j.eswa.2008.01.030

De Martonne, E., 1925: Traité Géographie. Physique: 3 tomes. Max leclcrc and H. Bourrclier, proprietors of Librairic Armard Colin: Paris. (In France)

Dewangan, C.L., Singh, S.N., and *Chakrabarti, S.,* 2020: Combining forecasts of day-ahead solar power. *Energy 202*, 117743. https://doi.org/10.1016/j.energy.2020.117743

Dhiman, H.S., Deb, D., and *Guerrero, J.M.,* 2019: Hybrid machine intelligent SVR variants for wind forecasting and ramp events. *Renew. Sustain. Energy Rev. 108*, 369–379. https://doi.org/10.1016/j.rser.2019.04.002

Emberger, C., Gaussen, H., Kassas, M., and *de Philippis, A.,* 1963: Bioclimatic map of the Mediterranean Zone, explanatory notes. UNESCO-FAO. Paris.

Fazelpour, F., Tarashkar, N., and *Rosen, M.A.,* 2016: Short-term wind speed forecasting using artificial neural networks for Tehran, Iran. *Int. J. Energy Environ. Engin. 7,* 377–390. https://doi.org/10.1007/s40095-016-0220-6

Frietas, P.S.A., and *Rodrigues A.J.L.,* 2006: Model combination in neural-based forecasting. *Eur. J. Oper. Res. 173,* 801–814. https://doi.org/10.1016/j.ejor.2005.06.057

Jamil, M., and *Zeeshan, M.,* 2019: A comparative analysis of ANN and chaotic approach-based wind speed prediction in India. *Neur. Comput. Appl. 31,* 6807–6819. https://doi.org/10.1007/s00521-018-3513-2

Jaseena, K.U., and *Kovoor, B.C.,* 2021: Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks. *Energy Conv. Manage. 234*, 113944. https://doi.org/10.1016/j.enconman.2021.113944

Khakzad, S.F., Golian, S., and *Elmi, M.,* 2017: Predicting wind speed and direction using the output of numerical meteorological models. MSc Thesis in Shahrood University of Technology.

Khosravi, A., Koury, R.N., Machado, L., and *Pabon, J.J.,* 2018: Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system. *Sust. Energy Technol. Assess. 25*, 146–160. https://doi.org/10.1016/j.seta.2018.01.001

Kumar, G., and *Malik, H.,* 2016: Generalized regression neural network based wind speed prediction model for western region of India. *Procedia Comput. Sci. 93*, 26–32. https://doi.org/10.1016/j.procs.2016.07.177

Liu, Z., Hara, R. and *Kita, H.,* 2021: Hybrid forecasting system based on data area division and deep learning neural network for short-term wind speed forecasting. *Energy Conver. Manage. 238*, 114136. https://doi.org/10.1016/j.enconman.2021.114136

Luo, L., Li, H., Wang, J., and *Hu, J.,* 2021: Design of a combined wind speed forecasting system based on decomposition-ensemble and multi-objective optimization approach. *Appl. Math. Model. 89*, 49–72. https://doi.org/10.1016/j.apm.2020.07.019

Minaeian, B. and *Ahangari, K.,* 2013: Estimation of uniaxial compressive strength based on P-wave and Schmidt hammer rebound using statistical method. *Arabian J. Geosci. 6,* 1925–1931. https://doi.org/10.1007/s12517-011-0460-y

Newbold, P. and *Granger, C.W.J.,* 1974: Experience with forecasting univariate time series and the combination of forecasts. *J. Roy.Stat. Soc.: General, 137(2)*: 131–146. https://doi.org/10.2307/2344546

Pan, L., Liu, Y., Roux, G., Cheng, W., Liu, Y., Hu, J., Jin, S., Feng, S., Du, J., and *Peng, L.,* 2021: Seasonal variation of the surface wind forecast performance of the high-resolution WRF-RTFDDA system over China. *Atmos. Res. 259*, 105673. https://doi.org/10.1016/j.atmosres.2021.105673

