Journal of Computers Vol. 35 No. 6, December 2024, pp. 1-14
doi: 10.53106/199115992024123506001

A Modular Data Analysis for Ocean Surface Wind Prediction using Artificial Intelligence-based Semantic Verification

Yu-Ge Liu, Bin-Feng Tang, and Ying Huang*

Liuzhou Railway Vocational Technical College, China
{rico-penny, tangbf}@ltzy.edu.cn, huangying800816@163.com

Received 30 June 2023; Revised 11 July 2023; Accepted 11 July 2023

Abstract. Predicting ocean surface winds aids in forecasting precise weather conditions and tides for residential and commercial purposes. A Multi-Modular Semantic Data Analysis (M2SDA) method is proposed to address the missing data errors across different accumulation regions to improve the forecast precision. Large volumes of data from multiple oceanic regions are gathered by sensors deployed across the ocean bed and buoy sense. Artificial intelligence-based analysis was used to identify missing data errors. Considering that the time factor is confined, the forecast endurance based on sensing and aggregation time factors is considered in the identification process, which is required for preventing breaks in data analysis. The M2SDA's performance is validated using precision, error, analysis time, identification ratio, and analysis rate. Experimental results showed that the suggested M2SDA enhances precision, identification, and analysis ratio of 9.16%, 9.9%, and 8.19%. Error and analysis time are decreased by 8.75% and 10.63%.

Keywords: data analysis, regression learning, semantics verification, wind prediction

1 Introduction

The task of wind prediction in an algorithm for forecasting is challenging. Severe and damaging wind prediction is a very difficult task to forecast. Wind prediction accuracy rate differs due to changes in climate and surface [1]. Wind prediction for the ocean surface plays a major role in ensuring users' safety during the sailing period. Ocean wind prediction is based on a certain set of instruments and measurements. Identifying wind direction and speed are the two most important tasks to perform in the wind prediction process [2]. Wind speed and direction over the ocean surface impact the weather forecasting system. Remote sensing technology is mostly used for the wind perdition process on ocean surfaces [3]. Wireless sensors and radars provide appropriate information for the prediction process. Both upper and lower-level winds are identified via signals produced by radar scans [4]. Numerical weather prediction (NWP) The practice of predicting the wind is most frequently utilized with models. NWP first addressed coastal regions and surfaces to collect necessary information regarding winds. Both sharp and smooth changes in winds are classified based on a certain set of features and functions. The weather prediction method's performance is improved due to NWP's increased prediction accuracy rate [5].

An examination of data acquired or communicated from sensors and Internet of Things (IoT)-capable devices is known as the analysis of sensor data. This analytical procedure identifies important data, which is then provided in a database for additional processing and use [6]. Sensor data analysis is an essential component of wind prediction since it identifies the pertinent information required for precise forecasting. Additionally, sensor data classification and identification yields crucial information that makes wind prediction easier [7]. The study of sensor data for wind prediction uses ML methods and techniques. These methods are used to pinpoint sensor data's specific significance and content, providing pertinent information for additional processing. The efficiency and efficacy of the system are increased by utilizing ML approaches, which also greatly increase the wind prediction process' overall accuracy rate [8, 9]. The sensor data analysis process analyses the quality and feasibility level of information that is presented in the database. A weather forecasting system gathers the information collected during the purpose of data evaluation procedure via a wireless sensor network (WSN). Wireless sensors capture exacts data of winds over the ocean surface This lowers the wind forecasting process' delay rate [10, 11].

Artificial intelligence (AI) technology is a subset of machine learning (ML) techniques that utilize human

* Corresponding Author

intelligence to perform a certain task in computers and machines. AI technique improves the effectiveness and efficiency of an application. AI is widely used for the wind prediction process over the ocean surface [12, 13]. The Long Short-Term Memory (LSTM) model is mostly employed in AI-based wind prediction process. LSTM identifies the important set of data that are presented in the historical wind dataset. LSTM trains the datasets that provide the necessary set of information for wind prediction [14]. In the wind forecast procedure, an artificial neural network (ANN) method is also employed. ANN technique calculates datasets that are presented in the database. ANN first identifies the data that is required for the wind prediction process [15, 16]. ANN improves the system's effectiveness and productivity by raising the general rate of accuracy in the wind forecasting procedure. ANN improves both performance and also feasibility rate of wind prediction process. A genetic algorithm (GA) is also used for over surface wind prediction process. Also introduced Deep Learning (DL) with probabilistic prediction for wind power plants speed and direction failure prediction. Using prior weather data as feed could help forecasts of power production since it allows the model used for prediction to take into account inaccuracies in forecasting the weather [17]. Also, Deep attention convolutional recurrent network [18] offered more precise short-term WSP. Initially designed to extract hidden representations to effectively capture spatial-temporal knowledge among wind speeds observed across the wind farm.

