DOI: https://doi.org/10.52756/ijerr.2024.v42.004 Int. J. Exp. Res. Rev., Vol. 42: 40-49 (2024)

Original Article | Peer Reviewed | Open Access

International Journal of Experimental Research and Review (IJERR)

ISSN: 2455-4855 (Online)
www.iaph.in

Forecasting Wind Speed Using Clustering of Trend-Based Time Series Data

Aditya Dubey[1*], Shirish Mohan Dubey[2], Jyoti Kumari[3], Pradeep Yadav[3] and Gaurav Sharma[2]

[1]Centre for Internet of Things, Madhav Institute of Technology & Science, Gwalior-474005, Madhya Pradesh, India; [2]Department of Computer Engineering and Science, Poornima College of Engineering, Jaipur -302022, Rajasthan, India; [3]Department of Computer Science Engineering, ITM College, Gwalior-474001, Madhya Pradesh, India

E-mail/Orcid Id:

AD, dubeyaditya65@gmail.com, https://orcid.org/0000-0002-4885-0632; *SMD*, dubeys78@gmail.com, https://orcid.org/0000-0001-6330-304X; *JK*, jyotikumari.asst@itmgoi.in, https://orcid.org/0009-0004-0533-4306; *PY*, pradeep.yadav@itmgoi.in, https://orcid.org/0000-0003-1073-7076; *GS*, gaurav801212@gmail.com, https://orcid.org/0009-0003-3089-6899

Article History:

Received: 06th Apr., 2024
Accepted: 28th Jul., 2024
Published: 30th Aug., 2024

Keywords:

Generalized autoregressive score, clustering, autoregressive integrated moving average, wind time series

How to cite this Article:
Aditya Dubey, Shirish Mohan Dubey, Jyoti Kumari, Pradeep Yadav and Gaurav Sharma (2024). Forecasting Wind Speed Using Clustering of Trend-Based Time Series Data. *International Journal of Experimental Research and Review, 42,* 40-49.
DOI: https://doi.org/10.52756/ijerr.2024.v42.004

Abstract: Accurate forecasting of wind speed is crucial for the efficient operation of wind energy systems. As a time-series concern, wind forecasting may help determine how much electricity a proposed wind farm might produce annually. The majority of forecasting techniques perform differently depending on seasonal and trends variation. For this reason, time series data frequently have seasonal and nonlinear trend components eliminated in order to simplify wind forecasting approach. The application function used to remove the seasonality and trend determines accuracy. The proposed method begins by identifying and extracting underlying trends from historical wind speed data, segmenting the time series into distinct trend-based components. This paper proposes a hybrid method for predicting time series. A method for clustering data has been designed that identifies clusters of time series data with similar trend components. Statistical procedures, such as generalized autoregressive and autoregressive integrated moving average scoring approaches, are used to each individual cluster after the appropriate clusters of related trend components have been identified. Ultimately, the components that are made are combined. The datasets collected from the NREL site. The experiment demonstrates that when compared to current statistical approaches, the cluster-based forecasting approach performs better. This research makes contribution towards the field of renewable energy forecasting by providing a robust and scalable method for wind speed prediction, which can be integrated into existing energy management systems for improving the efficiency and stability of wind energy generation. The research paper examines the trend features of time series data of wind employing the suggested hybrid technique on wind forecasting. Performance calculated in terms of RMSE and MAE shows that the proposed technique succeeds as compared to other state of the art techniques.

Introduction

One important factor in addressing the global energy dilemma is wind energy. A precise model for calculating the power produced by wind power facilities is necessary for it to be a dependable source of electricity. According to research published by Thapar et al. (2011), Wadhvani and Shukla (2018), Morshedizadeh et al. (2017) and Dongre and Pateriya (2019), there is a significant correlation between the electricity produced by a wind farm and the site's wind speed. The accuracy of power prediction may be increased using a reliable wind speed forecast model. Time series forecasting approaches may be utilised to predict wind speed at a specific location. By identifying the time series data's hidden pattern, these approaches employ the values of the time series data already in existence to forecast values in the future. For wind speed prediction, a number of statistical models were developed, including the generalized autoregressive score (GAS) (Creal et al., 2013), autoregressive integrated moving average (ARIMA) (Torres et al., 2005), and autoregressive moving average (ARMA) (Yang et al., 2015). Time series data typically includes seasonal as well as trend components, which can be both heterogeneous and homogenous in manner. While the

*Corresponding Author: dubeyaditya65@gmail.com

Table 1. Identified research gaps for forecasting techniques using clustering techniques.