Raza, A., Shoaib, M., Faiz, M.A., Baig, F., Khan, M.M., Ullah, M.K., and *Zubair, M.,* 2020: Comparative assessment of reference evapotranspiration estimation using conventional method and machine learning algorithms in four climatic regions. *Pure Appl. Geophys. 177,* 4479–508. https://doi.org/10.1007/s00024-020-02473-5

Razzaghi, H., Madandoust, R., and *Aghabarati, H.,* 2021: Point-load test and UPV for compressive strength prediction of recycled coarse aggregate concrete via generalized GMDH-class neural network. *Constr. Build. Mater. 276*, 122143. https://doi.org/10.1016/j.conbuildmat.2020.122143

Song, C. and *Fu, X.,* 2020: Research on different weight combination in air quality forecasting models. *J. Clean. Product. 261*, 121169. https://doi.org/10.1016/j.jclepro.2020.121169
Temeng, V.A., Ziggah, Y.Y., and *Arthur, C.K.,* 2020: A novel artificial intelligent model for predicting air overpressure using brain inspired emotional neural network. *Int. J. Mining Sci. Technol. 30*(5), 683-9. https://doi.org/10.1016/j.ijmst.2020.05.020
Ul-Saufie, A.Z., Yahaya, A.S., Ramli, N.A., Rosaida, N., and *Hamid, H.A.,* 2013: Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principal component analysis (PCA). *Atmos. Environ. 77*, 621–630. https://doi.org/10.1016/j.atmosenv.2013.05.017
Ul-Saufie, A.Z., Yahya, A.S., and *Ramli, N.A.,* 2011: Improving multiple linear regression model using principal component analysis for predicting PM10 concentration in Seberang Prai, Pulau Pinang. *Int. J. Environ. Sci. 2,* 403–410.
Vapnik, V.N., 1998: Statistical learning theory. John Wiley & Sons.
Wang, J., Li, Q., and *Zeng, B.,* 2021a: Multi-layer cooperative combined forecasting system for short-term wind speed forecasting. *Sust. Ener. Technol. Assess. 43*, 100946. https://doi.org/10.1016/j.seta.2020.100946
Wang, J., and *Wang, J.,* 2015: Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks. *Neurocomputing 156*, 68–78. https://doi.org/10.1016/j.neucom.2014.12.084
Wang, J., Zhou, H., Hong, T., Li, X., and *Wang, S.,* 2020: A multi-granularity heterogeneous combination approach to crude oil price forecasting. *Energy Econ. 91*, 104790. https://doi.org/10.1016/j.eneco.2020.104790
Wang, Z.W., Zhang, W.M., Zhang, Y.F., and *Liu, Z.,* 2021b: Circular-linear-linear probabilistic model based on vine copulas: An application to the joint distribution of wind direction, wind speed, and air temperature. *J. Wind Engineer. Indust. Aerodynam. 215*, 104704. https://doi.org/10.1016/j.jweia.2021.104704
Winkler, R.L. and Makridakis, S., 1983: The combination of forecasts. J. Roy. Stat. Soc.: General 146(2), 150–157. https://doi.org/10.2307/2982011
Xie, A., Yang, H., Chen, J., Sheng, L., and *Zhang, Q.,* 2021: A Short-Term Wind Speed Forecasting Model Based on a Multi-Variable Long Short-Term Memory Network. *Atmosphere 12*(5), 651. https://doi.org/10.3390/atmos12050651
Yaghoubi, B., Hosseini, S.A., and *Nazif, S.,* 2019: Monthly prediction of streanflow using data-driven models. *J. Earth Syst. Sci. 128*(6), 1–15. https://doi.org/10.1007/s12040-019-1170-1
Yang, H., Zhu, Z., Li, C., and *Li, R.,* 2020: A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight. *Appl. Soft Comput. 87*, 105972. https://doi.org/10.1016/j.asoc.2019.105972