The development of an auto-updated memory module follows, which reconstructs latent representations from previous data. The reconstructed latent representations are grouped into K patterns using a K-shape clustering technique. The final prediction layer is created to produce the WSP outcome for latent representations that are consistently assigned to one of the K patterns out of K. Hence [19] provided a hybridized framework WSF for a moment based on meta-learning. The meta-learning and individual predictor components comprise the collective forecasting system. Three previously trained distinct predictors centered on the multi-output neural network featuring backpropagation with many concealed layers, the LSTM-RNN, and the recurrent gated units, make up the separate prediction section built using a multi-feed. One can estimate the direction of the wind value that needs to be expected by balancing the combined values of the hybridized units that depend on past wind speed data. Multimodal DL [20] to recover the wind speed after the variation of the information provided to get around the difficulty of hand-engineered feature integration and successfully do so. The results showed improvement in prediction accuracy, with error prediction metrics, and demonstrated that accuracy is closely correlated with sample size but unrelated to wind speed. Bidirectional long short-term memory (Bi-LSTM) [21] predicted the short-term typhoon wind speed accurately by taking into account both the real-world model and the artificial neural network (ANN) model. The bi-LSTM model is optimized using the particle swarm optimization (PSO) technique and then employs a variation mode decomposition (VMD) approach for breaking speed downwind. The findings demonstrate that the developed model can predict typhoon wind speed and has great tolerance for making uncertainty predictions, showing good results for MSE, RMSE, and R2. The approach's practical applicability may be impacted by the approach's computational requirements for scaling up. In [22], the multiple sizes pattern adaptive retrieval hybridized method generates wind speed estimations. The suggested approach builds six networks with various convolution operator lengths and collects. It exercises the deep autocorrelation pattern at various point levels hidden in data with great resolution. A versatile cuckoo-search-moth-flame fusion optimized method is then used to aggregate the forecasting results. To rectify errors, the model employs a multifaceted error regression technique. [23] provided a multiple CNN architecture for temporal wind nature assessments, multiple LSTM, tightly connected convolutional layers, and numerous input features. Multiple features of Densely Connected Convolutional Neural Network with Multifaceted LSTM Architecture is the name of the designed architecture. The input planes of the MCLT contain 58 features created using values for wind speed and direction. Past supervisory management, as well as acquiring information about the system collected by a lot of turbines (WTs) across multiple windmills, and Wasserstein distance-based adversarial training is subsequently used for estimating the average wind speed likelihood density function associated with the focus on wind generator at the subsequent date and time [24]. An enhanced heterogeneity density network is given to preserve better the variability structure seen in previous wind velocity series and to figure out the parameters utilised in a probabilistic heterogeneity approach to estimate the wind's speed. The wind speed intervals estimation model is proposed using a quasi RNN in [25] and the lower upper bound estimate method. Because of Meta-heuristic efficiency approaches with a non-differentiable function of loss. must be used to train the model. An additional variance that contains two target functions and enables regular stochastic-based gradient descent for network training. The recommended method produced short intervals with high coverage, according to the results of the computer experiments, which enabled it to enhance the coverage width criterion by 33.2%. Sensitive in choosing hyperparameters such as training rates, number of batches, or network designs. AR and SVM based on the hybrid method [26] used the wavelet decomposition approach, the wind speed time series are divided into several frequency components, and each frequency component is modeled separately. Even though wind speed oscillations exhibit various dynamics over a range of time scales, effective modeling techniques can greatly reduce the error in the prediction. Attention-Based Graph Networks and Deep Learning [27] have been used for wind speed prediction, which increased the forecasting method's effectiveness and dependability by increasing prediction accuracy. According to the results, the model greatly outperforms a multilayer perceptron and a BiLSTM model, while a simple GNN model is performed on par. Furthermore, the suggested graph attention architecture is easily adaptable to various applications by providing flexibility in the preferred attention operations, which may vary depending on the particular application. Particularly, it appeared that the attention networks realized turbine connections consistent with some physical sense regarding wake losses.

Different factors influence the accuracy of predicting ocean surface winds. A Multi-Modular Semantic Data Analysis (M2SDA) method is proposed to address the missing data errors across different accumulation regions to improve the forecast precision. The primary contribution of this study is the proposal and evaluation of the M2SDA method, which identifies and addresses missing data issues in ocean surface wind forecasting by combining artificial intelligence-based analysis and semantics methodologies. This system provides a fresh and creative way to tackle a significant problem in weather forecasting. The objective is to increase the precision of predicting weather and tides for household and commercial uses. The main objective of this study involves:

- Assemble massive amounts of data from many oceanic locations utilizing buoy sensing and sensors placed all over the ocean surface.
- Use semantic analysis methods to find missing data errors utilizing artificial intelligence-based analysis.
- To avoid pauses in data analysis, consider estimating durability based on sensing and aggregating temporal components.
- Analyze the M2SDA method's performance in comparison to other methods by using objective measures like precision, error, analysis of period identification proportions, and analysis ratio.

The manuscript is organized into sections in the manner described below: Section 2 discusses the suggested data analysis procedure. Section 3 analyzes the performance evaluation of the proposed method, and finally, a summary of the study's work is given in Section 4.

2 Proposed Data Analysis Method

Ocean surface wind represents the magnitude and direction present in the ocean, based on which the forecast of the ocean surface wind is modeled with the sensors placed at the bed and buoys in the oceans. The winds in the oceans are considered the main factor in the study of oceans, which identifies the lifestyle of the marine environment. The ecosystems protracting sea currents and waves and the sediments of materials on the sea floor are used to analyze the different stages of erosion procedure at the seafloor. The ocean surface winds play a vital role in determining the momentum flux and energy of the winds, including the stresses caused by waves and the circulation of oceans. It provides heating and cooling effects and increases the salinity of the ocean with an exchange of gases between the atmosphere and oceans. A consistent dataset with higher resolution within the specific intervals is needed to analyze and predict the forecast of the wind conditions. The dataset aggregated from the sensor must also address the fluctuations representing climatic conditions such as the el-Nino effect. Fig. 1 depicts the suggested M2SDA process.

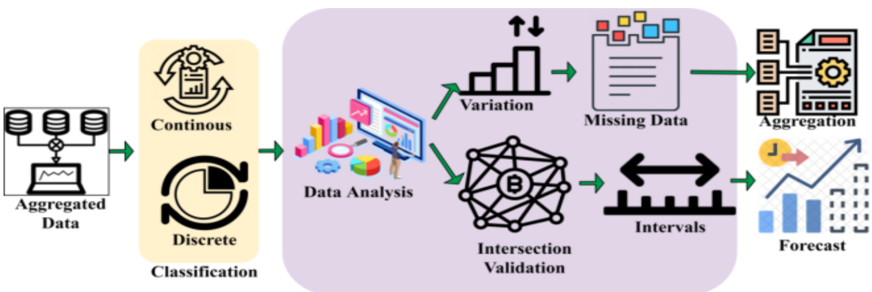

Fig. 1. Implementation of the M2SDA process

A sensor at the buoys for monitoring the ocean surface wind is capable of multi-tasking operations. It requires different configurations for periodic sensing and data aggregation with sufficient energy. The hardware and mechanical design must be fault tolerant for the efficient monitoring of temperatures in the ocean resulting in analyzing the ocean surface winds. The data from these sensors are monitored for different parameters: wind direction, wind speed, wind adjustment, and wind components. From these sorts of data, the ocean surface winds are forecasted. The wind data from the buoys are extracted, which are of vector quantities. These data are divided based on the orthogonal components representing the wind direction, namely x direction and y direction. These directions are denoted as A and B. The surface winds and their components in the equatorial regions represent warm pool conditions which are tedious for monitoring weather and climatic conditions. A time interval of data is extracted, mentioning the wind burst, wave propagation, and wave storms. The observations from buoys sensors are sparsely distributed, and the coverage area is too small.