Research Gap	Description
Integration with Other Predictive Models	There is often a lack of integration between clustering approaches and other advanced predictive models (e.g., deep learning), which could enhance forecasting accuracy by leveraging the strengths of multiple methodologies.
Scalability Issues	Many clustering-based methods may struggle to scale effectively when applied to large, high-resolution wind speed datasets. This limits their applicability in real-world scenarios where data size and complexity are significant.
Lack of Model Adaptation to Changing Conditions	Wind patterns can change rapidly due to various factors like weather conditions, terrain, and other environmental influences. Existing methods may not adapt quickly enough to these changes, as they often rely on static models that do not account for evolving trends in the data.
Insufficient Handling of Nonlinearity and Seasonality	Wind speed data is highly nonlinear and exhibits strong seasonal effects. Many current methods do not adequately address these complexities within the clustering framework, leading to models that struggle to generalize well across different time periods and conditions.
Inadequate Clustering Techniques	While clustering is used to group similar time series patterns, the techniques applied are often rudimentary, lacking sophistication in distinguishing subtle differences in trends. This can lead to suboptimal clustering, where important nuances in the data are overlooked, resulting in less accurate forecasts.
Limited Integration of Trend Dynamics	Existing approaches often fail to adequately capture the complex and dynamic nature of trends in wind speed data. Most methods focus on raw time series data or simplistic trend extraction, which may not fully account for the temporal patterns and variations inherent in wind speed.

seasonality and trend components are homogeneous, they can be adequately modeled by the classical models that are currently in use (Maatallah et al., 2015; Lydia et al., 2015; Kavasseri and Seetharaman, 2009). However, when there are heterogeneous seasonality and trend components, they must first be eliminated using the appropriate techniques before being modeled. In this case, Certain valuable patterns may be lost from the time series data if the seasonal and trend aspects are removed.

In practice, statistical approaches are often employed since they yield predictive results faster (Chih-Hung et al., 2024). The inability of these methods to handle diverse time series data is one of its drawbacks. A number of hybrid techniques have been devised recently to accurately represent the heterogeneous time series data. For wind speed prediction, Kushwah and Wadhvani (2019) provided hybrid modeling methodologies based on neural networks and GAS. The previous GAS models had performed effectively with desirable levels of errors in prediction when a neural network was added. According to Inniss (2006), a nonlinear seasonality and trend pattern may be present in any time series data, and these features may be utilized to separate heterogeneous wind speed data into homogenous data. A wind time series' trend component displays the data's typical propensity to either rise or decrease over an extended period of time. On the other hand, seasonality refers to the occurrence of regular variations across the time series data. For the transformation of nonlinear trends and seasonal components into linear patterns, statistical techniques are utilized for transforming non-stationary time series into stationary ones. The clustering approach has been proposed by Vilar et al. (2018) and Kuznetsov and Mohri (2020) as a means of dividing heterogeneous wind speed data into a homogenous sample.

An unsupervised learning technique called clustering separates the data samples into an equal number of clusters. Numerous clustering techniques that work with both sequential and non-sequential data samples may be found in the existing research. Techniques designed for non-sequential data are not effective when used for time series data because of the differences in their features. When methods like clustering are used for non-sequential data, the number of clusters produced is minimal and is identified by a distance measure for data values. One methodology utilised for determining the ideal number of clusters involves the K-mean clustering technique (Zhu et al., 2019). Whereas, serial correlation between succeeding observations is a feature of time series data. Similar data cannot be combined into clusters using the distance measure without affecting the serial correlation. According to Lim et al. (2018), time series data may be utilized to find comparable structured data values for creating clusters since they may show seasonality and a pattern over time.

The first section of this article introduces proposed clustering techniques for locating time series data segments with similar trend shapes. Following the creation of clusters of related trends, the time series data for each cluster was modeled using statistical time series prediction techniques, specifically GAS and ARIMA. This is a hybrid technique of forecasting, where the final values are predicted by combining the output from models that are specific to each cluster. Lastly, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are employed for assessing the accuracy of suggested hybrid models, such as C-GAS and C-ARIMA, and statistical approaches, such as GAS and ARIMA.

DOI: https://doi.org/10.52756/ijerr.2024.v42.004

Gaps in the existing research

The gap in current wind speed forecasting methods using clustering of trend-based time series data is described in table 1. Addressing these gaps would involve developing more sophisticated clustering techniques, better trend extraction methods, and integrating adaptive models that can handle the nonlinearity, seasonality, and evolving nature of wind speed data.

Research Objectives

- Develop methods to accurately identify and extract underlying trends from historical wind speed data.
- Segment the time series into meaningful trend-based components that reflect different patterns in wind speed behavior.
- Implement advanced clustering algorithms to group similar trend-based components of the time series.
- Ensure that the clustering process captures subtle variations and similarities in wind speed trends, enabling more precise pattern recognition.
- Construct predictive models tailored to each cluster, taking into account the specific temporal relationships and characteristics within each group.
- Develop an ensemble of models based on the clustered trends to improve the overall accuracy and robustness of wind speed forecasts.
- Test and validate the ensemble approach against traditional forecasting methods to demonstrate its superiority in handling complex wind speed data.
- Integrate mechanisms within the clustering and forecasting models to account for the nonlinear and seasonal nature of wind speed data.
- Ensure that the models adapt to these characteristics dynamically, enhancing their generalizability across different time periods and conditions.

