Taylor & Francis
Taylor & Francis Group

Engineering Applications of Computational Fluid Mechanics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tcfm20

Wind speed prediction over Malaysia using various machine learning models: potential renewable energy source

Marwah Sattar Hanoon, Ali Najah Ahmed, Pavitra Kumar, Arif Razzaq, Nur'atiah Zaini, Yuk Feng Huang, Mohsen Sherif, Ahmed Sefelnasr, Kwok wing Chau & Ahmed El-Shafie

To cite this article: Marwah Sattar Hanoon, Ali Najah Ahmed, Pavitra Kumar, Arif Razzaq, Nur'atiah Zaini, Yuk Feng Huang, Mohsen Sherif, Ahmed Sefelnasr, Kwok wing Chau & Ahmed El-Shafie (2022) Wind speed prediction over Malaysia using various machine learning models: potential renewable energy source, Engineering Applications of Computational Fluid Mechanics, 16:1, 1673-1689, DOI: 10.1080/19942060.2022.2103588

To link to this article: https://doi.org/10.1080/19942060.2022.2103588

Published online: 16 Aug 2022.

ENGINEERING APPLICATIONS OF COMPUTATIONAL FLUID MECHANICS
2022, VOL. 16, NO. 1, 1673–1689
https://doi.org/10.1080/19942060.2022.2103588

Wind speed prediction over Malaysia using various machine learning models: potential renewable energy source

Marwah Sattar Hanoon[a], Ali Najah Ahmed [b], Pavitra Kumar[c], Arif Razzaq[d], Nur'atiah Zaini[e], Yuk Feng Huang [f], Mohsen Sherif [g,h], Ahmed Sefelnasr [h], Kwok wing Chau [i] and Ahmed El-Shafie [j,h]

[a]College of Technical Engineering, Islamic University, Najaf, Iraq; [b]College of Engineering, Universiti Tenaga Nasional (UNITEN), Selangor, Malaysia; [c]Department of Geography and Planning, University of Liverpool, Liverpool, United Kingdom; [d]College of Science, Al Muthanna University, Al-Muthanna, Iraq; [e]Department of Civil Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), Kajang, Malaysia; [f]Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Selangor, Malaysia; [g]Civil and Environmental Engineering Department, College of Engineering, United Arab Emirates University, Al Ain, United Arab Emirates; [h]National Water and Energy Center, United Arab Emirates University, Al Ain, United Arab Emirates; [i]Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hung Hom, Hong Kong; [j]Department of Civil Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia

ABSTRACT

Modeling wind speed has a significant impact on wind energy systems and has attracted attention from numerous researchers. The prediction of wind speed is considered a challenging task because of its natural nonlinear and random characteristics. Therefore, machine learning models have gained popularity in this field. In this paper, three machine learning approaches – Gaussian process regression (GPR), bagged regression trees (BTs) and support vector regression (SVR) – were applied for prediction of the weekly wind speed (maximum, mean, minimum) of the target station using other stations, which were specified as reference stations. Daily wind speed data, gathered via the Malaysian Meteorological Department at 14 measuring stations in Malaysia covering the period between 2000 and 2019, were used. The results showed that the average weekly wind speed had superior performance to the maximum and minimum wind speed prediction. In general, the GPR model could effectively predict the weekly wind speed of the target station using the measured data of other stations. Errors found in this model were within acceptable limits. The findings of this model were compared with the measured data, and only Kota Kinabalu station showed an unacceptable range of prediction. To investigate the prediction performance of the proposed model, two models were used as the comparison models: the BTs model and SVR model. Although the comparison of GPR with the BTs model at Kuching station showed slightly better performance for the BTs model in maximum and minimum wind speed prediction, the prediction outcomes of the other 13 stations showed better performance for the proposed GPR model. Moreover, the proposed model generated smaller prediction errors than the SVR model at all stations.

ARTICLE HISTORY
Received 7 May 2021
Accepted 15 July 2022

KEYWORDS
Bagged regression trees; Gaussian process regression; support vector regression; machine learning; wind speed prediction

1. Introduction

Wind power is one of the most common renewable energy resources (Başakın et al., 2022), as the energy it generates is clean, and it also aims to minimize global warming and environmental pollution because it does not release toxic emissions (such as those produced by fossil fuel power stations, which cause many health problems for humans). In this respect, reliable wind speed prediction is critical for the future of renewable energy use (Barhmi & El Fatni, 2019).

The world's overall consumption of electrical energy is quickly increasing; however, the emission of greenhouse gases is rising owing to energy being produced from fossil fuels. Furthermore, the global electrical energy production percentage (2.7% average annually) increased from 2003 to 2015 and it will continue to rise until at least 2030 (Shafiullah et al., 2013). Nevertheless, about 40% of the world's greenhouse gas emissions are generated by electricity, which mainly uses fossil fuels, i.e. carbon and oil (Shafiullah, 2016). The greenhouse gas emissions are harmful to humans; however, renewable power sources, including wind, solar, biomass and rain, could reduce our dependence on fossil fuels. The demand for wind energy is increasing, to overcome the greenhouse impact and use the other energy resources efficiently. Wind is considered the most effective and technologically innovative

CONTACT Ali Najah Ahmed mahfoodh@uniten.edu.my

renewable energy resource available, owing to its free nature and availability (Shafiullah et al., 2013).

1.1. Motivation for the study

Nowadays, in Malaysia, renewable energy such as wind energy is considered the most attractive source of energy. Therefore, researchers need to prepare an inventory of the availability of wind energy in a region where there is a lack of measured wind speed datasets. This motivation would be helpful in predicting the wind speed potential utilizing various machine learning approaches. In this context, it is important to establish a reliable model for wind speeds at a particular station based on other stations. This would aid in the management of wind energy resources for electricity production, as well as in case of any issues arising related to the output of a station, where the output of other stations could be considered. For these reasons, in this article, three machine learning methods – support vector regression (SVR), bagged trees (BTs) and Gaussian process regression (GPR) – were employed to predict the wind speed at 14 stations in Malaysia.

1.2. Literature review

Several methods have been tested for forecasting wind speed. Machine learning is an excellent multidisciplinary area in which techniques for wind speed can be applied. Therefore, several studies have focused on the implementation of machine learning algorithms to forecast wind speed values (Khosravi et al., 2018). Barhmi and El Fatni (2019) used various hybrid models based on the support vector machine (SVM), and artificial neural network or neural network machines were used to predict hourly wind speed. Md Abul Ehsan et al. (2019) used multiple non-parametric tree-based machine learning approaches to predict the extreme wind speed. Moreover, to predict short–long-term wind speed, a hybrid model was applied in the short–long-term prediction period of 10 min (L. Wang et al., 2018). M. R. Chen et al. (2019) developed a new two-layer nonlinear combination approach termed EEL-ELM for short-term wind speed prediction problems, e.g. 10 min or 1 h in advance. The first layer is based on ELM, Elman neural network (ENN) and long–short-term memory (LSTM) to individually predict wind speed, and this model achieved good performance. Furthermore, the LSTM method was applied via a nonlinear training ensemble of deep learning prediction based on SVR, LSTM and extremal optimization (EO) algorithm to predict wind speed (J. Chen et al., 2018).

The GPR model is gaining popularity in multiple scientific applications. The GPR was presented to improved forecasting accuracy for near-surface wind speed prediction (Hoolohan et al., 2018). The hybrid model based on auto-regression and GPR was applied to achieve better probabilistic wind speed prediction; it was compared with other models and provided good performance (Zhang et al., 2016). The GPR, as an effective nonlinear modeling approach, could efficiently interpret the complicating features of industrial datasets via the combined covariance function derived from the base kernel. The GPR model is equivalent to the conventional soft sensors (Liu et al., 2018). The GPR model predicted total electron content (TEC) values more accurately (1 or 2 days) ahead of day in order to forecast the daily ionosphere TEC using the GPR model and multiple linear regression (Inyurt et al., 2020). The European Union chemical property maps were created using a GPR model. This model has been chosen because of its ability to measure model uncertainty and its ability to add prior information to modeling in the form of covariance functions (Ballabio et al., 2019). H. Wang et al. (2020) introduced a probabilistic method to predict wind gusts applying ensemble learning. This model included random forest, LSTM and GPR models, and the results indicated that this model had higher precision and efficiency of generalization.

In recent years, ensemble models, which combine many individual models into an ensemble modeling method, have presented an efficient replacement for the traditional machine learning methods (James et al., 2013). The potential of ensemble modeling methods to reduce the variance of the sample while achieving low bias make them very desirable to improve the accuracy of prediction (Breiman, 1996). The bagging ensembles of decision trees are implemented to efficiently predict wind energy (Breiman, 1996). A preference for this regression method is based on its ability to use multiple comparatively poor single trees to obtain a high degree of predictability compared to single regressors. In addition, it decreases the total error and has the potential to combine multiple models. Real measurements taken from a real wind turbine every 10 min were used to demonstrate the estimation accuracy of the tested methods. The researchers used regression tree approaches to accurately model load (Srivastava et al., 2020).

Another popular group is SVR methods, which have good generalization capability. Jiang et al. (2017) introduced the hybrid short-term model for predicting wind speed using the ν-SVM, optimized using the cuckoo search algorithm. The state-space SVM with Kalman unscented filter was developed by K. Chen and Yu (2014) to predict wind speed. Furthermore, a combined or hybrid model has been tested to boost the prediction efficiency of a single model (Liye Xiao et al., 2017; Ling

Xiao et al., 2015). For the combined approaches, the various single models are used in forecasting, and their projected outcomes are combined with the necessary weight coefficients for the final prediction. In a combined paradigm based on the principle of no negative limitation and artificial intelligence, J. Wang et al. (2017) proposed defining the optimum weight coefficients using an algorithm for maximizing chaos particle swarm. J. Wang et al. (2017) developed an integrated predictive model using multitarget bat algorithms for wind speed prediction to simultaneously attain high accuracy and consistency. J. Wang and Hu (2015) introduced the stable combination model of the autoregressive integrated moving average (ARIMA), SVM, ELM and least square support vector machine (LSSVM) for the short-term probabilistic estimation of wind speed, using GPR to integrate the outcomes of each predictor.

Citakoglu and Aydemir (2019) used the gray estimation technique for the monthly wind speed dataset measured for the years 2000–2017 at three weather stations in Kayseri. Three different gray model (GM) (1,1) equations were obtained from the results of their analyses. Their results showed that prediction of wind speed at locations without measurements can be precisely achieved with the GM(1,1) model obtained from other nearby stations.

1.3. Problem statement

Optimum forecasting of wind speeds is an ongoing research issue since it is necessary to predict the viability of harnessing wind speed from particular locations. Consequently, the technologies relating to wind speeds must be developed in depth. Because wind speed has characteristics of nonlinearity and randomness, the production of wind energy is unpredictable. The large-scale grid integration of wind energy to the electricity network can be challenging in terms of energy conversion and management (Xiwei Mi et al., 2019). To tackle this issue, very efficient technologies are needed to predict wind speeds. Wind speeds can be affected by many factors. Simple forecasting methods face a challenge to capture the sophisticated characteristics of wind speeds in order to achieve precise prediction results. Thus, high-accuracy wind speed forecasts have gained worldwide attention.