2.1 Aggregation Process

Let Av(pj, t) and Bu(pj, t) denote the wind direction components placed at the location {where j=1,…..,k} and observed for the time interval {t where t= 1,….., T}. The data extracted from the sensors' wind direction are variable for the spatial and time intervals denoted as mt. The vectorization for the wind direction components is represented as At and Bt. These vectorizations of wind components tend to be the combined list which is denoted as k+ mt, and the overall collection of vectors which is denoted as $\{A\}_r^q$ representing $\{A_t = q, \ldots, r\}$ at time intervals. The wind data observed are said to be independent of time $\{A\}_1^T$ and $\{B\}_1^T$ which denotes the true values of observations which are as shown in Eqn (1)

$$\{A\}_1^T \{B\}_1^T ;\theta_1 = A_t \mid \theta_1][B_t \mid \theta_1] \quad \textbf{(1)}$$

θ is the parameter used in data. The covariance of the matrices Σt is diagonal with unknown variances σ. The elements k which are diagonal to σ are equal to mt. For each interval t, H_t is represented as shown in Eqn (2)

$$H_t = (k + m_t) \times n \quad \textbf{(2)}$$

It represents the data observations for the location of the buoy's sensors. Based on the data, it is partitioned for mapping matrices which are denoted by the Eqn (3)

$$H_t = \left[H_a', H_s'(t)\right]' \quad \textbf{(3)}$$

The above representation of Eqn (3) denotes the matrices H_a' and $H_s'(t)$ for k × n and m_t × n set of observations. H_a' denotes the conditional means of the data i.e., from the observation location within the coverage distance ∝ . The weights are assigned based on the linearity as shown in Eqn (4)

$$W = {(\alpha - \alpha_i)} \big/ {w^*} \quad \textbf{(4)}$$

From the above Eqn (4) α_i is the distance between the observation location and the ocean surface and w^* are the weights assigned based on the linearity, $H_s'(t)]'$ is the matrix that simplifies the mean of the observation from the location of surface buoys sensors. The data representation for continuous and discrete sequences is presented in Fig. 2.

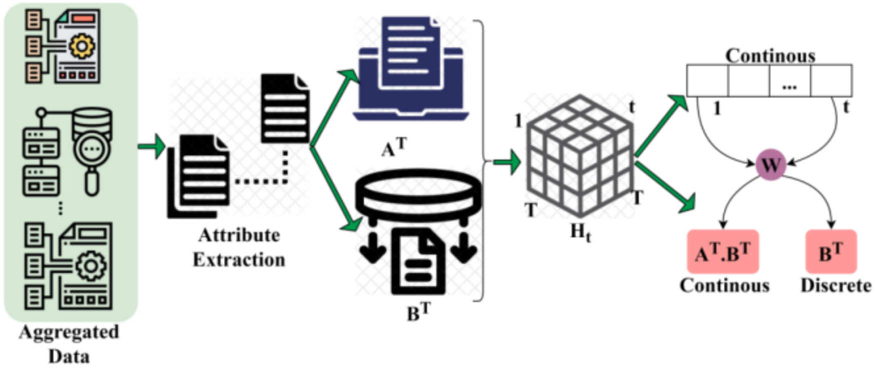

Fig. 2. Data representation for continuous and discrete sequences

The aggregation data is represented using the A^T and B^T for 1 to t and 1 to T ∀ such that the representation is valid. In this process, the continuous representations are split using A^T. B^T. If only B^T is available, then H_t representations are confined within (t, T). Therefore W is subtracted for extending α for which H(t) serves as the continuous representation (Refer to Fig. 2). The variability of the wind fields on the ocean surface is represented by considering the atmosphere depth, which is approximated as a thin fluid as shown in Eqn (5)

$$A_t = \gamma_A + A_t^F + \hat{A}_t \tag{5}$$

$$B_t = \gamma_B + B_t^F + \hat{B}_t \tag{6}$$

From the above Eqn (5) and (6), γ_A and γ_B are the spatial mean for the wind components, A_t^F and B_t^F are the wind component approximation related to the atmosphere depth, $\widehat{A}_t$ and $\widehat{B}_t$ denotes the motions of the sea waves. The symbols γ_A and γ_B are the spatial mean for the wind direction components described in terms of γ_A and γ_B with the aggregated information. The wind direction elements with γ_A as low, γ_B as high and $(\gamma_A . \gamma_B)$ as non-linear identification for conditional analysis in the linear regression basis is employed for identifying the classifications with three basic conditions as $\gamma_A > \gamma_B$, $\gamma_A < \gamma_B$ and $(\gamma_A = \gamma_B)$. To prevent the difference in the condition $(\gamma_A = \gamma_B)$ the following parameter A_t and B_t ∀ β variation is validated. Similarly, the representation for classification $\gamma_A . \gamma_B \in \{1 \text{ to } T\}$ is validated for further analysis of reducing mean square error.

To maintain the linearity of the equation by eliminating non-linear terms for momentum and simplification of the atmosphere depth, an approximation is related to satisfy the shallow linear equations on the equatorial plane in two-dimensional form. It is denoted as a (X, Y) plane based on which the equatorial mode on an orthogonal basis set is derived as shown in Eqn (7)

$$b^E(X,Y;t) = \sum_m b_m^E(X,Y;t) \tag{7}$$

The Eqn (7) represents the plane of the equatorial mode represented in a two-dimensional plane satisfying the momentum of the sea surface for the atmosphere depth. These data observations are used to analyze the dynamic nature of the atmosphere and the ocean surface. These observations of data, it is analyzed for the continuous set of data and discrete set of data. It is based on the sequence of data aggregated from the beds and the sensors' buoys for forecasting the ocean surface winds. It is analyzed for different conditions based on which the data is classified as continuous or discrete. Continuous data is the data that tends to replicate data observed at different time intervals. In contrast, discrete data is the data that identifies a set of missing variables when observed at different time intervals. Some of the conditions based on which the data is classified as continuous or discrete are identified as follows:

1. The observed data is analyzed for variations if the condition tends to be low. Then there is the least possibility for variation of observed data from buoys sensors.
2. If the condition tends to be high, then there is the highest possibility for the variation of observed data to forecast ocean surface winds.
3. If the variation of data happens to be low and high, then the observed data is considered to be the linear information data observed using multiple sensors employed to forecast the ocean surface winds.

From the observed data, a dataset is used for linear regression learning to identify data's intersection validation and forecast the ocean surface winds. The data processing based on the intersection of variables at different time intervals is analyzed. The conditional analysis is illustrated in Fig. 3.

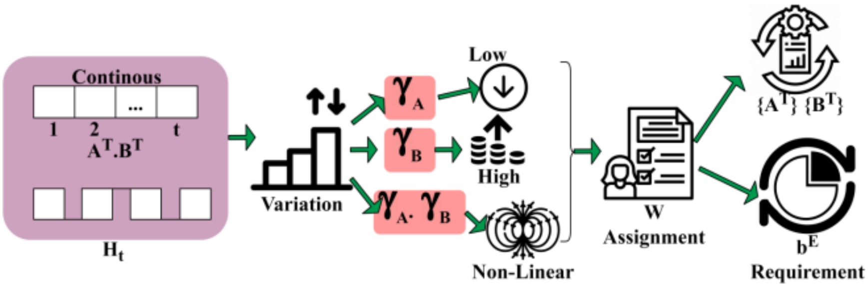

Fig. 3. Conditional analysis

The $H_t \forall A^T. B^T$ is used for analyzing the variation for which γ_A, γ_B and $(\gamma_A \cdot \gamma_B)$ are used for low, high, and non-linear identification. Based on ω assignment, $\{A^T\}.\{B^T\}$, and b^E requirements are classified. The above-represented conditions are balanced in detecting multiple variations for H_t. This is required for the linear analysis discussed in the following section.