Design models that can quickly adapt to changes in wind patterns due to environmental or meteorological factors.

Trend-based analysis

A trend in time series data refers to the long-term movement or direction in the data, which may indicate an overall increase, decrease, or cyclical behavior over a period of time. Wind speed data often exhibits complex patterns due to the influence of various meteorological factors, seasonal changes, and environmental conditions. Identifying these trends is essential for accurate forecasting. Trend-based analysis involves identifying and leveraging the underlying patterns in wind speed data over time to improve forecasting accuracy.

Extraction of Trends

The first step in the trend-based analysis is to decompose the wind speed time series data into its constituent components, typically trend, seasonality, and noise. Techniques like moving averages, polynomial fitting, or more sophisticated methods like Seasonal-Trend decomposition using LOESS (STL) can be used. Once the trend component is extracted, the data is segmented into trend-based time series segments. These segments represent different phases of wind speed behavior, such as periods of steady increase, decrease, or stability.

Clustering of Trend-Based Segments

The extracted trend-based segments are then clustered based on their similarity (Azhar and Huzaifa, 2024). Clustering algorithms like K-means, hierarchical clustering, or more advanced methods like DBSCAN can be employed to group segments with similar trend patterns. Each cluster represents a distinct pattern or behavior in wind speed trends, such as steady growth, rapid decline, or cyclical fluctuations. This clustering helps in simplifying the complexity of the time series data by categorizing it into a finite number of recognizable patterns.

Modeling Based on Clusters

For each cluster identified, a specific forecasting model is developed. Since each cluster represents a different type of trend, the models can be fine-tuned to capture the unique characteristics of the trend within that cluster. The individual cluster-based models are then combined into an ensemble to produce a comprehensive wind speed forecast. This ensemble approach enhances the overall forecasting accuracy by leveraging the strengths of different models tailored to specific trend patterns.

Advantages of Trend-Based Analysis

By focusing on trends, the analysis reduces noise and focuses on the fundamental patterns that drive wind speed changes. This leads to more accurate predictions, especially in the short to medium term. Trend-based analysis allows the forecasting models to adapt to new trends as they emerge, making the approach more responsive to changes in wind patterns over time. Clustering trend-based segments makes the forecasting process more interpretable, as it is easier to understand and analyze the specific patterns driving the predictions.

Application in Wind Speed Forecasting

Accurate wind speed forecasts are crucial for optimizing the operation and integration of wind energy into the power grid. Trend-based analysis provides a more reliable tool for predicting wind speeds, leading to better energy management decisions. The insights gained from trend-based analysis can be used to support decisions in various sectors, including renewable energy production, weather forecasting, and climate research.

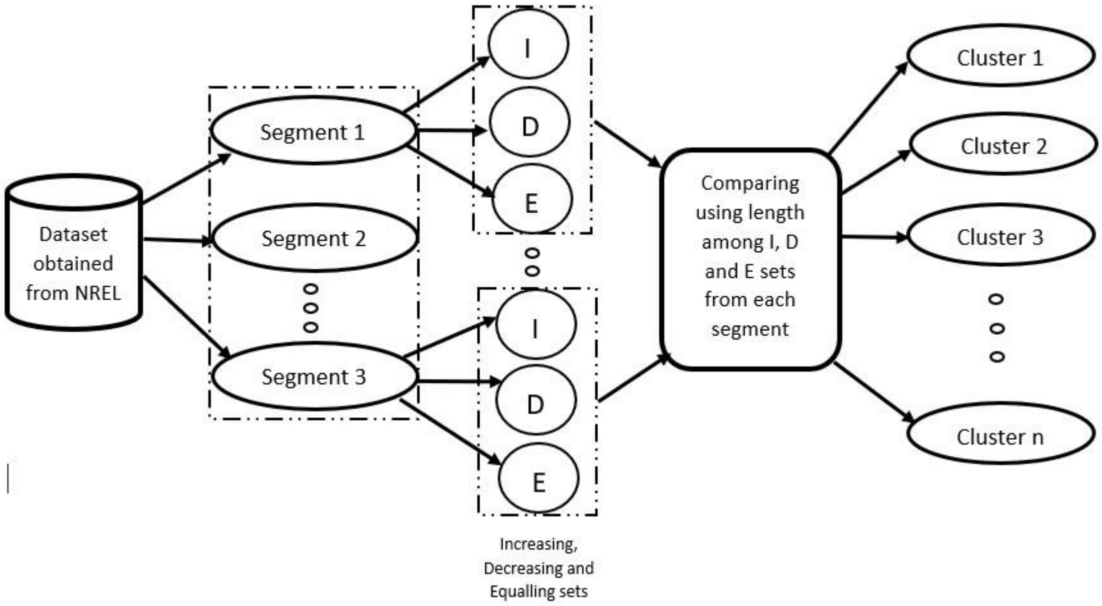

Figure 1. Clustering technique using trend component analysis.