1.4. Objectives

The effectiveness and accuracy of the approaches used in predicting wind speed are important to predict the performance of wind energy. The overarching purpose of this article is to build a method enabling the reliable and efficient prediction of wind speed for a particular station based on time-series supervisory control on other stations in Malaysia. Thus, such time-series prediction would describe a model that predicts a future value of any station only using the preceding values of the reference stations. Another contribution in this study is to bring this method to the attention of the renewable energy community and show how it could be employed in a new field.

2. Wind energy in Malaysia

In their growth cycles, all nations rely heavily on the power market, and the world's demand for power is rising daily. According to British Petroleum, the use of prime power increased by 2.2% from 2013 to 2017. The largest increases in energy consumption among fuel forms have been in pure gas and oil. However, renewable energy do not yet have large quotas of the overall energy portfolio compared to non-renewable energy. Malaysia is heavily reliant on fossil fuels, which generate over 90% of the country's energy, because of a lack of renewables. For example, the Malaysian government emphasizes renewable energies for power production, especially wind energy projects; however, wind energy development is at a major disadvantage, as Malaysia is situated in a low wind-speed area (Albani & Ibrahim, 2017).

Many projects do not achieve their desired goal. At Pulau Terumbu Layang in Sabah, S. K. Najid et al. presented a 150 kW wind turbine project. According to them, this was Malaysia's first wind turbine (Najid et al., 2009). Universiti Kebangsaan Malaysia (UKM) extended this proposed project and combined it with diesel systems that generate power to support the army and the nearest resort. Tenaga Nasional Berhad (TNB) in Perhentian Island installed the most famous wind turbine project. With 100 kW of photovoltaic and 100 kW of diesel generators, this was a hybrid project, with recorded wind speeds of 3.6 m/s and 15.6 m/s, respectively (Karim et al., 2009).

In Malaysia, the gross littoral region is approximately 4675 km, the longest in the world (Ahmad & Tahar, 2014). Malaysia, therefore, accepts the value of renewable energy as a source of electricity, rather than petrol. Small Renewable Energy Power (SREP) was initiated to improve renewable energy evolution, but the results were disappointing. With the overall electricity produced from renewable energy, the production rate of renewable energy is sluggish. The Malaysian parliament subsequently enacted the 2011 Clean Energy Act (Act 725) for adoption in 2011. In 2015, the target for wind power generation was stated to be 985 MW. However, about 400 MW was generated earlier in 2015. In comparison, the success rate (50%) reached the initial target, as the objective was reported to be unattainable. Furthermore, the 2015 target was 985 MW, whereas 2020 and 2030 are expected to contribute 2080 and 4000 MW, respectively. In Malaysia, the wind power project is used only for educational investigations (Sarkar et al., 2019).

3. Methods and materials

3.1. Study area and wind speed data

Malaysia is a country that lies completely in the equatorial zone or equatorial low Intertropical Convergence Zone. The Asian monsoon, including the weather in Malaysia, is related to a greater than average seasonal migration of the Intertropical Convergence Zone (Ibrahim et al., 2014).

Malaysia's climate is split into seasons: the inter-monsoon in April, the monsoon in the south-west from mid-May to September, the inter-monsoon in October, and the monsoon in the north-east from November to March. Also, in Malaysia, wind usually is regular, with the highest wind flows occurring in the afternoon and the lowest wind flows before morning. During the south-west monsoon, the prevailing wind flow is generally south-westerly and light, lower than 7.5 m/s. During the north-east monsoon, a stable easterly or north-easterly wind of 5–10 m/s prevails. Nonetheless, on the east coast of Malaysia wind speeds may reach 15 m/s or more. Through the two inter-monsoon seasons, winds are generally light and variable. In a normal year, high pressure over the eastern Pacific causes warm equatorial water and the surface wind to flow westward, and converge near Indonesia. Consequently, from April to November, the south-westerly winds on the north-west coast of Sabah and Sarawak regions may intensify to reach 10 m/s or higher (Ibrahim et al., 2014). However, the wind flow in Malaysia is influenced by mesoscale or local wind. Because Malaysia is mostly a marine country, the influence of land and sea breezes on the overall wind flow pattern is important, particularly on clear days. Sea and land breezes of 5–7.5 m/s frequently arise on bright sunny afternoons or clear nights, reaching speeds of up to tens of kilometers inland and along the shore. Daily wind in many mountainous places is known as mountain and valley breeze, and it is similar to land and sea breeze.

During the day, the valley breeze is warm air rising down the mountain slope, but after sunset, the mountain breeze is cool air draining into the valley. The average wind speed in the valley is more than 9 m/s, although the mountain breeze is generally stronger, with winds exceeding 11 m/s.

Existing wind speed data or secondary data are valuable in assessing wind resources in the early stages of wind energy development (Ibrahim et al., 2014). The study area selected in this article was Malaysia, so the wind speeds must be obtained in Malaysia. The 14 selected stations are Alor Setar, Bayan Lepas, Cameron Highlands, Ipoh, Klia Sepang, Kota Bharu, Kota Kinabalu, Kuantan, Kuching, Lubok Merbau, Muadzam Shah, Pulau Langkawi, Sitiawan and Subang). Daily wind data at these 14 stations were obtained from the Malaysian Meteorological Department. The locations of these 14 stations are shown in Figure 1.

Figures 2 and 3 show the monthly median wind speeds at the 14 stations in Malaysia. The 24 h wind data have been recorded for 19 years, between 2000 and 2019.

Two sets of average wind speed time series sampled with wind speed data are used to validate the predictive efficacy of the proposed model. Each wind speed series of stations comprises 6400 tests, 80% of which are used for preparation and 20% for research. For example, one of the 14 stations is shown in Figure 1 with the time-series data of maximum, mean and minimal wind speeds, and statistics for the 14 stations are measured and presented in Table 1.

3.2. Gaussian process regression (GPR) approach

Gaussian processes (GPs) are described as sets of random variables, some of which have a multiple-variable Gaussian distribution. is the input domain, Y is the output domain, and a pairs (x_ι, y_ι) are independent, then extracted and distributed equally. It is supposed that the GP on X is defined by the average and covariance functions (μ: $Y \rightarrow Re$ and A: $X * X \rightarrow Re$), respectively. The main theory in GPR is that y is determined through $y = f(x) + \zeta_\iota$, where ζ represents the Gaussian noise, with variance of σ^2. There are random variables $f(x)$ for the input variables x in GPR, in which the values of the random functions f are in certain locations. In the current work, it is presumed that the observation errors are independent and have a similar distribution to the zero mean value ($\mu\ (x) = 0$) and the variance (σ^2), and $f(x)$ of the GP on X (denoted by A) as:

$$y = (y_1, \ldots, y_a) \sim N\,(o,\ A + \sigma^2 K) \tag{1}$$

where K is the identity matrix and $A_{\iota j} = A(x_\iota,\ y_\iota)$; $y/x \sim N\ (o,\ A + \sigma^2 K)$ is normal. More details can be found in Shabani et al. (2020).

3.3. Support vector regression (SVR) approach

This section describes the general concept of SVR modeling. The SVM was initially created by Vapnik (Sain,

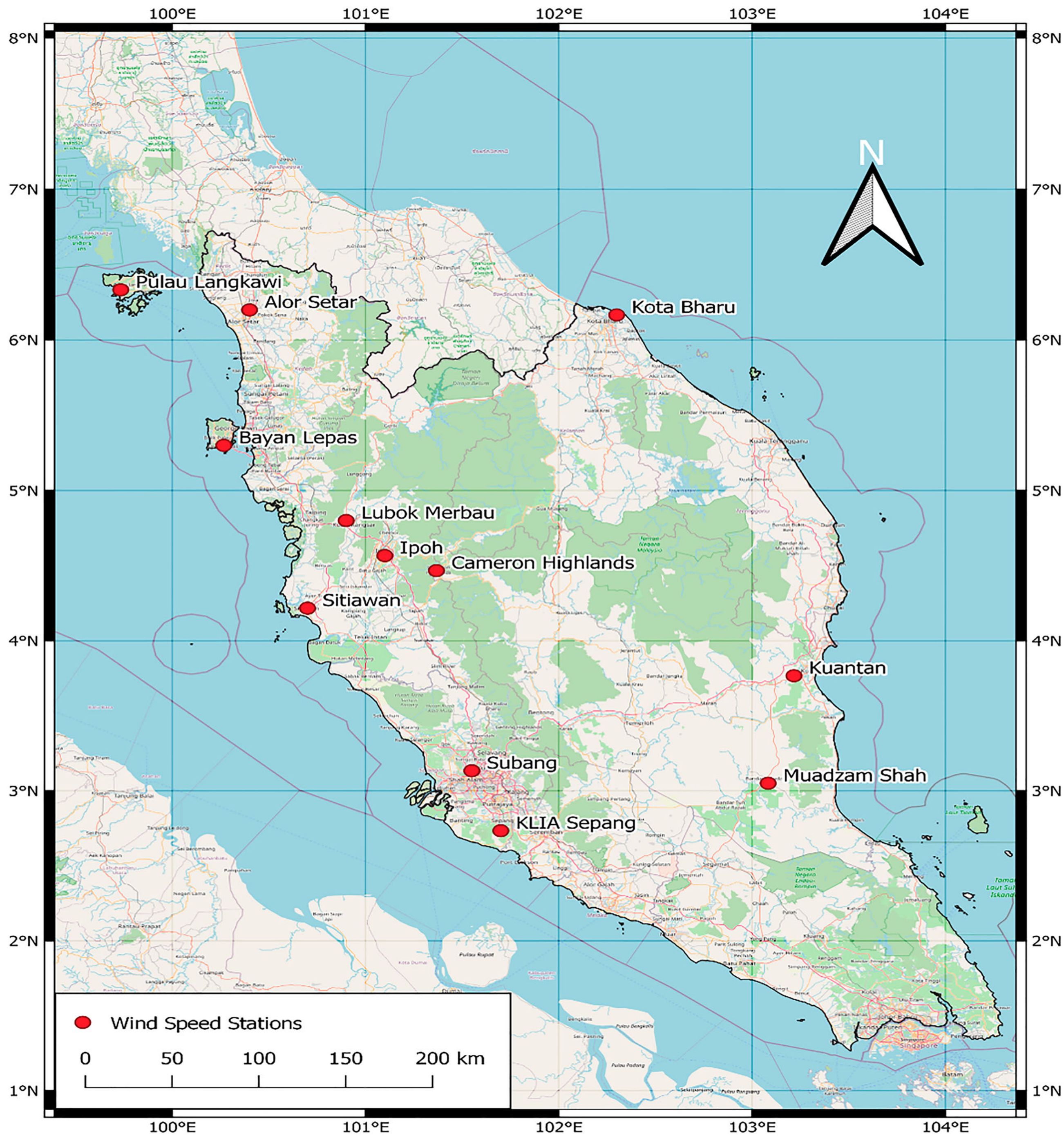

Figure 1. Location of the 14 selected wind speed stations in Malaysia.

1996), and has been widely used for classification (Yin & Hou, 2016; Citakoglu, 2021) and regression (Olatomiwa et al., 2015) problems. In this study, the regression type of SVM will be used, since it has been demonstrated that it has great properties in training limited samples (Smola & Schölkopf, 2004). Moreover, this method can be used to uncover nonlinearity features in multivariate input and output datasets using the kernel trick. In epsilon-SVR, the goal is to achieve with a function $f(x)$, which can deviate from the present target y_i at most by epsilon for the whole learning data, and is as flat as possible. At this time, the error remains appropriate because it is less than epsilon, but any deviations larger than epsilon are not satisfactory. In ordinary least square regressions, the goal is to reduce errors, whereas in SVR modeling the aim is to fit errors in a specific threshold. Additional details regarding SVR are described in Smola and Schölkopf (2004).