2.2 Data Analysis

The input data at different time intervals are aggregated which is represented as ∂_A, ∂_B and the output is δ. The input of the data is considered as the vector which is denoted as $\partial_{A,B} = [\partial_{AB1}, \partial_{AB2}, \dots, \partial_{ABn}]$. These input data vectors belong to the respective location of observation. It is observed for time intervals which are represented as $T = [T_1, T_2, \dots, T_n]^d$. The functions of mapping the data at different time intervals T into patterns for the forecast of ocean surface winds are defined as shown in Eqn (8)

$$\partial_{AB,d} = \frac{T_{AB,d} - \underbrace{T_d}}{\sqrt{\sum_t^n (T_{AB,t} - \underbrace{T_d})^2}} \tag{8}$$

From the above Eqn (8), t = 1, 2, …, T is the time intervals at which the observation of data occurs with $T_{AB,d}$ is the time series of data with a mean load of $\underbrace{T_d}$ for the period of t. From Eqn (8), the vector Td is subtracted from the components and then it is divided by its length. Thus, the normalized vector ∂_{AB} is obtained. The data at different time intervals represent the same mean and variance. Thus, it maintains continuous data of observation from surface buoy sensors. The linearity in input to forecast data analysis is presented in Fig. 4.

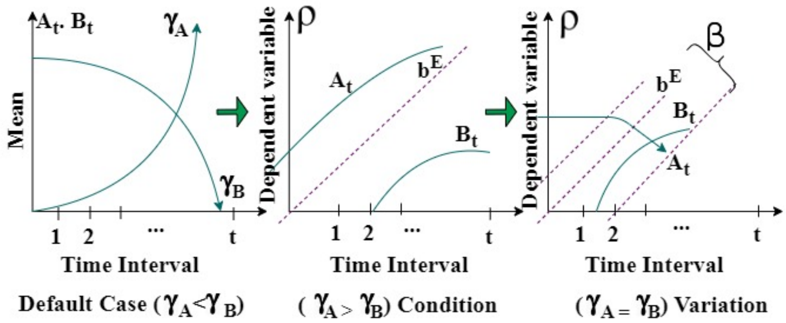

Fig. 4. Input to forecast data analysis

As shown in Fig. 4 the y axis 1, 2, …, t represents the time intervals and the A_t and B_t in the horizontal representation of default case represents the mean load used to observe the data points time series. In the second and third condition ($\gamma_A < \gamma_B$) and $\gamma_A = \gamma_B$ with respect to time interval the horizontal axis ρ represents the dependent variable represents the best line of fit for relationship between target and forecasted data.

The conventional to variation conditions are analyzed using $\partial_{AB,d} \forall \rho$. This ρ across distinct t requires A_t and $B_t \forall b^E$ separation. Post this separation, A_t and $B_t \forall \beta$ (variation) is estimated to preventing $\gamma_A = \gamma_B$ condition. Therefore, $\gamma_A > \gamma_B$ to ($\gamma_A < \gamma_B$) condition is validation is validated for linear analysis (Refer to Fig. 4). The variables in the vector ∂_{AB} represents the input data with O as the output variables of data which denotes the forecast of ρ. The output of the forecast vector with data is $\rho_d = [\rho_{d1}, \rho_{d2}, \rho_{d3}, \ldots, \rho_{dn}]^T$ which is as shown in Eqn (9)

$$\rho_{AB,d} = \frac{T_{\rho,d} - \underbrace{T_d}}{\sqrt{\sum_t^n (T_{AB,t} - \underbrace{T_d})^2}} \tag{9}$$

Eqn (9) forecasts the database on the input vector. From Eqn (8) and (9), the input data and the corresponding data forecast are unified according to the time intervals of data observations. From these unified data, the relationship between the input and the output is analyzed using linear regression learning by analyzing the neighborhood variables in the input data. It is used to find the intersection of variables in the data. It is being modeled based on the predictors, i.e., input data, to highlight the relationship of data. The best line of fit represents the mapping of target data to predicted data. i.e., input data to forecasted data is given by

$$\rho = \vartheta_0 + \vartheta_1 \partial_{AB} \tag{10}$$

From the above Eqn (10), the best line of fit denoting the relationship between the target input data and the predicted forecast data, which denotes the ocean surface winds, is obtained. This relationship denotes the positive linearity between the input and forecast data. The variable in Eqn (10) ρ represents the dependent variable and ∂_{AB} represents the independent input data, ϑ_0 is the intercept of the line and ϑ_1 is the coefficient of linear regression. The forecast process is illustrated in Fig. 5

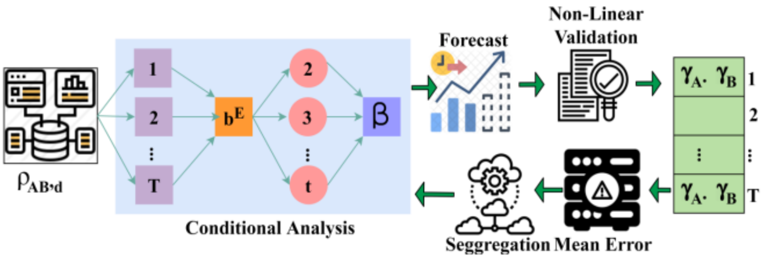

Fig. 5. Forecast process illustration

The $\rho_{AB,d}$ is used for b^E conditional analysis $\forall$ t $\in \beta$; this correlates with the three conditions discussed above. Based on the forecast, the non-linear validation $\forall\rho$ is performed. The identification of $\gamma_A.\gamma_B \in [1 \text{ to } T]$ from which the mean square error is further validated. The b^E (segregation) is analyzed further for H_t in the forecast process (Fig. 5). The cost function for the linear regression learning is calculated, which denotes the average of the squared errors between the target input data and the forecast of ocean surface wind. The mean square error function is represented as shown in Eqn (11)

$$\beta = \frac{1}{\varepsilon}\sum_{\pi=1}^{\varepsilon}\rho - (\varphi\partial_{AB} + \tau)^2 \tag{11}$$

Eqn (11), denotes the cost function used to find the accuracy of the mapping function used to map the input target data and the predicted forecast data. This mapping function is also considered a hypothesis function. This cost function parameter is determined by using a gradient descent method which minimizes the value of the cost function. This forecast is possible for continuous data from the sensors. To use the discrete data for the forecast of ocean surface winds, the data is again aggregated from the sensors which are again analyzed for their type of data. These aggregated continuous data obtained from the buoy sensors are used as input data for the linear regression learning algorithm. This further optimizes the performance of the proposed method by using the most available data for ocean surface wind forecast.