In summary, trend-based analysis in this research paper involves identifying, segmenting, and clustering trends in wind speed data to build more accurate and adaptable forecasting models. This approach addresses the complexity of wind speed data, allowing for better prediction and management of wind energy resources.

Clustering of time series data

The time series statistics on wind speed often include seasonal and trend elements. The statistical feature across the period is extracted from wind time series data using a seasonal and trend analysis (Johnpaul et al., 2020). The main goal of this research is to classify time series segments according to comparable trend behavior. Any time series' statistical data values can exhibit one of three trend features: rising, falling, or equalling. Each segment of the data could follow one or more distinct patterns. Clusters may be created based on the order in which these patterns appear in the segments (Wang et al., 2006).

A comprehensive detail of the clustering method, which is predicated on locating related trend components, is provided in Figure 1. The trend displays the overall pattern of the data values, which might be in a fashion of equal, decreasing, or growing. S represents the segment, and E, D, and I stand for equaling, decreasing, and increasing sets, respectively. Assume that statistics data are used to produce m segments (s1, s2, s3,..., sm). Determine the E, D, and I set for every segment of S first, and then determine the length of all of them. Find the comparable segments lately using the computed length. Allocate segments to the same cluster if two or more exhibit the same pattern. Place E, D, and I in the same cluster if their lengths are equal; place E, D, and I in different clusters if their lengths differ. Put D and I in different clusters if their lengths are greater than I and E's. Similarly, a varied number of clusters are identified depending on the length of the sets.

Forecasting wind speed using statistical models

The statistical models may be utilized for forecasting the time series data with respect to accuracy, horizon and data. Utilizing variables including temperature, air density, wind speed, direction, and so on, wind time series data are forecasted. Predicted wind speed data make up the majority of the predicted wind time series data. The time frame during which the parameter will occur in the future is known as the forecast horizon, and it typically varies from long-term (several days ahead) to short-term (one day ahead). The modeling technique's effectiveness is measured by forecasting accuracy, which may be assessed using suitable performance measures. These statistical methods are used to assess the predicted accuracy: GAS and ARIMA.

ARIMA model

For time series data, ARIMA model was presented by Maatallah et al. (2015) and is utilized in the prediction. A more comprehensive variant of the ARMA model is ARIMA model, which is utilised to forecast values in the future in time series data by employing values from the past. In cases when the given time series lacks stationarity, the ARIMA model might be employed. In this instance, the non-stationarity from time series data could be eliminated by applying the differencing approach once or more times. Regression error, developing variables, and integrated (differencing) over present and past values are indicated by the letters AR, MA, and I, respectively. The ARIMA model (q, d, p) may be expressed mathematically as

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \cdots + \alpha_p \alpha_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \cdots + \beta_q \varepsilon_{t-q} \quad (1)$$

DOI: https://doi.org/10.52756/ijerr.2024.v42.004

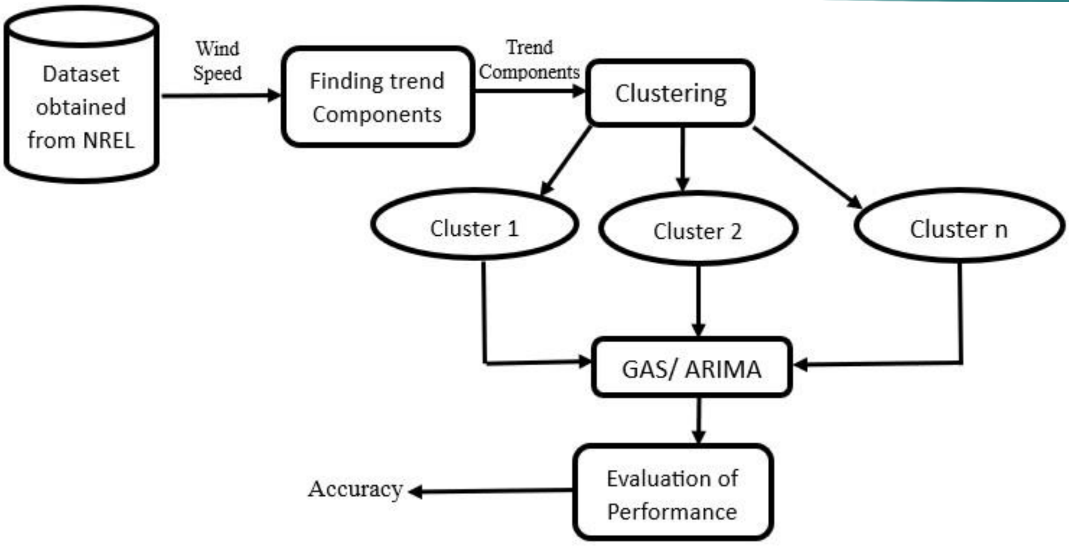

Figure 2. Wind speed forecasting model.