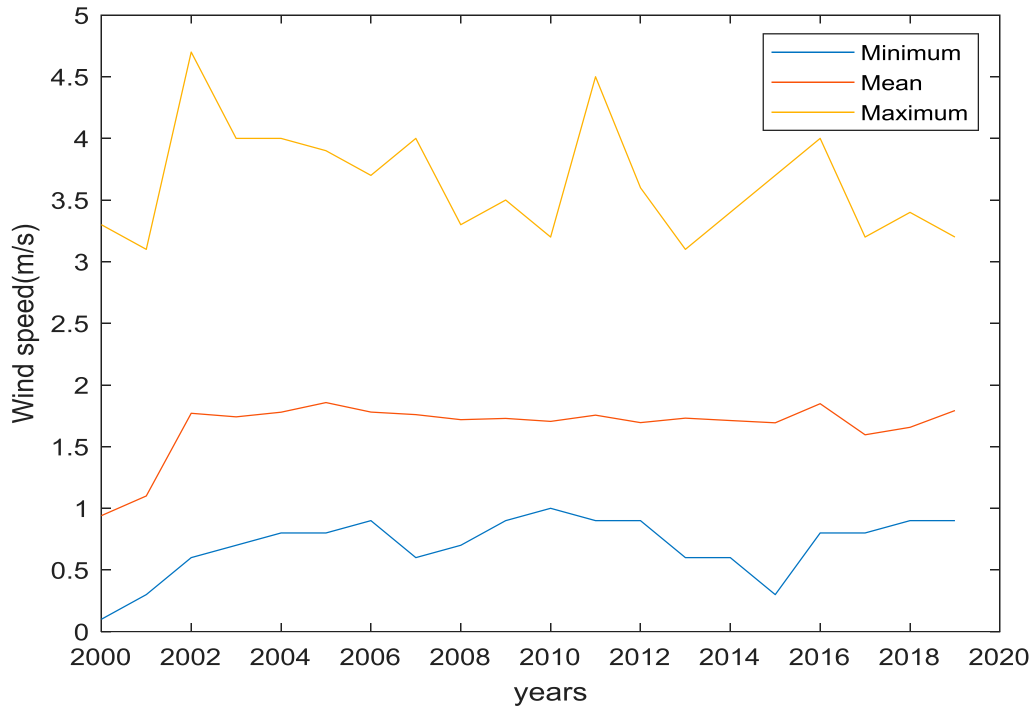

Figure 2. Annual wind speed in Malaysia (Alor Setar).

3.4. Bagged regression trees (BTs) approach

The core of bootstrap aggregating (bagging) trees, which was constructed by Breiman (1996), is to construct many similar independent predictors and average the output of the predictors to obtain the final predictions. This allows for decreasing variance errors (Sutton, 2005). In the bagging ensemble of the decision trees method, numerous trees (single models) are combined to improve the predictive quality of the model. The bagging trees prediction approach is used to decrease the variation of the regression trees and to handle overfitting problems in individual trees. The initial stage in bagging trees is to create A new training datasets of equal size by selecting samples in a uniform way with replacement from actual training data. Thus, every tree in the ensemble is trained independently on the corresponding new training sets. Lastly, the average of the entire predictions is calculated to obtain the final forecast. The predictions of the BTs approach are defined as (Harrou et al., 2019):

$$\hat{y} = \frac{1}{A}\sum_{\iota=1}^{A} f_\iota(X) \tag{2}$$

where every tree model f_ι is trained on the bootstrapped dataset ι. The important steps used in calculating the BTs predictions are defined in Algorithm 1:

Inputs: Training dataset and testing dataset, D.

Outputs: Prediction outputs.

For $\iota = 1, \ldots, A$ in training dataset do:

- Take a bootstrap replica D, from D_ι
- Call decision trees with D_ι and received predictions $\widehat{y_i}$
- Add y to ensemble $\hat{y}$
- Calculate final prediction: $\hat{y} = \frac{1}{A}\sum_{\iota=1}^{A}\widehat{y_i}$

End.

$$Prediction_{Bagged\ Trees} \leftarrow \hat{y};$$

Return $Prediction_{Bagged\ Trees}$

3.5. Model evaluation

To evaluate the prediction accuracy of the models proposed in the current work, various performance indicators were applied, as shown in Table 2, including: the correlation coefficient (R) (Shabani et al., 2020), mean absolute error (MAE), root mean squared error (RMSE) (Bokde et al., 2020), determination coefficient (R^2), Nash–Sutcliffe efficiency (NSE), mean squared error (MSE), improvement percentage of mean absolute error (PMAE), improvement percentage of root mean squared error (PRMSE) (Xiwei Mi et al., 2019; Xi wei Mi et al., 2017) and improvement percentage of mean squared error (PMSE).

y_i and $\widehat{y_i}$ represent the observed and predicted values of wind speed in time step ι, respectively; $(\bar{y})$ and $\overline{\widehat{y}_i}$ represent the mean values of the actual and predicted values

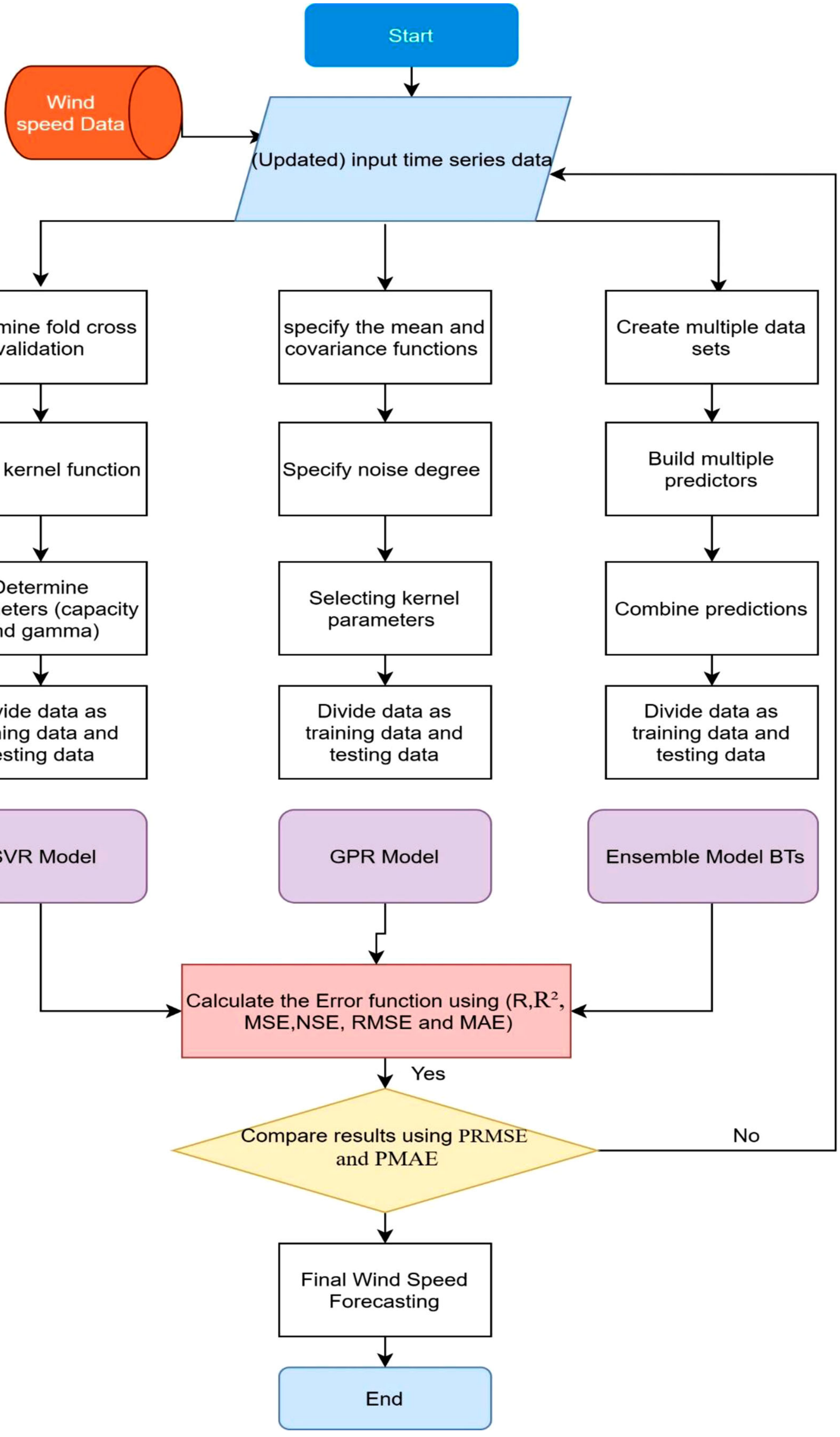

Figure 3. General data flow diagram of the regression models. SVR = support vector regression; GPR = Gaussian process regression; BTs = bagged regression trees; R = correlation coefficient; MSE = mean squared error; NSE = Nash–Sutcliffe efficiency; RMSE = root mean squared error; MAE = mean absolute error; PRMSE = improvement percentage of root mean squared error; PMAE = improvement percentage of mean absolute error.

Table 1. Simple statistics of wind speed data.

	Alor Setar	Bayan Lepas	Cameron Highlands	Ipoh	Klia Sepang	Kota Bharu	Kota Kinabalu	Kuantan	Kuching	Lubok Merbau	Muadzam Shah	Pulau Langkawi	Sitiawan	Subang
Mean	1.72	1.96	1.92	1.56	1.86	2.30	2.06	1.65	1.67	1.34	0.81	2.07	1.12	1.58
Standard error	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.01
Median	1.70	1.90	1.60	1.60	1.80	2.10	2.00	1.60	1.70	1.30	0.70	1.80	1.10	1.50
Mode	1.50	1.70	1.50	1.60	1.60	2.00	1.90	1.50	1.70	1.30	0.70	1.60	1.10	1.30
Standard deviation	0.51	0.62	0.96	0.46	0.58	0.90	0.50	0.49	0.43	0.39	0.46	0.90	0.38	0.47
Sample variance	0.26	0.38	0.92	0.21	0.33	0.82	0.25	0.24	0.19	0.15	0.21	0.81	0.14	0.22
Kurtosis	1.99	3.37	3.47	1.00	4.60	5.72	5.27	0.57	1.47	3.78	5.36	4.48	0.41	1.29
Skewness	0.83	1.35	1.63	−0.05	1.51	2.00	1.61	0.51	0.07	1.00	1.61	1.76	−0.01	0.79
Range	4.60	5.60	8.40	4.40	6.10	8.20	5.90	4.00	4.00	4.70	4.30	8.20	2.80	4.20
Minimum	0.10	0.20	0.20	0.00	0.20	0.40	0.40	0.00	0.10	0.10	0.00	0.50	0.00	0.20
Maximum	4.70	5.80	8.60	4.40	6.30	8.60	6.30	4.00	4.10	4.80	4.30	8.70	2.80	4.40
Sum	10,960.50	12,539.90	12,240.90	9980.40	11,852.50	14,693.30	13,140.70	10,566.20	10,650.10	8591.40	5176.10	13,239.70	7162.90	10,094.40
Count	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00	6389.00

Table 2. Statistical indices.