3 Performance Assessment

The section presents the discussion on wind prediction from [28]. The dataset contains 17 fields and 613392 entries for different time intervals. In this observation, wind speed, temperature, and degree are precisely used for predicting; the data is continuous and discrete. The intersection is observed for 24 sensings (observation) intervals and 60 min of aggregation. First, the data utilized for interval forecasting is represented in Fig. 6.

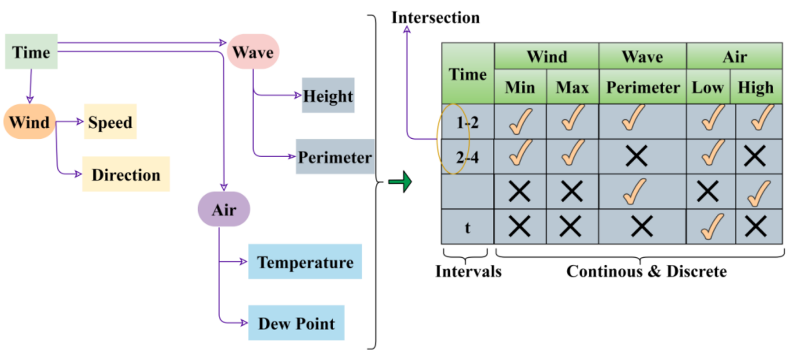

Fig. 6. Data utilization for continuous & discrete

The wind (speed, direction), wave (height, perimeter), and air (temperature, dew point) factors and their attributes across different intervals are used for the prediction process. In consecutive intervals, the intersections based on data availability are used for prediction. Discreteness is observed if the previous observations say high (air) or max (wind) does not fit the consecutive time. This representation is used for different utilization intervals for maximizing forecast. Following the above representation, the missing data (Variation) is estimated and presented in Fig. 7 for wind, wave, and Air attributes. Refer to Fig. 7 for calculating missing data (variation) represented in the x-axis concerning three components: wind, wave, and air. For wind, the speed and direction in terms of km/*h*r of the component are calculated based on the variations. For wave, the x-axis represents the height and perimeter as m about missing data variation. For the air component, the temperature with due point is measured as °C is measured concerning missing data (variation).

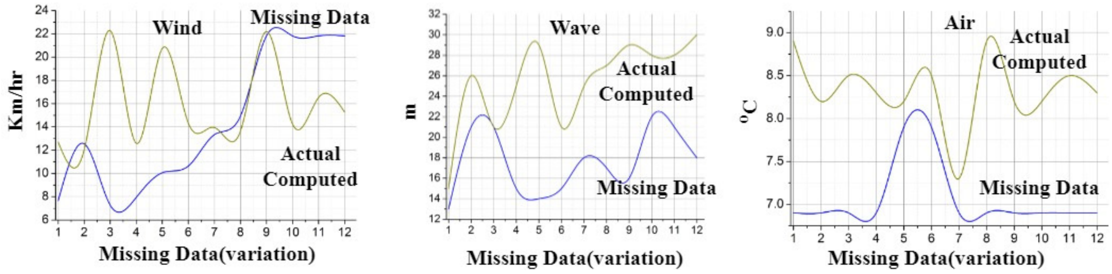

Fig. 7. Missing data estimation analysis

Fig. 7 analyzes the missing data for wind, wave, and air components. The missing data is estimated based on the three conditions illustrated in Fig. 7 $\forall\ \partial_{AB}$ such that $\gamma_A = \gamma_B$ is analyzed. In this analysis, the linearity interrupting features are used for validation. By utilizing the intersection in $T\{T_1, T_2, \ldots, T_n\}^d\ \forall\ t$ the H_t is extracted from Fig. 6 representation. The aim of $\{A^T\}.\{B^T\}$ from distinct intervals is estimated for leveraging the b^E for which missing data is computed. Therefore, the aggregation features are renewed such that the intersections are varied. From the missing data detection and aggregation recommendations, the classification demands are analyzed in Fig. 8.

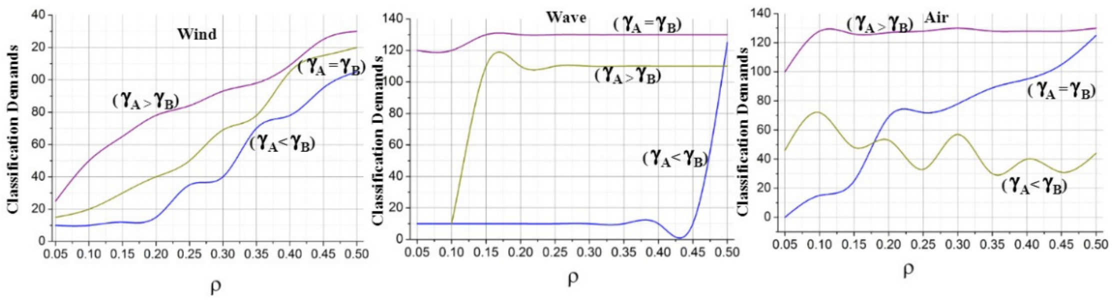

Fig. 8. Classification demands ∀ ρ

As shown in Fig. 8, the classification demands are represented in the x-axis for three categories wind, wave, and air, with y-axis representation as dependent variable ρ is analyzed based on three variant conditions.

The classification demands for wind, wave and air attributes are analyzed as in Fig. 8. The 3 conditions ($\gamma_A > \gamma_B$), ($\gamma_A = \gamma_B$) and ($\gamma_A < \gamma_B$) are used for identifying the classifications. The classifications include missing and intersection data across distinct ρ. As the ρ increases, classification increases. Based on the convenient variation, the ∂_{ABn} is validated for ($\gamma_A.\gamma_B$) maximization. The ω assignment is performed for stabilizing variation across distinct interactions. Based on this process, the H_t is reformed for b^E such that A_t. B_t is retained for which ρ is reduced. Therefore the classifications are increased for preventing mean errors.

This subsection provides a discussion on the proposed method's performance using comparative analysis. In the comparative analysis, precision, error, analysis time, identification ratio, and analysis rate are validated. The variations considered are sensing intervals between 2 and 24, and aggregation time from 5min to 60mins. For proving the proposed method's efficiency, a comparison with HMDL [20], CSWSF [13], and WSIP [29] is performed.