The moving average model's order is represented by q. the MA coefficients are β_1 ,...., and β_q . d is a representation of the number of differencing factors that are subtracted to create the stationary time series data from current values to past values. The autoregressive model's proportion of time lags is represented by the symbol p. The values of the parameters q, d, and p are all non-negative. White noise is depicted by ε_t. The AR coefficients are α_0 , α_1 ,...., and α_p ; where the legged variable of interest y_t is a pth-valued.

GAS model

For analyzing nonlinearity in time series data, the score function is utilized as part of the score-driven GAS model. Because the GAS model is of observation-driven nature, it may be used for long-term data that lacks future complexity and asymmetric data having complex dynamics (Creal et al., 2013). When dealing with wind time series having fluctuating density, the GAS model is employed. The conditional observation density $P(yt\theta t/)$ may be used to characterize this model, wherein θt represents a latent time-varying component and yt is an interest variable that depends on it. But modifying the time-varying variable recursively utilizing the autoregressive calculation, which is expressed as

$$\theta_t = \omega + \sum_{i=1}^{p} \emptyset_i \theta_{t-i} + \sum_{j=1}^{q} \alpha_j S(\theta_{j-1}) \frac{\partial logp(y_{t-j}|\theta_{t-j})}{\partial \theta_{t-j}}$$

Wherein, s strictly positive scaling factor $\emptyset$ represents an autoregressive coefficient and ω is a vector of constants. The conditional densities P's first derivative is multiplied by s, which yields one observation at time j, and α represents a scaling parameter. S is dependent upon the time-varying parameter θ_t and the observation y_t. The primary motive of the GAS model is selecting driving mechanism(s), that are relevant to various nonlinear modeling methods. When it comes to nonlinear data, the GAS model outperforms the ARIMA. The GAS model uses whole density structure rather than just means and higher moments as it depends on the score.

Proposed Technique

The NREL website provided the dataset that was utilized in the research. Several variables, such as air density, air temperature, wind power, wind speed, and so on, are included in the resource file. Since the primary goal of the research is univariate time series prediction, mainly wind speed data have been taken into account while modeling. Figure 2 provides a thorough summary of the proposed technique.

The entire set of data is split up into pairs for testing and training. The approach that is being proposed is a hybrid one that integrates statistical forecasting techniques with time series data clustering. Specifically, the training set is first split up into equal-sized segments. The overall dimension of the completed data set used for modeling determines the segment's size. Following that, the clusters were formed by applying the proposed clustering approach, as described in the section. This has led to the formation of several groups with a linear trend aspect in time series data. This data has been modeled utilizing statistical time series forecasting techniques, specifically GAS and ARIMA, because each cluster has a linear trend aspect. The hybrid techniques have been named C-GAS (GAS simulation with clustering methodology) and C-ARIMA (ARIMA simulation with clustering methodology). In the C-ARIMA, clustering is followed by the use of the ARIMA statistical approach. Similar to the previous, the GAS statistical procedure was used following clustering. The hybrid forecasting approach is developed on the training set, and its performance is assessed on testing data using the RMSE and MAE metrics. The section under " Results and analysis" has covered the assessment method's whole procedure.

Results and discussion

This section describes experiments carried out on a variety of dataset before going into the results of research. The details of the datasets utilized in the

Table 2. Summary of the dataset.

Dataset	Year	Site Id	Max	Min	Standard deviation	Mean
#1	2012	16833	21.875	0.189	4.407	9.668
#2	2011	16833	17.829	0.067	3.911	7.816
#3	2010	16833	20.798	0.46	4.093	8.553
#4	2009	16833	19.128	0.653	3.792	8.772
#5	2008	16833	20.841	0.14	4.357	9.831
#6	2007	16833	20.84	0.161	3.963	7.967
#7	2012	124693	26.178	0.048	5.132	5.772
#8	2011	68003	16.772	0.094	3.566	8.646
#9	2010	72509	21.418	0.262	4.498	9.267
#10	2009	72509	27.689	0.253	5.818	13.925
#11	2008	72509	27.832	0.415	5.676	13.604
#12	2007	72509	28.422	0.088	6.055	11.513

Table 3. RMSE and MAE results utilizing the clustered ARIMA and ARIMA models.

Dataset	C3-ARIMA		C2-ARIMA		C1-ARIMA		ARIMA	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
#1	4.949	3.917	12.232	11.128	4.808	4.220	6.593	4.751
#2	6.956	6.471	2.609	2.048	2.699	2.144	2.743	2.188
#3	2.591	2.294	4.983	4.725	2.869	2.528	3.771	3.455
#4	9.877	9.358	10.010	9.483	12.062	11.621	4.785	4.207
#5	5.645	4.880	4.252	3.554	4.049	3.362	4.950	4.246
#6	123.995	92.738	2.106	1.714	3.833	3.257	3.291	2.796
#7	7.241	5.679	7.331	5.653	7.254	5.563	7.124	5.558
#8	7.697	7.404	6.685	6.344	11.136	10.319	4.359	4.074
#9	14.061	13.893	9.383	9.113	6.757	6.374	6.976	6.593
#10	6.964	6.156	5.021	3.821	3.897	2.747	4.281	2.968
#11	6.047	4.726	5.206	4.191	8.430	7.029	5.972	4.675
#12	6.326	5.570	6.937	5.940	5.973	5.159	8.649	7.346

Table 4. RMSE and MAE results utilizing the clustered GAS and GAS models.