Evaluation metric	Equation
MAE	$\mathrm{MAE} = \frac{\sum_{\iota=1}^{n} \lvert y_\iota - \widehat{y}_i \rvert}{n}$
RMSE	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{\iota=1}^{n}(y_\iota - \widehat{y}_i)^2}$
R	$R = \frac{\sum_{\iota=1}^{n}(y_\iota - \bar{y})(\widehat{y}_i - \overline{\widehat{y}}_i)}{\sqrt{\sum_{\iota=1}^{n}(y_\iota - \bar{y})^2}\sqrt{\sum_{\iota=1}^{n}(\widehat{y}_i - \overline{\widehat{y}}_i)^2}}$
R^2	$R^2 = \left[\sum_{\iota=1}^{n}(y_\iota - \bar{y})(\widehat{y}_i - \overline{\widehat{y}}_i)\right]^2 \left[\sum_{\iota=1}^{n}(y_\iota - \bar{y})^2 \sum_{\iota=1}^{n}(\widehat{y}_i - \overline{\widehat{y}}_i)^2\right]^{-1}$
NSE	$NSE = 1 - \frac{\sum_{i=1}^{n}(\widehat{y}_i - y_\iota)^2}{\sum_{i=1}^{n}(y_\iota - \bar{y})^2}$
MSE	$\mathrm{MSE} = \frac{1}{n}\sum_{\iota=1}^{n}(y_\iota - \hat{y}_\iota)^2$

Note: MAE = mean absolute error; RMSE = root mean squared error; R = correlation coefficient; NSE = Nash–Sutcliffe efficiency; MSE = mean squared error.

of wind speed in time step ι, respectively; and n is the number of sets.

4. Results and discussion

This section contains two subsections: the results and discussion of the performance evaluation for the SVR, BTs and GPR models, and the results and discussion of the comparisons of these models.

4.1. Performance machine learning techniques

The three machine learning methods were trained using historical wind speed data obtained between 2000 and 2019. These data were arranged into maximum, mean and minimum weekly wind speed values of the reference stations, and corresponding weeks were used. Every dataset was split into training and checking divisions. This process was proposed to prevent overfitting of the systems to the training dataset (Jang et al., 1997).

In general, the goal of prediction is to investigate the correlation between the input process variables (13 stations) and the output variable (one station) in the training stage, thus providing the prediction value of the output for the given unseen input dataset. More precisely, initially, the methods (SVR, BTs and GPR models) were constructed using the training input and output datasets X and y. After that, for the unseen inputs test dataset, the built model was used to predict the input variable "wind speed". Lastly, the performance of prediction was checked; R, MAE, MAE and RMSE are the most commonly used metrics for checking the quality of predictions.

The K-fold cross-validation (CV) technique, which is often performed to provide an appropriate estimate of a

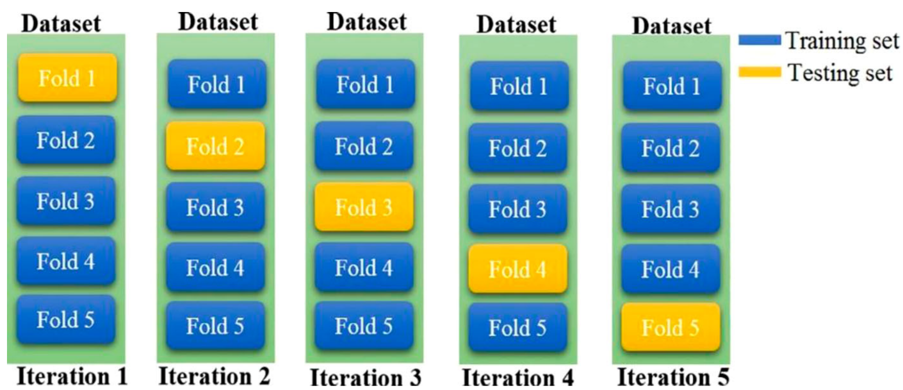

Figure 4. Five-fold cross-validation process (Harrou et al., 2019).

model's prediction errors, was used to construct prediction models (Hastie et al., 2009). A training dataset was split into K-folds, where each portion was used as test data at some point (Figure 4). This allows the construction of trustworthy prediction models. Figure 4 presents a K-fold cross-validation with $K = 5$, as used in this paper. The presented dataset was randomly divided into K equal subsets. As shown in Figure 4, in iteration 1, fold 1 was using for testing and the other folds for training. In iteration 2, fold 2 was used to test the model and the others were used for training.

This process was repeating until it was confirmed that all folds had been used as the test dataset. The MSE was calculated for every test sub-dataset, MSE_ι. The cross-validation error is calculated as the mean of the prediction errors, as:

$$\mathrm{CVerror} = \frac{1}{\mathrm{k}} \sum_{\iota=1}^{\mathrm{k}} \mathrm{MSE}_\iota \qquad (3)$$

In practice, cross-validation with $K = 5$ usually uses $\iota equal1$ in model construction (Hastie et al., 2009).

To predict the maximum values of weekly wind speed for each station, we applied the SVR model. The ability of the SVR model to achieve the performance goal depends on predefined internal SVR parameters, such as capacity and gamma of kernel functions. The performance values for each station using the SVR model (RMSE, MSE, NSE, MAE, R and R^2) for both training and test phases are specified in Table 3. The MAE values, ranging from 0.2 to 0.6, vary from the observed value for each station and test procedure. The highest MAE was observed to be 0.7 for Cameron Highlands station in the test phase, whereas the best result was 0.197785 for Muadzam Shah station in the test phase. The best R and R^2 values between the target and output values reached 0.768 and 0.59, respectively, for Muadzam Shah station in the training stage. In the test phase, the highest values of R and R^2 were 0.773 and 0.597, respectively, for Kota Bharu station. At this station a satisfactory range of NSE was reached, at 0.55476. A value of NSE equal to 1.0 represents an ideal fit, NSE greater than 0.75 is a very good fit, NSE of 0.64–0.74 represents a good fit, NSE of 0.5–0.64 is a satisfactory value and NSE less than 0.5 refers to unsatisfactory prediction (Moriasi et al., 2007). Concerning the testing phase, five stations showed an unacceptable range of performance (Ipoh, Kota Kinabalu, Kuching, Lubok Merbau and Sitiawan). Lower R values, of 0.3 and 0.4, were found for Kota Kinabalu station for the training and testing stages, respectively. At Sitiawan station, for the training procedure, preferable values of RMSE and MSE, of 0.29 and 0.088, respectively, were obtained, while regarding the testing procedure a satisfactory value of NSE, at 0.534, and the lowest (better) values of RMSE and MSE were found, at 0.265645 and 0.070567, for Muadzam Shah station.

The results in Table 3 show that the predictive model BTs demonstrates good performance, with the lowest RMSE in the training phase, with correlation coefficients of more than 0.5 in all stations except for one, Kota Kinabalu, which showed an unsatisfactory fit, with NSE and R^2 values of 0.15 and 0.095, respectively, while in the testing phase, in addition to this station, the predicted maximum weekly wind speed values at three other stations (Ipoh, Kuching and Lubok Merbau,) using the other 13 stations as input, displayed unacceptable performance.

The third model implemented in this study, GPR, showed better performance than the other models in predicting maximum wind speed values, as shown in Table 3. As seen from the training procedure, the maximum and minimum values of RMSE range between 0.3 and 0.8 at all stations. It can also be noted that the lowest value of MAE is 0.22, at Sitiawan station. Among all the output results for prediction at all stations, the R, R^2 and NSE at Kota Kinabalu station did not achieve satisfactory values in either the training or testing stages. Better values

Table 3. Performance evaluation of the training and testing procedures at maximum wind speed.

	Training procedure					Testing procedure					
Station	RMSE	MSE	MAE	R^2	R	RMSE	MSE	NSE	MAE	R^2	R
SVR											
Alor Setar	0.36292	0.13171	0.26235	0.57	0.75498	0.30476	0.09288	0.54672	0.23686	0.55295	0.74361
Bayan Lepas	0.51854	0.26888	0.37586	0.46	0.67823	0.51401	0.26421	0.28309	0.37973	0.36704	0.60584
Cameron Highlands	0.84839	0.71976	0.64937	0.45	0.67082	1.05910	1.12170	0.33047	0.71446	0.42482	0.65178
Ipoh	0.35734	0.12769	0.27085	0.35	0.59161	0.38492	0.14816	0.02445	0.27368	0.10743	0.32776
Klia Sepang	0.50318	0.25319	0.36647	0.47	0.68557	0.48569	0.23589	0.47774	0.34772	0.48231	0.69449
Kota Bharu	0.79622	0.63397	0.54859	0.5	0.70711	0.82080	0.67371	0.55476	0.49413	0.59768	0.77310
Kota Kinabalu	0.56189	0.31572	0.40349	0.15	0.38730	0.58728	0.34490	0.04397	0.44740	0.08978	0.29963
Kuantan	0.36290	0.13170	0.27548	0.41	0.64031	0.31146	0.09700	0.43024	0.24817	0.43251	0.65765
Kuching	0.35244	0.12421	0.26064	0.37	0.60828	0.33982	0.11548	0.08501	0.25828	0.12591	0.35484
Lubok Merbau	0.32709	0.10699	0.24581	0.41	0.64031	0.32928	0.10842	0.11624	0.22074	0.17141	0.41402
Muadzam Shah	0.35510	0.12610	0.24111	0.59	0.76812	0.26565	0.07057	0.53409	0.19779	0.55121	0.74244
Pulau Langkawi	0.76255	0.58149	0.48582	0.58	0.76158	0.58553	0.34285	0.50145	0.45167	0.54057	0.73523
Sitiawan	0.29354	0.08617	0.22699	0.4	0.63246	0.27081	0.07334	0.13605	0.20793	0.20916	0.45734
Subang	0.37250	0.13876	0.28581	0.42	0.64807	0.37970	0.14417	0.32660	0.28406	0.34383	0.58637
BTs											
Alor Setar	0.34667	0.12018	0.25436	0.61	0.78103	0.28811	0.08301	0.59641	0.22624	0.59651	0.77234
Bayan Lepas	0.50979	0.25989	0.37872	0.48	0.69282	0.47041	0.22128	0.40264	0.35481	0.44031	0.66356
Cameron Highlands	0.83248	0.69303	0.64357	0.47	0.68557	1.05203	1.10678	0.34159	0.71727	0.40623	0.63736
Ipoh	0.36224	0.13122	0.27333	0.34	0.58310	0.36640	0.13425	0.11605	0.27032	0.17453	0.41776
Klia Sepang	0.51615	0.26641	0.38182	0.44	0.66333	0.45386	0.20599	0.54394	0.33018	0.55684	0.74622
Kota Bharu	0.76545	0.58592	0.52703	0.54	0.73485	0.78343	0.61377	0.59437	0.50626	0.62574	0.79104
Kota Kinabalu	0.56024	0.31387	0.40870	0.15	0.38730	0.57029	0.32524	0.09716	0.44241	0.10236	0.31994
Kuantan	0.35129	0.12341	0.26784	0.44	0.66333	0.30105	0.09063	0.56668	0.23494	0.47672	0.69045
Kuching	0.34027	0.11578	0.25388	0.41	0.64031	0.32037	0.10264	0.40034	0.24812	0.22697	0.47641
Lubok Merbau	0.32010	0.10246	0.24407	0.43	0.65574	0.32983	0.10879	0.26766	0.23370	0.19977	0.44696
Muadzam Shah	0.36366	0.13225	0.25941	0.57	0.75498	0.23526	0.05535	0.63457	0.17967	0.64454	0.80283
Pulau Langkawi	0.75463	0.56946	0.50723	0.59	0.76812	0.61377	0.37671	0.45221	0.48272	0.56378	0.75085
Sitiawan	0.29712	0.08828	0.22913	0.38	0.61644	0.26092	0.06808	0.19796	0.19577	0.28074	0.52985
Subang	0.37244	0.13871	0.28947	0.42	0.64807	0.35250	0.12426	0.41961	0.26389	0.42069	0.64860
GPR											
Alor Setar	0.33012	0.10898	0.24860	0.65	0.80623	0.29108	0.08473	0.58650	0.23322	0.60072	0.77506
Bayan Lepas	0.48239	0.23270	0.36234	0.53	0.72801	0.47155	0.22236	0.39663	0.35631	0.44347	0.66593
Cameron Highlands	0.78779	0.62061	0.61154	0.53	0.72801	1.00190	1.00381	0.40084	0.68729	0.45585	0.67517
Ipoh	0.34725	0.12058	0.26534	0.39	0.62450	0.36487	0.13313	0.12340	0.26738	0.16592	0.40734
Klia Sepang	0.49881	0.24881	0.36656	0.48	0.69282	0.46718	0.21826	0.51677	0.35293	0.51817	0.71984
Kota Bharu	0.75850	0.57532	0.52655	0.55	0.74162	0.76225	0.58102	0.61601	0.49378	0.62433	0.79015
Kota Kinabalu	0.55094	0.30354	0.40351	0.18	0.42426	0.58198	0.33870	0.06116	0.45237	0.08214	0.28660
Kuantan	0.34582	0.11595	0.26444	0.46	0.67823	0.31387	0.09851	0.42139	0.25011	0.42644	0.65303
Kuching	0.34876	0.12163	0.26540	0.38	0.61644	0.34396	0.11831	0.06259	0.26577	0.13804	0.37154
Lubok Merbau	0.30416	0.09251	0.23482	0.49	0.70000	0.31545	0.09951	0.18890	0.21831	0.25447	0.50445
Muadzam Shah	0.32019	0.10252	0.24670	0.66	0.81240	0.25950	0.06734	0.55541	0.19852	0.57420	0.75776
Pulau Langkawi	0.72367	0.52370	0.49030	0.62	0.78740	0.66184	0.43803	0.36304	0.49807	0.48444	0.69602
Sitiawan	0.28694	0.08234	0.22447	0.42	0.64807	0.26874	0.07222	0.14921	0.21052	0.22041	0.46948
Subang	0.36782	0.13529	0.28518	0.44	0.66333	0.35722	0.12761	0.40398	0.27292	0.40810	0.63883