3.1 Precision

The proposed multi-modular semantics data analysis technique optimizes the precision by incorporating data from surface buoys and analyzing the data for its nature. The sensors are placed for the ocean wind direction increasing the coverage of the sensors. This increases the chances of aggregating the data under different circumstances. From this aggregation of data, this set of data is also used for the forecast of ocean surface winds which enhances the accuracy of the proposed methods. The conditional classification for the varying representations and H_t are used for confining the variations. In the variation suppression process, the data from distinct sensing intervals are used for forecasting. The forecasting process then relies on missing data and forecast sequences for improving the aggregation/ recommendation factor. Therefore, the conventional learning from the regression process requires identified features for which the classifications under discreteness are suppressed. From the suppressed intervals, the t and T based representations are identified for A^T and B^T illustrations. This is required for γ_A and γ_B balancing from the previous recommendation. The learning process is instigated from this point provided the condition fails. The linear non-regressive process identifies the intervals (intersection) for leveraging the forecast precision (Refer to Fig. 9).

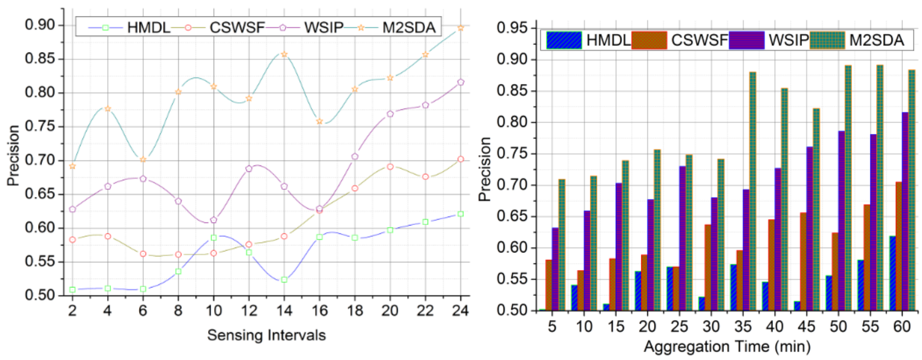

Fig. 9. Comparison of precision

3.2 Error

The proposed technique deals with the data based on the nature of data i.e., continuous or discrete type of data. Initially, the continuous set of data is used for a forecast which attains a state of the forecast using a linear regression learning algorithm. To further reduce the chance of errors in forecasting the ocean surface winds, the missing values from a discrete set of data are determined. To improvise the discrete type of data, aggregation of data from surface buoys is again initiated which further reduces the errors present within the data and increases the accuracy of the forecast of ocean surface winds. The error-causing condition from $\gamma_A = \gamma_A$ post the ρ and $\partial_{AB,d}$ is required for preventing precision drops. The identified variations $\in M_T$ and A^T. B^T in t is disintegrated for non-linearity analysis. Based on the ω assignment, the b^E the requirement is estimated. This estimation is used for monitoring $\gamma_A < \gamma_B \forall$ B (non-resulting) case. Therefore, the γ_A above the γ_B or vice Versa is induced for suppressing errors. Depending on the segregation error, $\gamma_A.\gamma_B \in$ 1 to T, the independent t is identified. Therefore, the error is conventionally reduced for the proposed method (Refer to Fig. 10).

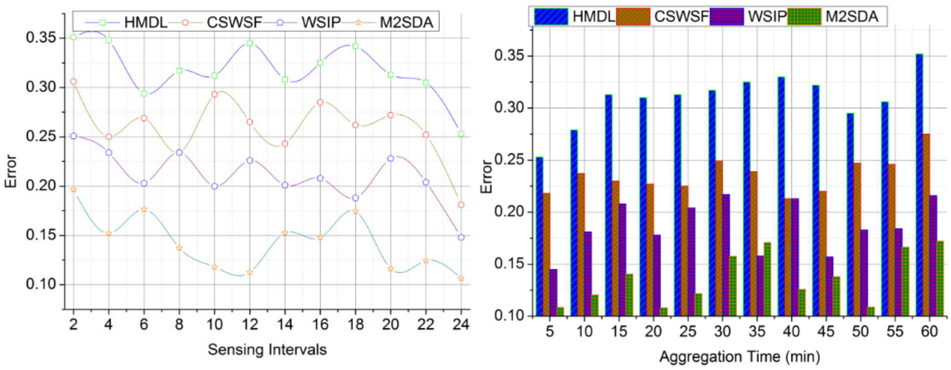

Fig. 10. Comparison of error

3.3 Analysis Time

The analysis time for the forecast of ocean surface wind with the proposed technique is less due to prior analysis of data. Prior study of the data entails reviewing past wind data gathered from buoys and sensors placed over the ocean floor. Data analysis is further enhanced by considering the average information from a specific observation point. The prior analysis of collected data can be observed using the conditional analysis approach. Then by offering a representative value that denotes the overall behavior of the wind data at that point, this addition helps shorten the analysis time. The optimal fit line is determined between the desired input data and the anticipated prediction data, representing the ocean surface winds. This relationship shows that the input information and the forecast data are positively linear. The analyzed data is used to train the linear regression learning algorithm which makes the forecast of ocean surface winds easier with less analysis time. This is achieved by assigning weights based on the linearity of data i.e., it corresponds to the distance between the observation location and the ocean surface. The analysis of data is further enhanced by the mean observation of data from the location of observation with the reduction in analysis time. The data analysis time increases if a non-linear require further aggregation. Therefore the analysis of previous variation and the current intersection is required. The $\rho_{A,B,d} \forall \partial_{ABd}$ is required for classifying discrete sequences from continuous ones. Therefore, the required classifications are suppressed for preventing additional b^E demands. This however confines the error within specific $t \in T$ for which mean errors are less. The conditional validation extends the above to the available $\gamma_A.\gamma_B \in T$ for which missing data is computed. The computation is suppressed for reducing the analysis time; the process is consistent throughout the sensing intervals and different observations (Fig. 11).