Dataset	C3-GAS		C2-GAS		C1-GAS		GAS	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
#1	7.259	6.093	6.823	5.213	6.899	5.900	5.951	4.378
#2	8.760	8.334	3.078	2.420	3.468	2.988	3.164	2.474
#3	3.159	2.820	8.800	8.576	2.426	2.112	5.371	4.952
#4	8.227	7.387	5.370	4.321	6.553	5.520	8.019	7.068
#5	5.777	5.029	6.077	5.341	6.258	5.487	7.166	6.324
#6	7.559	7.266	8.579	8.110	7.107	6.928	5.373	5.212
#7	7.955	6.143	7.771	5.940	9.369	7.175	9.359	7.268
#8	5.402	5.188	2.016	1.715	3.470	3.155	2.599	1.851
#9	10.250	9.919	5.940	4.767	2.449	1.794	3.431	2.738
#10	3.356	2.887	2.395	1.785	15.224	13.468	3.820	3.003
#11	6.133	4.802	5.672	4.510	7.304	5.776	6.176	4.788
#12	7.342	6.454	7.377	6.324	3.398	2.955	5.767	5.017

experiments were provided in the first sub-section. The measures used to assess performance are covered in the second sub-section. The results are analyzed in the final sub-section.

Dataset Description

The datasets utilized in the research are indeed from NREL and have the site IDs 124693, 68003, 16883, and 72509 (NREL, 2007). The site ID 72509 has 9.241 m/s average wind speed, located at latitude 41:7768 and longitude ~106:2598. Table 2 includes a comprehensive summary of every dataset that has been addressed. With 105,120 observations, the wind plant system known as supervisory control and data acquisition (SCADA) at a height of 100 meters yields an average 5-min wind speed. 8500 samples were collected for our research study and split into pairs for testing and training purposes.

Performance measuring criteria

The appropriate criteria that evaluate the models' abilities are used to determine how well hybrid models and statistical function. The wind speed predicting performance is measured for the experiment using RMSE and MAE. In this case, $\hat{y}$ represents the predicted variable, y indicates the input variable, and N denotes the total number of observations. The model that performs best has the lowest possible RMSE and MAE values. The Python 3.6 variant is employee for implementing GAS, ARIMA, and hybrid models.

DOI: https://doi.org/10.52756/ijerr.2024.v42.004

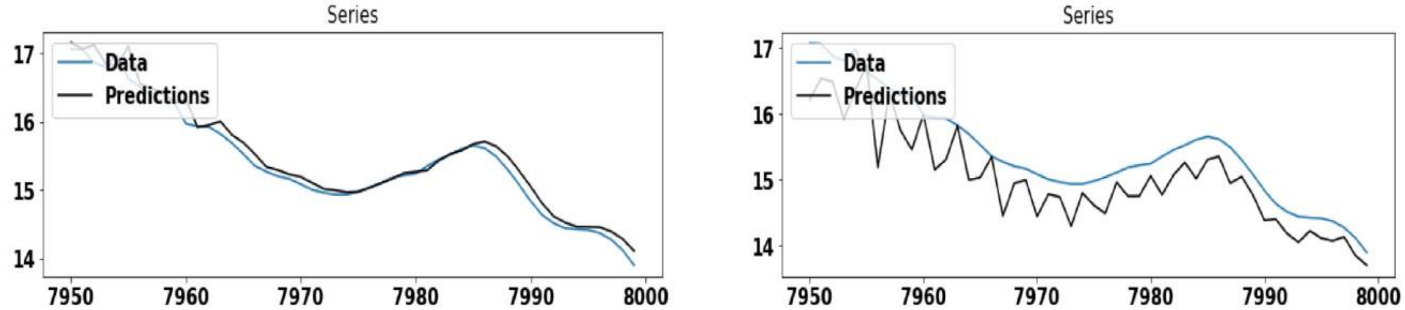

Figure 3. The wind speed forecast on Dataset #1 employing the GAS and ARIMA models, respectively, is depicted in the left and right columns.