Note: RMSE = root mean squared error; MSE = mean squared error; MAE = mean absolute error; R = correlation coefficient; NSE = Nash–Sutcliffe efficiency; SVR = support vector regression; BTs = bagged regression trees; GPR = Gaussian process regression.

of NSE and R^2 of more than 0.6 were found at Kota Bharu station in the testing procedure.

Regarding the predicted mean wind speed, the wind speeds of 13 independent stations were utilized as inputs, whereas the wind speeds of the target station were utilized as the output in the SVR model. The findings obtained with this approach were compared with the measurement dataset. Errors found in the SVR method are within satisfactory limits for prediction at each station in the training procedure, where R values ranged between 0.5 and 0.8 (Table 4). Regarding the best value of NSE in the testing procedure, it was found to be 0.7, representing good fit, at Pulau Langkawi station. However, as mentioned before, in the testing procedure, forecasting of three stations (Kota Kinabalu, Ipoh and Kuching) did not achieve an acceptable range of performance, with R values of 0.187306, 0.441994 and 0.44375, respectively.

By applying the BTs ensemble method, better results were obtained compared with the previous model, SVR, at the 14 forecasting stations. The training procedure saw the maximum determination coefficient of 0.73 obtained at Pulau Langkawi station, and the minimum determination coefficient was found to be 0.23 at Kota Kinabalu station. The values of MAE ranged from 0.16 at Muadzam Shah to 0.34 at Cameron Highlands station. In the testing stage, a good value of NSE, with

Table 4. Performance evaluation of the training and testing procedures at mean wind speed.

	Training procedure					Testing procedure					
Station	RMSE	MSE	MAE	R^2	R	RMSE	MSE	NSE	MAE	R^2	R
SVR											
Alor Setar	0.25259	0.06380	0.17784	0.64	0.8	0.21027	0.04421	0.62775	0.16769	0.63189	0.79492
Bayan Lepas	0.29786	0.08872	0.20958	0.58	0.76158	0.29066	0.08449	0.41419	0.22376	0.50391	0.70986
Cameron Highlands	0.44483	0.19787	0.32709	0.56	0.74833	0.55208	0.30479	0.44680	0.38211	0.55301	0.74365
Ipoh	0.22849	0.05221	0.17153	0.57	0.75498	0.22575	0.05096	0.01388	0.16948	0.19536	0.44199
Klia Sepang	0.27978	0.07828	0.20489	0.49	0.70000	0.25141	0.06321	0.48841	0.18622	0.49720	0.70513
Kota Bharu	0.42433	0.18005	0.28188	0.6	0.77460	0.50480	0.25483	0.55146	0.32589	0.56504	0.75169
Kota Kinabalu	0.27082	0.07334	0.19670	0.22	0.46904	0.36478	0.13306	−0.00877	0.27203	0.03508	0.18731
Kuantan	0.23479	0.05513	0.17343	0.61	0.78103	0.23614	0.05576	0.48893	0.19088	0.49889	0.70632
Kuching	0.21733	0.04723	0.15740	0.62	0.78740	0.20983	0.04403	−0.13595	0.16463	0.19691	0.44375
Lubok Merbau	0.20823	0.04336	0.15951	0.5	0.70711	0.17246	0.02974	0.30864	0.12976	0.39410	0.62777
Muadzam Shah	0.23310	0.05433	0.14874	0.65	0.80623	0.17411	0.03031	0.56669	0.13227	0.58021	0.76172
Pulau Langkawi	0.38938	0.15162	0.25266	0.74	0.86023	0.32000	0.10240	0.70013	0.25229	0.73758	0.85882
Sitiawan	0.19499	0.03802	0.15002	0.57	0.75498	0.17281	0.02986	0.21989	0.12494	0.28879	0.53739
Subang	0.22497	0.05061	0.16772	0.51	0.71414	0.22022	0.04850	0.40586	0.16183	0.41530	0.64444
BTs											
Alor Setar	0.23840	0.05684	0.17309	0.68	0.82462	0.19851	0.03941	0.66822	0.16043	0.66857	0.81766
Bayan Lepas	0.28944	0.08378	0.21884	0.60	0.77460	0.29066	0.08449	0.41419	0.22376	0.50391	0.70986
Cameron Highlands	0.46258	0.21398	0.34943	0.53	0.72801	0.56920	0.32399	0.41196	0.37878	0.47282	0.68762
Ipoh	0.23935	0.05729	0.17713	0.53	0.72801	0.21560	0.04648	0.10060	0.16213	0.19980	0.44699
Klia Sepang	0.28412	0.08072	0.21171	0.48	0.69282	0.24798	0.06149	0.50228	0.17932	0.51853	0.72009
Kota Bharu	0.43468	0.18895	0.29302	0.58	0.76158	0.52481	0.27543	0.51520	0.32989	0.56045	0.74863
Kota Kinabalu	0.26884	0.07228	0.19702	0.23	0.47958	0.35634	0.12698	0.03738	0.27188	0.05432	0.23306
Kuantan	0.24023	0.05771	0.18135	0.59	0.76812	0.20582	0.04236	0.61172	0.16482	0.61175	0.78214
Kuching	0.20747	0.04304	0.15380	0.65	0.80623	0.19391	0.03760	0.02996	0.15512	0.31880	0.56462
Lubok Merbau	0.21434	0.04594	0.16735	0.47	0.68557	0.18558	0.03444	0.19944	0.13627	0.30549	0.55272
Muadzam Shah	0.22047	0.04861	0.15344	0.68	0.82462	0.15122	0.02287	0.67312	0.11262	0.68707	0.82890
Pulau Langkawi	0.39678	0.15743	0.26194	0.73	0.85440	0.36587	0.13386	0.60799	0.28510	0.68722	0.82899
Sitiawan	0.20905	0.04370	0.16184	0.51	0.71414	0.16406	0.02691	0.29693	0.12367	0.38368	0.61942
Subang	0.23514	0.05529	0.17776	0.46	0.67823	0.21294	0.04535	0.44445	0.15738	0.44989	0.67074
GPR											
Alor Setar	0.21181	0.04486	0.16032	0.74	0.86023	0.21410	0.04584	0.61406	0.16826	0.62464	0.79034
Bayan Lepas	0.26071	0.06797	0.19741	0.68	0.82462	0.27863	0.07763	0.46170	0.21548	0.56082	0.74888
Cameron Highlands	0.41693	0.17383	0.31324	0.62	0.78740	0.51782	0.26814	0.51332	0.36038	0.56029	0.74853
Ipoh	0.22109	0.04888	0.16933	0.60	0.77460	0.21339	0.04553	0.11895	0.16417	0.24501	0.49498
Klia Sepang	0.27239	0.07420	0.20057	0.52000	0.72111	0.25735	0.06623	0.46396	0.19204	0.49365	0.70260
Kota Bharu	0.41236	0.17004	0.27980	0.62	0.78740	0.50089	0.25089	0.55838	0.33143	0.56643	0.75261
Kota Kinabalu	0.26471	0.07002	0.19547	0.26	0.50990	0.35624	0.12691	0.03789	0.26849	0.05879	0.24246
Kuantan	0.22147	0.04905	0.16668	0.65	0.80623	0.22970	0.05276	0.51641	0.18061	0.52799	0.72663
Kuching	0.20542	0.04220	0.15346	0.66	0.81240	0.21838	0.04769	−0.23041	0.17501	0.25207	0.50206
Lubok Merbau	0.19572	0.03831	0.15464	0.56	0.74833	0.16794	0.02820	0.34444	0.12819	0.41432	0.64368
Muadzam Shah	0.20256	0.04103	0.14003	0.73	0.85440	0.16357	0.02676	0.61754	0.12375	0.64608	0.80379
Pulau Langkawi	0.36393	0.13244	0.24158	0.77	0.87750	0.34300	0.11765	0.65547	0.26741	0.72328	0.85046
Sitiawan	0.19066	0.03635	0.14803	0.59	0.76812	0.16779	0.02816	0.26453	0.12242	0.33322	0.57726
Subang	0.22100	0.04884	0.16581	0.53	0.72801	0.20846	0.04346	0.46761	0.15775	0.47076	0.68612

Note: RMSE = root mean squared error; MSE = mean squared error; MAE = mean absolute error; R = correlation coefficient; NSE = Nash–Sutcliffe efficiency; SVR = support vector regression; BTs = bagged regression trees; GPR = Gaussian process regression.

0.67, was found at Muadzam Shah station. In the testing stage, the results were less satisfactory compared with the training stage, but still obtained an acceptable range of errors.

Table 4 shows that the GPR technique could effectively predict the mean weekly wind speeds of any target station utilizing the measurement data of the other 13 stations as in the training phase for each station.