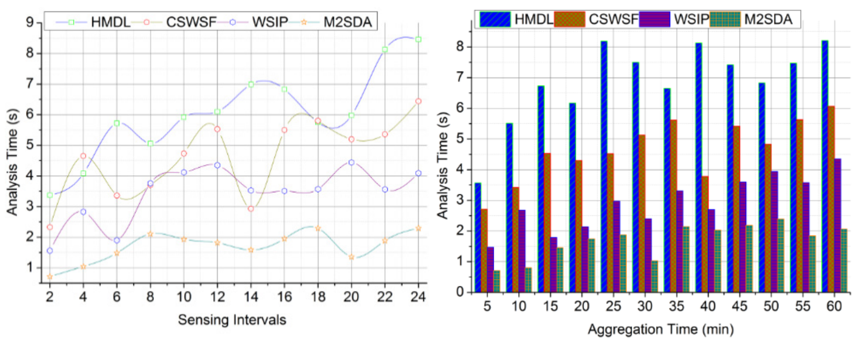

Fig. 11. Comparison of analysis time

3.4 Identification Ratio

The identification ratio of the proposed multi-modular semantics data analysis technique deals with the mapping of data observed at different locations and at different time intervals. This mapping of data between the input vectors and the forecast based on the data optimizes the identification ratio using a linear regression algorithm. In a linear regression learning algorithm, the neighborhood variables are analyzed based on the relationship of data. Thus, the proposed technique increases the identification ratio of the input data to forecast the ocean surface wind. The identification ratio increases by identifying precise intersections across different intervals. The existing intervals are validated using discrete classifications across multiple instances for which aggregation is induced. From this process, the consecutive instances are analyzed for the three conditions for identifying ρ. From the ρ, the missing information is located in past to present observation intervals. This proposed method relies on linear regression for preventing periodic ups and downs across the varying classifications. Therefore, for the varying sensing interval inputs and the aggregation time, the proposed method performs better for missing data detections (Refer to Fig. 12).

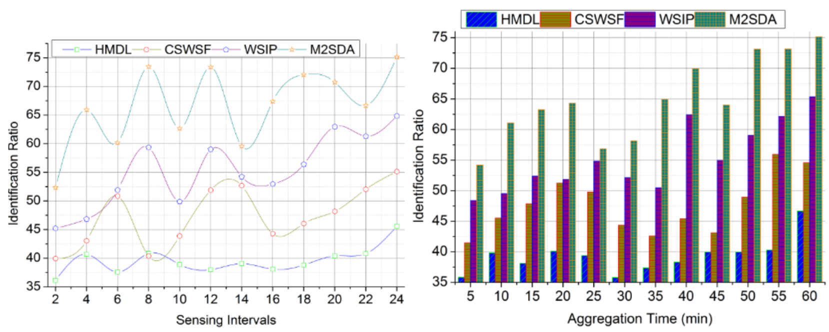

Fig. 12. Comparison of identification ratio

3.5 Analysis Rate

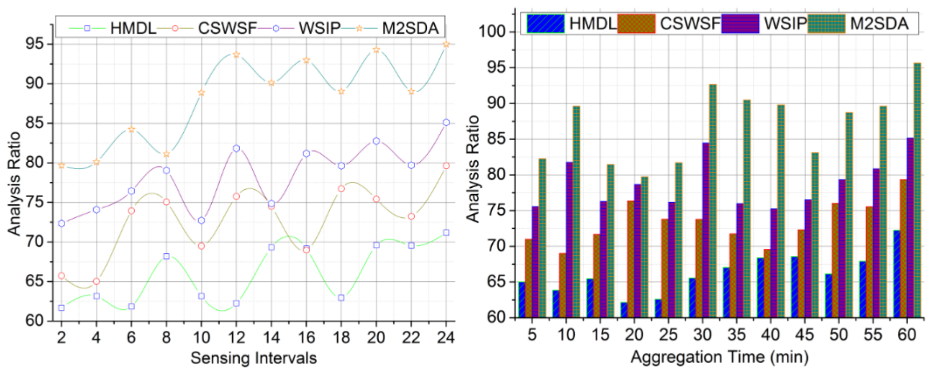

Fig. 13. Comparison of analysis rate

The cost functions of the linear regression learning algorithm to perceive the accuracy of the mapping function. It is also said to be the hypothesis function. The cost function denotes the average of the squared errors that occurs between the target input data and the forecast of ocean surface wind. It uses a gradient descent method to minimize the cost function. It intensifies the M2SDA method's analysis rate. The proposed method increases the analysis depending on the variation and intersection observed. In the proposed method, three conditions are analyzed for preventing ρ for distinct intervals. The missing data is analyzed for varying aggregation analysis and the consecutive intervals are validated from discrete classifications. Therefore the classifications are induced across multiple analytical instances. This is further validated using linear aggregation learning in failing and achievable conditions. The H_t and t for continuous and discrete instances are extracted for analysis. This further increases the w assignment for satisfying b^E assignment in suppressing variations. Therefore, this analytics also induces improvements across different intervals (Fig. 13). The below comparison study outcomes are discussed in Table 1 (Sensing Intervals) and Table 2 (Aggregation Time).

Table 1. Comparative analysis summary (Sensing intervals)

Metrics	HMDL	CSWSF	WSIP	M2SDA
Precision	0.621	0.702	0.816	0.8961
Error	0.253	0.181	0.148	0.1065
Analysis time (s)	8.46	6.44	4.09	2.2935
Identification ratio	45.58	55.13	64.84	75.165
Analysis ratio	71.16	79.63	85.12	95.008

Summary: The M2SDA enhances precision, identification, and analysis ratio with 9.16%, 9.9%, and 8.19%. Errors as well as analysis time are decreased by 8.75% and 10.63%, accordingly.

Table 2. Comparative analysis summary (Aggregation time)

Metrics	HMDL	CSWSF	WSIP	M2SDA
Precision	0.619	0.705	0.816	0.8838
Error	0.352	0.275	0.216	0.1722
Analysis time (s)	8.21	6.07	4.35	2.0648
Identification ratio	46.69	54.56	65.35	75.152
Analysis ratio	72.26	79.35	85.17	95.684

Summary: The proposed M2SDA maximizes precision, identification, and analysis ratio by 8.52%, 10.14%, and 8.38%. Error as well as analysis time are decreased by 10.88% and 11.13%, accordingly.

4 Conclusion

The main work we have done is summarized as follows:

(1) This article introduced a multi-modular semantic data analysis for predicting wind speed on ocean surfaces. The working principle and process of M2SDA are introduced, and the data analysis model is constructed.

(2) Considering that the time factor is confined, the forecast endurance based on sensing and aggregation time factors is considered in the identification process, which is required for preventing breaks in data analysis. The environmental data observed from distinct time intervals are used for retaining the linearity in forecasting wind speed. The modularity is analyzed using discreteness and linearity for validating the missing sequence.

(3) The performance of the proposed method was evaluated. The evaluation indicators are precision, error, analysis time, identification ratio, and analysis rate. For proving the proposed method's efficiency, a comparison with HMDL, CSWSF, and WSIP is performed. Experimental results showed that the M2SDA enhances precision, identification, and analysis ratio by 9.16%, 9.9%, and 8.19% accordingly. It decreases inaccuracy by 8.75% and analysis time by 10.63%, accordingly, which indicates that the practicality and efficacy of the system are increased by the M2SDA method's high rate of prediction accuracy.