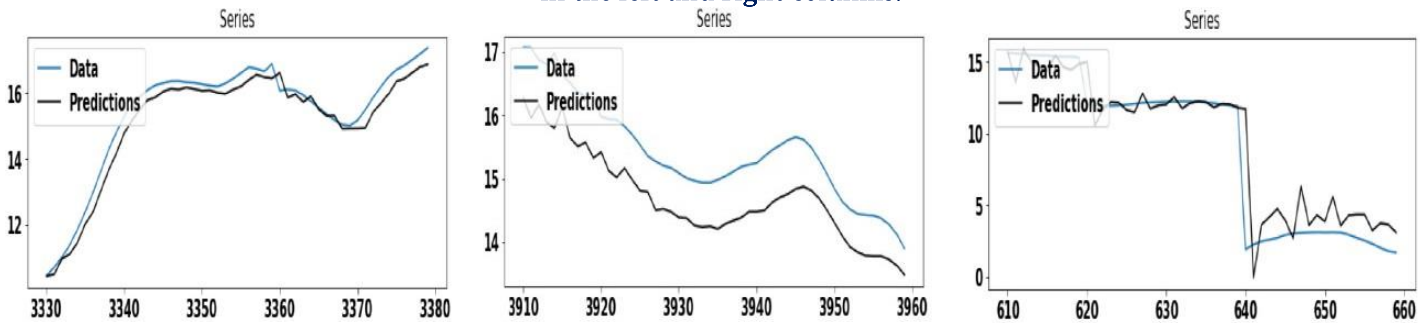

Figure 4. The first, second, and third clusters' wind speed predictions utilizing the ARIMA model on Dataset #1 are depicted in the images in the left to right sections, respectively.

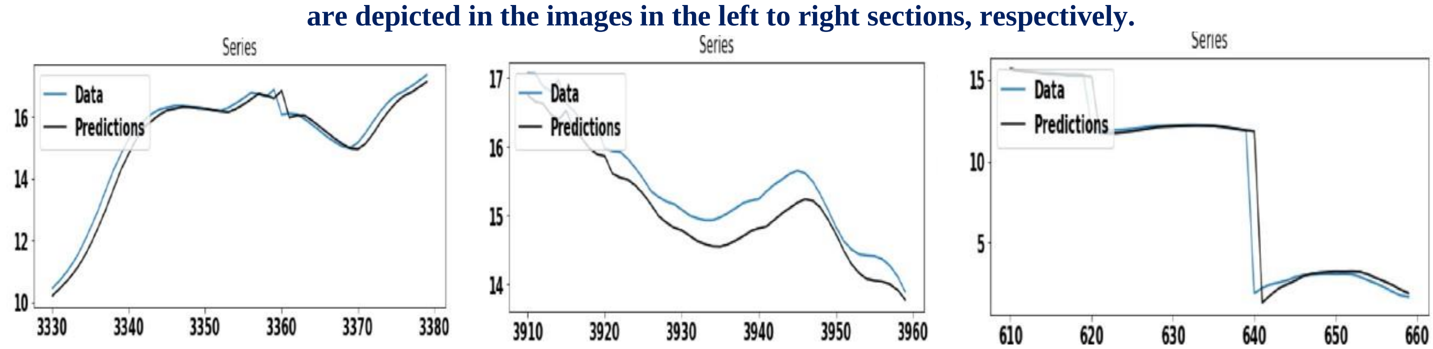

Figure 5. The first, second, and third clusters' wind speed predictions utilizing the GAS model on Dataset #1 are depicted in the images in the left to right sections, respectively. Wind speed forecast utilizing the GAS model for the 1st, 2nd and 3rd clusters over Dataset #1 is depicted in the left to right sections, respectively.

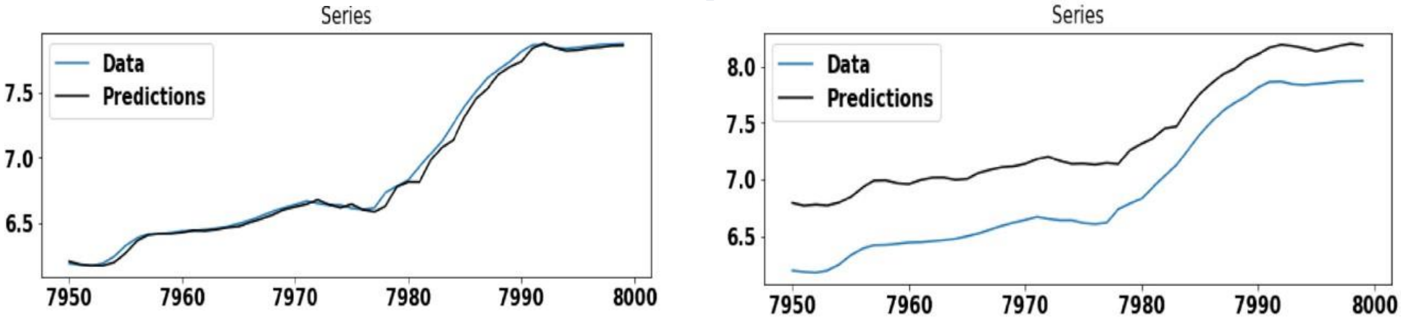

Figure 6. Wind speed forecast on Dataset #7 utilizing the GAS and ARIMA models, respectively, is depicted in the left and right sections.