In the context of predicting minimum wind speed values at the 14 independent measurement stations, SVR performed well and generated minor prediction errors in 13 stations (Table 5). The minimum value of RMSE was obtained at Sitiawan station, with 0.22203, which is the best value in the training phase. In contrast, the best R^2 value was moderate, at 0.59, at Pulau Langkawi station compared with the other stations. It can be observed that the lowest value of MAE was 0.153, at Muadzam Shah station. The output results of SVR in the testing phase showed a lower value of R^2 obtained at Kota Kinabalu station, which indicates unsuccessful prediction at this station.

According to the implementation of the ensemble BTs, the results obtained indicate that the correlation coefficients for only two stations (Kota Kinabalu and Subang) obtained less than the moderate value of 0.5 in the training stage. On the other hand, in the testing procedure,

Table 5. Performance evaluation of the training and testing procedures at minimum wind speed.

	Training procedure					Testing procedure						
Station	RMSE	MSE	MAE	R^2	R	RMSE	MSE	NSE	MAE	R^2	R
SVR											
Alor Setar	0.28011	0.07846	0.21084	0.46	0.67823	0.25937	0.06727	0.31334	0.19658	0.33852	0.58182
Bayan Lepas	0.30416	0.09251	0.22770	0.40	0.63246	0.29731	0.08839	0.20231	0.23260	0.29375	0.54199
Cameron Highlands	0.37358	0.13956	0.25832	0.26	0.50990	0.45279	0.20501	0.13933	0.30851	0.23351	0.48323
Ipoh	0.26819	0.07193	0.20990	0.44	0.66333	0.28242	0.07976	−0.25598	0.22005	0.02733	0.16531
Klia Sepang	0.28955	0.08384	0.22167	0.22	0.46904	0.23385	0.05468	0.01316	0.18616	0.13601	0.36880
Kota Bharu	0.43081	0.18560	0.32234	0.22	0.46904	0.51585	0.26610	0.13204	0.34020	0.19645	0.44322
Kota Kinabalu	0.23686	0.05610	0.18010	0.16	0.40000	0.34717	0.12053	0.00073	0.27464	0.01422	0.11924
Kuantan	0.28590	0.08174	0.22148	0.41	0.64031	0.26393	0.06966	0.33994	0.21302	0.35879	0.59899
Kuching	0.23851	0.05689	0.18326	0.55	0.74162	0.23724	0.05628	−0.48375	0.18993	0.01732	0.13159
Lubok Merbau	0.23437	0.05493	0.18539	0.33	0.57446	0.22077	0.04874	−0.07805	0.17320	0.14165	0.37636
Muadzam Shah	0.22237	0.04945	0.15364	0.44	0.66333	0.17671	0.03123	0.30718	0.13265	0.33803	0.58141
Pulau Langkawi	0.35903	0.12890	0.23858	0.59	0.76812	0.33646	0.11320	0.42494	0.26209	0.49775	0.70551
Sitiawan	0.22203	0.04930	0.17145	0.45	0.67082	0.20769	0.04313	−0.10026	0.15949	0.11633	0.34107
Subang	0.26005	0.06763	0.19370	0.24	0.48990	0.23235	0.05399	0.03885	0.18025	0.07563	0.27501
BTs											
Alor Setar	0.26461	0.07002	0.20024	0.52	0.72111	0.24224	0.05868	0.40104	0.18395	0.42016	0.64820
Bayan Lepas	0.30031	0.09019	0.22640	0.42	0.64807	0.28372	0.08050	0.27356	0.22738	0.34038	0.58342
Cameron Highlands	0.36667	0.13445	0.25823	0.29	0.53852	0.42375	0.17957	0.24617	0.29616	0.30271	0.55019
Ipoh	0.26549	0.07048	0.20819	0.45	0.67082	0.28354	0.06278	0.01143	0.19840	0.08040	0.25056
Klia Sepang	0.28888	0.08345	0.21940	0.22	0.46904	0.22376	0.05007	0.09646	0.17267	0.19134	0.43743
Kota Bharu	0.42130	0.17749	0.31689	0.25	0.50000	0.47527	0.22588	0.26323	0.33012	0.31952	0.56526
Kota Kinabalu	0.24276	0.05893	0.18319	0.12	0.34641	0.34645	0.12003	0.00490	0.27257	0.01383	0.11761
Kuantan	0.28658	0.08213	0.22477	0.40	0.63246	0.23249	0.05405	0.48781	0.18945	0.50178	0.70836
Kuching	0.22952	0.05268	0.17578	0.58	0.76158	0.21316	0.04544	−0.19791	0.17262	0.08450	0.29069
Lubok Merbau	0.24126	0.05821	0.19082	0.29	0.53852	0.20058	0.04023	0.11013	0.15580	0.19335	0.43972
Muadzam Shah	0.20854	0.04349	0.15231	0.51	0.71414	0.17135	0.02936	0.34855	0.12595	0.37087	0.60899
Pulau Langkawi	0.35180	0.12376	0.23314	0.60	0.77460	0.33911	0.11499	0.41586	0.26443	0.51442	0.71723
Sitiawan	0.22644	0.05128	0.17778	0.43	0.65574	0.20738	0.04301	−0.09700	0.16700	0.13396	0.36601
Subang	0.26749	0.07155	0.19944	0.20	0.44721	0.22276	0.04962	0.11657	0.17452	0.12666	0.35589
GPR											
Alor Setar	0.25776	0.06644	0.19874	0.55	0.74162	0.25689	0.06599	0.32643	0.19729	0.36558	0.60464
Bayan Lepas	0.29476	0.08688	0.22272	0.44	0.66333	0.28588	0.08173	0.26245	0.22381	0.31771	0.56366
Cameron Highlands	0.36218	0.13118	0.25954	0.30	0.54772	0.42375	0.17957	0.24617	0.29616	0.30271	0.55019
Ipoh	0.26237	0.06884	0.20467	0.46	0.67823	0.26305	0.06919	−0.08956	0.20900	0.06856	0.26183
Klia Sepang	0.28656	0.08212	0.21730	0.24	0.48990	0.22497	0.05061	0.08669	0.17615	0.18701	0.43245
Kota Bharu	0.41591	0.17298	0.31710	0.27	0.51962	0.47097	0.22181	0.27649	0.32407	0.33423	0.57812
Kota Kinabalu	0.23620	0.05579	0.17693	0.17	0.41231	0.34455	0.11871	0.01579	0.27053	0.02065	0.14369
Kuantan	0.28057	0.07872	0.21675	0.43	0.65574	0.23808	0.05668	0.46292	0.19209	0.46766	0.68386
Kuching	0.23272	0.05416	0.18114	0.57	0.75498	0.22585	0.05101	−0.34476	0.17853	0.06427	0.25351
Lubok Merbau	0.22991	0.05286	0.18310	0.35	0.59161	0.20661	0.04269	0.05581	0.16208	0.19800	0.44498
Muadzam Shah	0.20753	0.04307	0.15075	0.51	0.71414	0.18144	0.03292	0.26957	0.13498	0.38115	0.61737
Pulau Langkawi	0.33927	0.11510	0.22939	0.63	0.79373	0.31563	0.09962	0.49394	0.24572	0.56585	0.75223
Sitiawan	0.21592	0.04662	0.17060	0.48	0.69282	0.20663	0.04270	−0.08907	0.15939	0.14331	0.37856
Subang	0.25790	0.06651	0.19433	0.26	0.50990	0.22872	0.05231	0.06862	0.17862	0.10155	0.31868

Note: RMSE = root mean squared error; MSE = mean squared error; MAE = mean absolute error; R = correlation coefficient; NSE = Nash–Sutcliffe efficiency; SVR = support vector regression; BTs = bagged regression trees; GPR = Gaussian process regression.

we mentioned previously that only Kota Kinabalu station showed unacceptable prediction, so when applying the BTs model, the high error in prediction was obtained with the lower value of R^2. Moreover, BTs at Muadzam Shah station has the lowest error or, more precisely, the best value of MSE, at 0.0293.

Based on Table 5, the GPR model provides the best outcomes compared with BTs and SVR in predicting minimum wind speed values. So, among all stations, Pulau Langkawi displays the vest performance, with good values of R^2 in training and testing, of 0.63 and 0.57, respectively. The optimum value of RMSE were found at Muadzam Shah station, with 0.20 and 0.13, respectively, in the training and testing procedures.

The main important point to take from Table 5 is that the performance evaluation in the training phase showed more accurate results than the performance evaluation in the testing phase.

However, we can see from all of the above results that the GPR model can provide a good modeling method and show better performance than SVR and BTs in (minimum, mean and maximum) wind speed prediction. So, to better understand the relationship between actual measurement data and predicted values for each station plot, the observed values were drawn for each station. The daily wind speed varies rapidly from day to day, so even though the prediction of GPR for most of the stations showed good agreement with the observed values,

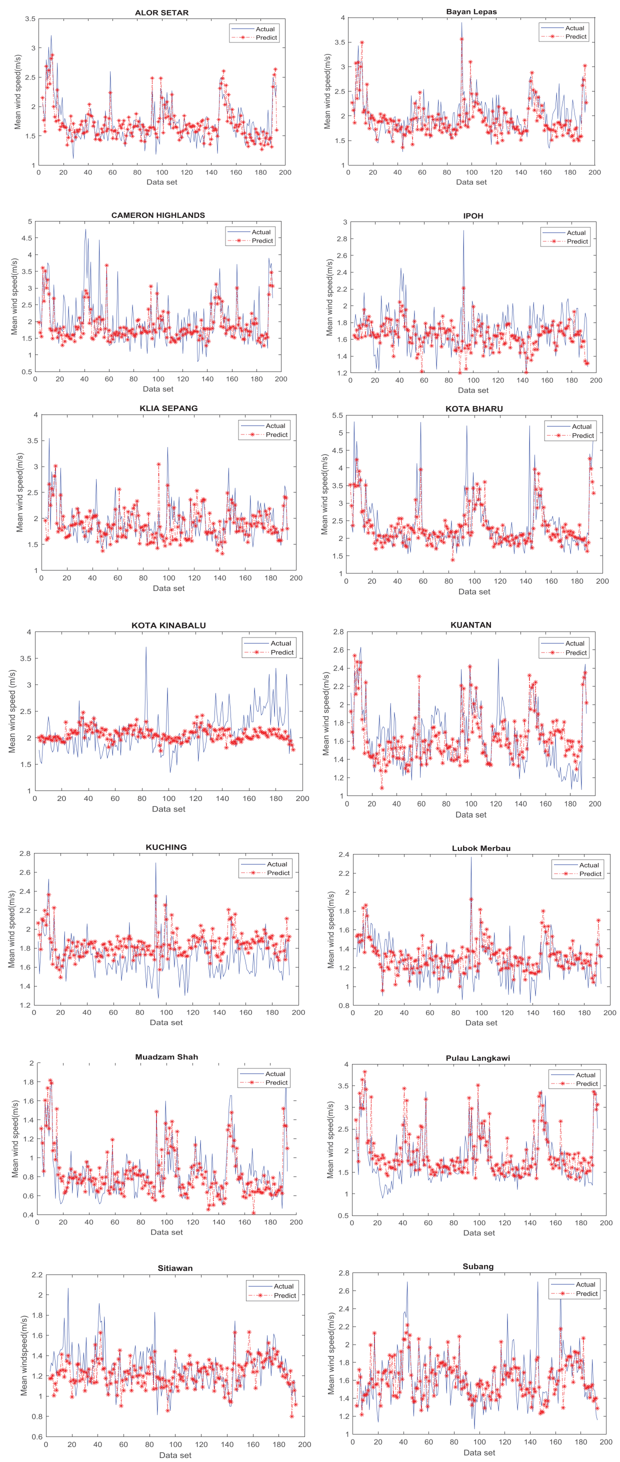

Figure 5. Predicted and actual in testing procedure using the Gaussian process regression (GPR) model.

the model at only one station, i.e. Kota Kinabalu, did not show good prediction performance. In other words, it can be seen from Figure 5 that the outcomes of the predictions have a close agreement with the corresponding measurements for all 13 stations.