References

[1] Y. Miao, X. Dong, M.-A. Bourassa, D. Zhu, Effects of ocean wave directional spectra on Doppler retrievals of ocean surface current, IEEE Transactions on Geoscience and Remote Sensing 60(2021) 1-12.

[2] H. Wang, J. Zhu, M. Lin, Y. Zhang, Y. Chang, Evaluating Chinese HY-2B HSCAT ocean wind products using buoys

and other scatterometers, IEEE Geoscience and Remote Sensing Letters 17(6)(2020) 923-927.
[3] C. Dutheil, S. Andréfouët, S. Jullien, R. Le Gendre, J. Aucan, C. Menkès, Characterization of south central Pacific Ocean wind regimes in present and future climate for pearl farming application, Marine Pollution Bulletin 160(2020) 111584.
[4] C.-W. Zheng, X.-H. Li, C. Azorin-Molina, C.-Y. Li, Q. Wang, Z.-N. Xiao, S.-B. Yang, X. Chen, C. Zhan, Global trends in oceanic wind speed, wind-sea, swell, and mixed wave heights, Applied Energy 321(2022) 119327.
[5] D. Olbers, P. Jurgenowski, C. Eden, A wind-driven model of the ocean surface layer with wave radiation physics, Ocean Dynamics 70(8)(2020) 1067-1088.
[6] Y. Wan, R. Qu, X. Shi, Y. Dai, A Joint Inversion Method of Wave and Wind Field Parameters Based on SAR SLC Data, IEEE Geoscience and Remote Sensing Letters 19(2022) 1-5.
[7] J. Bu, K. Yu, S. Han, N. Qian, Y. Lin, J. Wang, Retrieval of Sea Surface Rainfall Intensity Using Spaceborne GNSS-R Data, IEEE Transactions on Geoscience and Remote Sensing 60(2022) 1-16.
[8] B. He, L. Ye, M. Pei, P. Lu, B. Dai, Z. Li, K. Wang, A combined model for short-term wind power forecasting based on the analysis of numerical weather prediction data, Energy Reports 8(2022) 929-939.
[9] Z.-W. Wang, W.-M. Zhang, G.-M. Tian, Z. Liu, Joint values determination of wind and temperature actions on long-span bridges: Copula-based analysis using long-term meteorological data, Engineering Structures 219(2020) 110866.
[10] W. Dong, H. Sun, J. Tan, Z. Li, J. Zhang, Y. Zhao, Short-term regional wind power forecasting for small datasets with input data correction, hybrid neural network, and error analysis, Energy Reports 7(2021) 7675-7692.
[11] N. Salvação, A. Bentamy, C.-G. Soares, Developing a new wind dataset by blending satellite data and WRF model wind predictions, Renewable Energy 198(2022) 283-295.
[12] Z. Sun, M. Zhao, Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis, IEEE Access 8(2020) 134422-134434.
[13] T. Guo, L. Zhang, Z. Liu, J. Wang, A combined strategy for wind speed forecasting using data preprocessing and weight coefficients optimization calculation, IEEE Access 8(2020) 33039-33059.
[14] W. Wei, J. Wu, Y. Yu, T. Niu, X. Deng, Uncertainty quantification analysis of wind power: A data-driven monitoring-forecasting framework, IEEE Access 9(2021) 84403-84416.
[15] M.-M. Nezhad, A. Heydari, E. Pirshayan, D. Groppi, D.A. Garcia, A novel forecasting model for wind speed assessment using sentinel family satellites images and machine learning method, Renewable Energy 179(2021) 2198-2211.
[16] A.-T. Dosdoğru, A. B. İpek, Hybrid boosting algorithms and artificial neural network for wind speed prediction, International Journal of Hydrogen Energy 47(3)(2022) 1449-1460.
[17] O.-F. Eikeland, F.D. Hovem, T.E. Olsen, M. Chiesa, F.M. Bianchi, Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: An Arctic case, Energy Conversion and Management: X 15(2022)100239.
[18] L. Yang, Z. Zhang, A deep attention convolutional recurrent network assisted by k-shape clustering and enhanced memory for short term wind speed predictions, IEEE Transactions on Sustainable Energy 13(2)(2021) 856-867.
[19] Z. Ma, S. Guo, G. Xu, S. Aziz, Meta learning-based hybrid ensemble approach for short-term wind speed forecasting, IEEE Access 8(2020) 172859-172868.
[20] X.-H. Chu, J. He, H.-Q. Song, Y. Qi, Y.-Q. Sun, W.-H. Bai, W. Li, Q.-W. Wu, Multimodal deep learning for heterogeneous GNSS-R data fusion and ocean wind speed retrieval, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13(2020) 5971-5981.
[21] J. Li, Z. Song, X. Wang, Y. Wang, Y. Jia, A novel offshore wind farm typhoon wind speed prediction model based on PSO–Bi-LSTM improved by VMD, Energy 251(2022) 123848.
[22] J. Chen, H. Liu, C. Chen, Z. Duan, Wind speed forecasting using multi-scale feature adaptive extraction ensemble model with error regression correction, Expert Systems with Applications 207(2022) 117358.
[23] S. Harbola, V. Coors, Deep Learning Model for Wind Forecasting: Classification Analyses for Temporal Meteorological Data, PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science 90(2)(2022) 211-225.
[24] L. Yang, Z. Zheng, Z. Zhang, An Improved Mixture Density Network Via Wasserstein Distance Based Adversarial Learning for Probabilistic Wind Speed Predictions, IEEE Transactions on Sustainable Energy 13(2)(2022) 755-766.
[25] A. Saeed, C. Li, Z. Gan, Short-Term Wind Speed Interval Prediction using LUBE based Quasi-Recurrent Neural Network, in: Proc. 2021 International Conference on Communication Technology and Information Technology, 2021.
[26] G.-V. Drisya, K.S. Kumar, A Wavelet, AR and SVM based hybrid method for short-term wind speed prediction. <https://arxiv.org/abs/2203.15298>, 2022 (accessed 23.05.2023).
[27] L.-D. Bentsen, N.-D. Warakagoda, R. Stenbro, P. Engelstad, Wind Park Power Prediction: Attention-Based Graph Networks and Deep Learning to Capture Wake Losses, in: Proc. The Science of Making Torque from Wind (TORQUE 2022), 2022.
[28] Data.world, Weather Buoy Network, https://data.world/marineinstitute/35a18e98-2706-4eb3-b064-46a8a66baf91
[29] J. Wang, Z. Cheng, Wind speed interval prediction model based on variational mode decomposition and multi-objective optimization, Applied Soft Computing 113(2021) 107848.