4.2. Comparison models

Finally, after the wind speed had been predicted based on the three regression models, and performed in three stages, as summarized in Tables 3–5, to further demonstrate the prediction performance of the training procedure using the three models, the PRMSE, PMSE and PMAE were adopted for analysis, as given in Tables 6–8.

The improvement percentages of the comparison models using the GPR model in predicting maximum wind speed values are shown in Table 6. As mentioned earlier, the only unsatisfactory prediction was obtained at Kota Bharu station. Otherwise, the GPR model has acceptable prediction accuracy at all stations, as it could

Table 6. Improvement percentages for comparing models with the Gaussian process regression (GPR) model at maximum wind speed prediction.

	GPR over SVR			GPR over BTs		
Station	PRMSE	PMSE	PMAE	PRMSE	PMSE	PMAE
Alor Setar	9	17	5	5	9	2
Bayan Lepas	7	13	4	5	10	4
Cameron Highlands	7	14	6	5	10	5
Ipoh	3	6	2	4	8	3
Klia Sepang	1	2	0	3	7	4
Kota Kinabalu	5	9	4	1	2	0
Kota Bharu	2	4	0	2	3	1
Kuantan	5	12	4	2	6	1
Kuching	1	2	−2	−2	−5	−5
Lubok Merbau	7	14	4	5	10	4
Muadzam Shah	10	19	−2	12	22	5
Pulau Langkawi	5	10	−1	4	8	3
Sitiawan	2	4	1	3	7	2
Subang	1	3	0	1	2	1

Note: SVR = support vector regression; BTs = bagged regression trees; PRMSE = improvement percentage of root mean squared error; PMSE = improvement percentage of mean squared error; PMAE = improvement percentage of mean absolute error.

Table 7. Improvement percentages for comparing models with the Gaussian process regression (GPR) model at mean wind speed prediction.

	GPR over SVR			GPR over BTs		
Station	PRMSE	PMSE	PMAE	PRMSE	PMSE	PMAE
Alor Setar	16	30	10	11	21	7
Bayan Lepas	12	23	6	10	19	10
Cameron Highlands	6	12	4	10	19	10
Ipoh	3	6	1	8	15	4
Klia Sepang	3	5	2	4	8	5
Kota Kinabalu	3	6	1	5	10	5
Kota Bharu	2	5	1	2	3	1
Kuantan	6	11	4	8	15	8
Kuching	5	11	3	1	2	0
Lubok Merbau	6	12	3	9	17	8
Muadzam Shah	13	24	6	8	16	9
Pulau Langkawi	7	13	4	8	16	8
Sitiawan	2	4	1	9	17	9
Subang	2	4	1	6	12	7

Note: SVR = support vector regression; BTs = bagged regression trees; PRMSE = improvement percentage of root mean squared error; PMSE = improvement percentage of mean squared error; PMAE = improvement percentage of mean absolute error.

Table 8. Improvement percentages for comparing models with the Gaussian process regression (GPR) model at minimum wind speed prediction.

	GPR over SVR			GPRover BTs		
Station	PRMSE	PMSE	PMAE	PRMSE	PMSE	PMAE
Alor Setar	8	15	6	3	5	1
Bayan Lepas	3	6	2	2	4	2
Cameron Highlands	3	6	0	1	2	−1
Ipoh	2	4	2	1	2	2
Klia Sepang	1	2	2	1	2	1
Kota Kinabalu	3	7	2	1	3	0
Kota Bharu	0	1	2	3	5	3
Kuantan	2	4	2	2	4	4
Kuching	2	5	1	−1	−3	−3
Lubok Merbau	2	4	1	5	9	4
Muadzam Shah	7	13	2	0	1	1
Pulau Langkawi	6	11	4	4	7	2
Sitiawan	3	5	0	5	9	4
Subang	1	2	0	4	7	3

Note: SVR = support vector regression; BTs = bagged regression trees; PRMSE = improvement percentage of root mean squared error; PMSE = improvement percentage of mean squared error; PMAE = improvement percentage of mean absolute error.

gain more accurate results than the SVR model and BTs model in predicting the maximum weekly wind speed. The outcomes of the GPR model are better than those of the SVR approach. For example, compared to the SVM approach, the RMSE and MSE of those models are reduced at all stations (Table 6). For the MAE of prediction results for Kuching, Muadzam Shah and Pulau Langkawi, the SVR model increased by 2%, 2% and 1%, respectively.

On the other hand, for the BTs model, the RMSEs are greatly reduced by 22% at Muadzam Shah station. At Kuching station, the RMSE, MSE and MAE for the BTs model are increased by 2%, 5% and 5%, respectively. However, for all other 13 stations, the prediction RMSE, MSE and MAE are reduced for the proposed model BTs.

Regarding improvement in predicting the mean weekly wind speed, it can see from Table 7 that the prediction outcomes of the GPR method are better than those of SVR and BTs at all prediction stations. This phenomenon

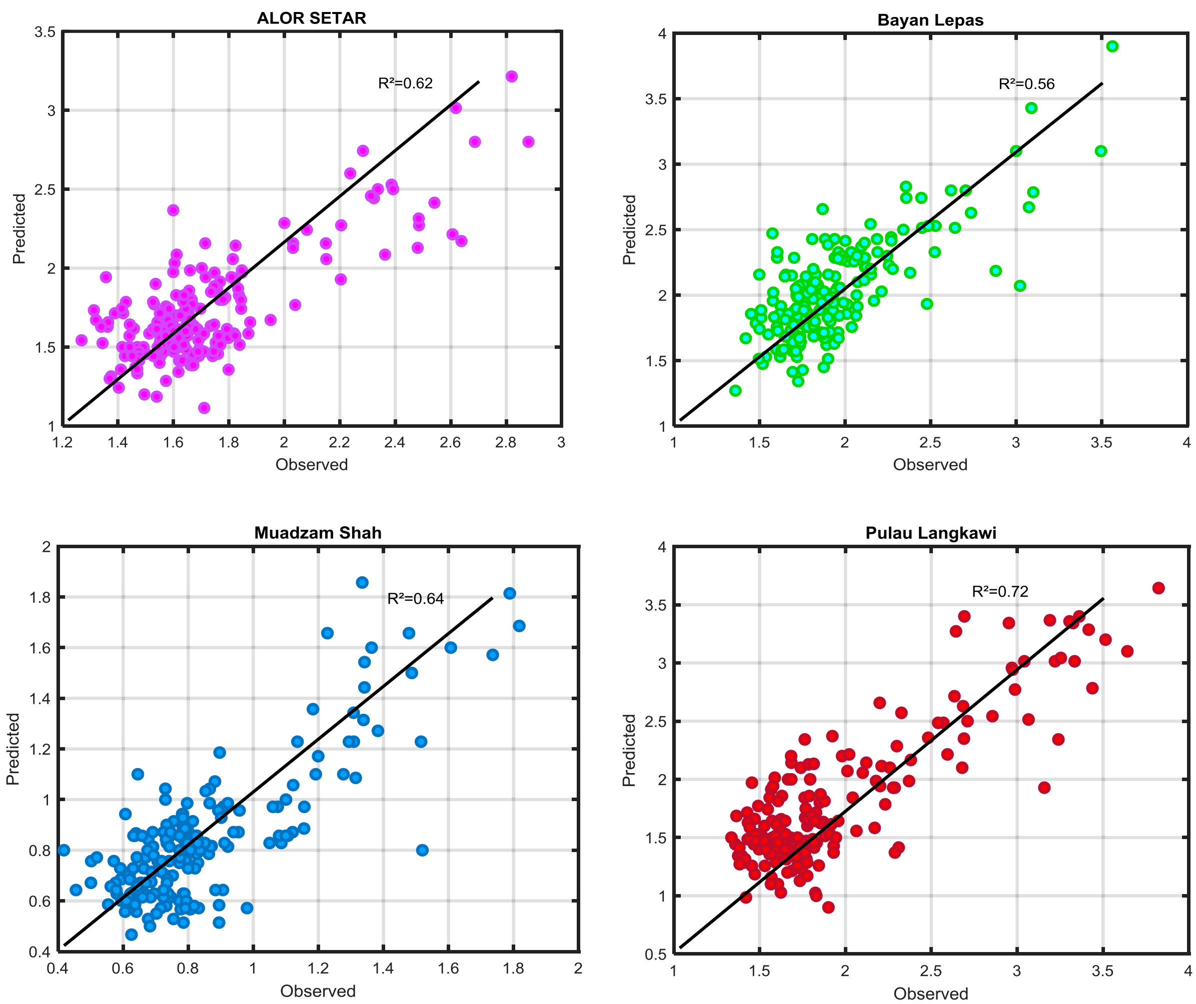

Figure 6. Scatter diagram of the predicted and measured values of wind speeds.

implies that GPR components could improve the prediction performance of the proposed method. For instance, it can be seen that the high MSE was reduced at Alor Setar, with 30% and 21% of the proposed SVR and BTs models, respectively.

Moreover, the prediction findings of the GPR approach are better than those of the SVR approach in predicting minimum weekly wind speed at all stations, as shown in Table 8. For instance, the MSE at Alor Setar station reduced by 15% compared with the proposed model SVR. Compared to the BTs model, only the prediction results of Kuching station have slightly increased, by 1%, 3% and 3% of the proposed model BTs, for RMSE, MSE and MAE, respectively.

Figure 6 shows the results of scatter diagrams for the best model, which is the GPR model at four selected stations. At the end of this research, we may conclude that this approach seems to be a promising tool for predicting wind speed. It could help management and control in the renewable energy field by allowing managers to choose an appropriate station to predict missing values of wind speed for the target station, based on the relationship between the target station and reference station. The benefit of this method is that if the requisite wind speed data for reference stations can be obtained, the future wind speeds of the target station could be predicted directly and acceptably without using any topographical details or further meteorological data.

5. Conclusion

As a crucial issue in the wind energy industry, wind speed prediction plays an essential role in optimum planning and wind energy production and conversion management. In this study, wind speed data, arranged into weekly minimum, mean and maximum values from 13 different stations, were applied as inputs, whereas the wind speed of the target station was applied as the output in three machine learning models. Even though prediction at Kota Kinabalu station showed an unsatisfactory level of accuracy using minimum and maximum wind speed values, this prediction model has been successfully employed at 14 measuring stations located in Malaysia using mean weekly wind speed data. Based on a five-fold cross-validation testing design, the GPR model performed better than the BT and SVR models. However, although the proposed model could provide good prediction accuracy, the coefficient of determination (R) was relatively low. Therefore, it is recommended to integrate the model proposed in this study with a preprocessing method or an optimization algorithm to improve the correlation between the predicted values and the actual values. In addition, future work could apply novel types of machine learning algorithms that are more appropriate for time-series prediction. Furthermore, implementing other data-driven techniques for predicting wind speed at stations can be useful.

Author contributions

Data curation: Yuk Feng Huang and Ali Najah Ahmed; formal analysis: Marwah Sattar Hanoon and Nur'atiah Zaini; methodology: Marwah Sattar Hanoon, Arif Razzaq, Mohsen Sherif and Ahmed Sefelnasr; writing – original draft: Marwah Sattar Hanoon, Ahmed El-Shafie, Nur'atiah Zaini, Yuk Feng Huang and Ali Najah Ahmed; writing – review and editing: Marwah Sattar Hanoon, Mohsen Sherif, Ahmed Sefelnasr, Pavitra Kumar, Kwok wing Chau and Yuk Feng Huang.

Acknowledgment

The authors would like to thank the Malaysian Meteorological Department (MMD) for providing the data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Data availability

The data available from the corresponding author on reasonable request.

ORCID

Ali Najah Ahmed http://orcid.org/0000-0002-5618-6663
Yuk Feng Huang http://orcid.org/0000-0001-6565-947X
Mohsen Sherif http://orcid.org/0000-0002-6368-8143
Ahmed Sefelnasr http://orcid.org/0000-0002-0281-6455
Kwok wing Chau http://orcid.org/0000-0001-6457-161X
Ahmed El-Shafie http://orcid.org/0000-0001-5018-8505

References

Ahmad, S., & Tahar, R. M. (2014). Selection of renewable energy sources for sustainable development of electricity generation system using analytic hierarchy process: A case of Malaysia. *Renewable Energy*, *63*, 458–466. https://doi.org/10.1016/j.renene.2013.10.001

Albani, A., & Ibrahim, M. Z. (2017). Wind energy potential and power law indexes assessment for selected near-coastal sites in Malaysia. *Energies*, *10*(3), 307. https://doi.org/10.3390/en10030307

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., & Panagos, P. (2019). Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma*, *355*(March), 113912. https://doi.org/10.1016/j.geoderma.2019.113912

Barhmi, S., & El Fatni, O. (2019). Hourly wind speed forecasting based on support vector machine and artificial neural networks. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *8*(3), 286–291. http://doi.org/10.11591/ijai.v8.i3.pp286-291. https://search.proquest.com/openview/85eca38ea47d5c9317c8b75479020dfe/1?pq-origsite=gscholar%cbl=1686339.

Başakın, E. E., Ekmekcioğlu, Ö, Çıtakoğlu, H., & Özger, M. (2022). A new insight to the wind speed forecasting: Robust multi-stage ensemble soft computing approach based on pre-processing uncertainty assessment. *Neural Computing and Applicications*, *34*(1), 783–812. https://doi.org/10.1007/s00521-021-06424-6

Bokde, N. D., Yaseen, Z. M., & Andersen, G. B. (2020). ForecastTB – An R package as a test-bench for time series forecasting-application of wind speed and solar radiation modeling. *Energies*, *13*(10), 2578. https://doi.org/10.3390/en13102578

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123–140. https://doi.org/10.1023/A:1018054314350.

Chen, J., Zeng, G. Q., Zhou, W., Du, W., & Lu, K. D. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy Conversion and Management*, *165*, 681–695. https://doi.org/10.1016/j.enconman.2018.03.098

Chen, K., & Yu, J. (2014). Short-term wind speed prediction using an unscented kalman filter based state-space support vector regression approach. *Applied Energy*, *113*, 690–705. https://doi.org/10.1016/j.apenergy.2013.08.025

Chen, M. R., Zeng, G. Q., Lu, K. D., & Weng, J. (2019). A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM. *IEEE Internet of Things Journal*, *6*(4), 6997–7010. https://doi.org/10.1109/JIOT.2019.2913176

Citakoglu, H. (2021). Comparison of multiple learning artificial intelligence models for estimation of long-term monthly temperatures in Turkey. *Arabian Journal of Geosciences*, *14*, 2131. https://doi.org/10.1007/s12517-021-08484-3

Citakoglu, H., & Aydemir, A. (2019). Determination of monthly wind speed of Kayseri region with gray estimation method. *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 415–420.

Harrou, F., Saidi, A., & Sun, Y. (2019). Wind power prediction using bootstrap aggregating trees approach to enabling sustainable wind power integration in a smart grid. *Energy Conversion and Management*, *201*(May), 112077. https://doi.org/10.1016/j.enconman.2019.112077

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media. https://doi.org/10.1007/978-0-387-84858-7

Hoolohan, V., Tomlin, A. S., & Cockerill, T. (2018). Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data. *Renewable Energy*, *126*, 1043–1054. https://doi.org/10.1016/j.renene.2018.04.019

Ibrahim, M. Z., Yong, K. H., Ismail, M., & Albani, A. (2014). Wind speed modeling for Malaysia. *International Journal of Renewable Energy Research*, *4*(4), 912–923. https://doi.org/10.20508/ijrer.36684

Inyurt, S., Hasanpour Kashani, M., & Sekertekin, A. (2020). Ionospheric TEC forecasting using Gaussian process regression (GPR) and multiple linear regression (MLR) in Turkey. *Astrophysics and Space Science*, *365*(6), 99. https://doi.org/10.1007/s10509-020-03817-2

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. Springer Science & Business Media. https://link.springer.com/book/10.1007%2F978-1-4614-7138-7.

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). Neuro-fuzzy and soft computing – A computational approach to learning and machine intelligence [book review]. *IEEE Transactions on Automatic Control*, *42*(10), 1482–1484. https://doi.org/10.1109/TAC.1997.633847

Jiang, P., Wang, Y., & Wang, J. (2017). Short-term wind speed forecasting using a hybrid model. *Energy*, *119*, 561–577. https://doi.org/10.1016/j.energy.2016.10.040

Karim, O. A., Zuhairuse, M. D., Nor Atikah, H., Siti Nurhidayah, A. M., Mohd Azhar, A. R., & Khairul Nizam, A. M. (2009). The development of hybrid integrated renewable energy system (wind and solar) for sustainable living at Perhentian. *European Journal of Social Sciences*, *9*(4), 557–563.

Khosravi, A., Machado, L., & Nunes, R. O. (2018). Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil. *Applied Energy*, *224*, 550–566. https://doi.org/10.1016/j.apenergy.2018.05.043

Liu, H., Yang, C., Huang, M., Wang, D., & Yoo, C. K. (2018). Modeling of subway indoor air quality using Gaussian process regression. *Journal of Hazardous Materials*, *359*(February), 266–273. https://doi.org/10.1016/j.jhazmat.2018.07.034

Md Abul Ehsan, B., Begum, F., Ilham, S. J., & Khan, R. S. (2019). Advanced wind speed prediction using convective weather variables through machine learning application. *Applied Computing and Geosciences*, *1*(September), 100002. https://doi.org/10.1016/j.acags.2019.100002

Mi, X., Liu, H., & Li, Y. (2019). Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. *Energy Conversion and Management*, *180*(November 2018), 196–205. https://doi.org/10.1016/j.enconman.2018.11.006

Mi, X. w., Liu, H., & Li, Y. f. (2017). Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. *Energy Conversion and Management*, *151*(September), 709–722. https://doi.org/10.1016/j.enconman.2017.09.034

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/10.13031/2013.23153

Najid, S. K., Zaharim, A., Razali, A. M., Zainol, M. S., Ibrahim, K., & Sopian, K. (2009). Analyzing the east coast Malaysia wind speed data. *International Journal of Energy and Environment*, *3*(2), 8.

Olatomiwa, L., Mekhilef, S., Shamshirband, S., Mohammadi, K., Petković, D., & Sudheer, C. (2015). A support vector machine-firefly algorithm-based model for global solar radiation prediction. *Solar Energy*, *115*, 632–644. https://doi.org/10.1016/j.solener.2015.03.015

Sain, S. R. (1996). The nature of statistical learning theory. *Technometrics*, *38*(4), 409–409. https://doi.org/10.1080/00401706.1996.10484565

Sarkar, R., Julai, S., Hossain, S., Chong, W. T., & Rahman, M. (2019). A comparative study of activation functions of NAR and NARX neural network for long-term wind speed forecasting in Malaysia. *Mathematical Problems in Engineering*, *2019*, 16403081. https://doi.org/10.1155/2019/6403081

Shabani, S., Samadianfard, S., Sattari, M. T., Mosavi, A., Shamshirband, S., Kmet, T., & Várkonyi-Kóczy, A. R. (2020). Modeling pan evaporation using Gaussian process regression k-nearest neighbors random forest and support vector machines; comparative analysis. *Atmosphere*, *11*(1), 66. https://doi.org/10.3390/ATMOS11010066

Shafiullah, G. M. (2016). Hybrid renewable energy integration (HREI) system for subtropical climate in central Queensland, Australia. *Renewable Energy*, *96*, 1034–1053. https://doi.org/10.1016/j.renene.2016.04.101

Shafiullah, G. M., Oo, A. M. T., Shawkat Ali, A. B. M., & Wolfs, P. (2013). Potential challenges of integrating large-scale wind energy into the power grid – A review. *Renewable and Sustainable Energy Reviews*, *20*, 306–321. https://doi.org/10.1016/j.rser.2012.11.057

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Srivastava, A. K., Singh, D., & Pandey, A. S. (2020). Short term load forecasting using regression trees: Random forest, bagging and M5P. *E3S Web of Conferences*, *4*(2), 15–21. http://dx.doi.org/10.30534/ijatcse/2020/152922020

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. *Handbook of Statistics*, *24*(04), 303–329. https://doi.org/10.1016/S0169-7161(04)24011-1

Wang, H., Zhang, Y. M., Mao, J. X., & Wan, H. P. (2020). A probabilistic approach for short-term prediction of wind gust speed using ensemble learning. *Journal of Wind Engineering and Industrial Aerodynamics*, *202*(April), 104198. https://doi.org/10.1016/j.jweia.2020.104198

Wang, J., Heng, J., Xiao, L., & Wang, C. (2017). Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. *Energy*, *125*, 591–613. https://doi.org/10.1016/j.energy.2017.02.150

Wang, J., & Hu, J. (2015). A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (autoregressive integrated moving average), ELM (extreme learning machine), SVM (support vector machine) and LSSVM (least square SVM) forecasts using a GPR (Gaussian process regression) model. *Energy*, *93*, 41–56. https://doi.org/10.1016/j.energy.2015.08.045

Wang, L., Li, X., & Bai, Y. (2018). Short-term wind speed prediction using an extreme learning machine model with error correction. *Energy Conversion and Management*, *162*(July 2017), 239–250. https://doi.org/10.1016/j.enconman.2018.02.015

Xiao, L., Shao, W., Yu, M., Ma, J., & Jin, C. (2017). Research and application of a combined model based on multi-objective optimization for electrical load forecasting. *Energy*, *119*, 1057–1074. https://doi.org/10.1016/j.energy.2016.11.035

Xiao, L., Wang, J., Dong, Y., & Wu, J. (2015). Combined forecasting models for wind energy forecasting: A case study in China. *Renewable and Sustainable Energy Reviews*, *44*, 271–288. https://doi.org/10.1016/j.rser.2014.12.012

Yin, Z., & Hou, J. (2016). Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. *Neurocomputing*, *174*, 643–650. https://doi.org/10.1016/j.neucom.2015.09.081

Zhang, C., Wei, H., Zhao, X., Liu, T., & Zhang, K. (2016). A Gaussian process regression based hybrid approach for short-term wind speed prediction. *Energy Conversion and Management*, *126*, 1084–1092. https://doi.org/10.1016/j.enconman.2016.08.086