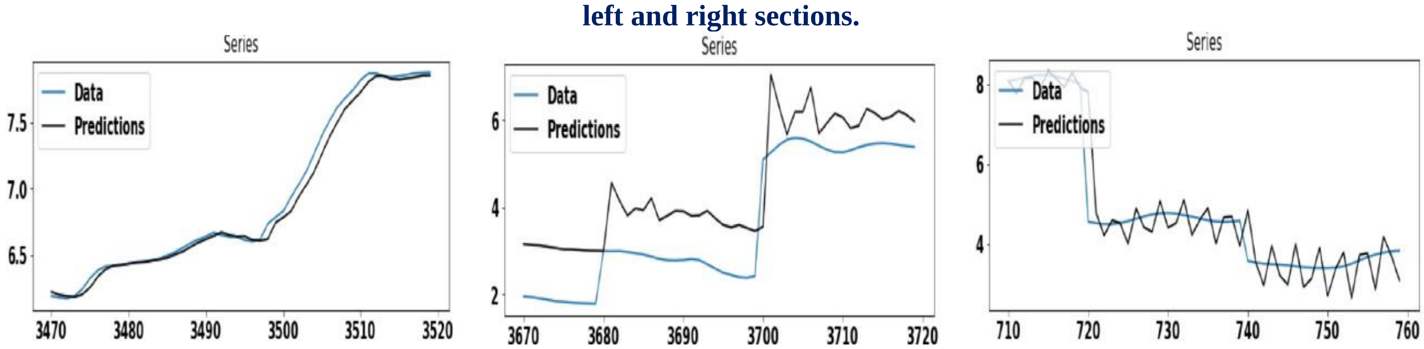

Figure 7. The first, second, and third clusters' wind speed predictions utilizing the ARIMA model on Dataset #7 are depicted in the images in the left to right sections, respectively.

DOI: https://doi.org/10.52756/ijerr.2024.v42.004

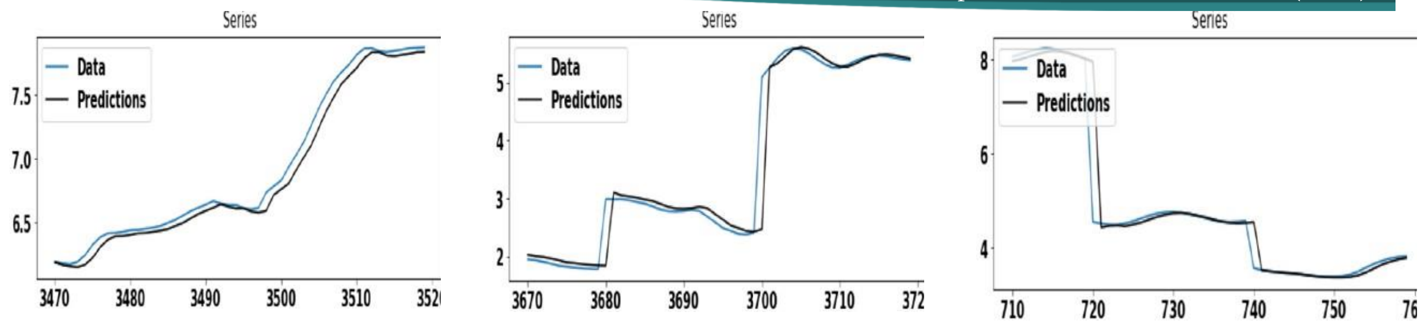

Figure 8. The first, second, and third clusters' wind speed predictions utilizing the GAS model on Dataset #7 are depicted in the images in the left to right sections, respectively.

Performance Comparison

This research paper applies hybrid models (C-GAS and C- ARIMA) and statistical techniques (GAS and ARIMA) to twelve distinct datasets taken via the NREL site. Table 3 displays the ARIMA models' forecasting results with regard to RMSE and MAE values. In the same way, Table 4 displays the GAS model's predictions.

Wind speed forecast employing GAS and ARIMA models, performed independently on Dataset #1, is shown in Figure 3. In this case, the GAS model has higher accuracy than ARIMA model. Wind speed forecast employing the ARIMA model, implemented separately to every cluster in Dataset #1, is shown in Figure 4. When compared to other clusters, the ARIMA model performs more accurately on cluster 1. In the same way, Figure 5 shows how the GAS model is used, one for every cluster in Dataset #1, to estimate wind speed.

The wind speed forecast employing the GAS and ARIMA models performed independently on Dataset #7 is shown in Figure 6. In this case, the GAS model outperforms the ARIMA model in terms of accuracy. The wind speed forecast employing the ARIMA model, implemented separately to every cluster in Dataset #7, is shown in Figure 7. When compared to other clusters, the ARIMA model performs more accurately on cluster 2. In an identical manner, Figure 8 shows how the GAS model is used, one for each cluster in Dataset #7, to estimate wind speed. The results of the experiment are presented in Tables 3 and 4, with regard to the RMSE and MAE values attained for the GAS and ARIMA variations, respectively. Tables 3 and 4 display the desired minimum values of RMSE and MAE, which signify superior performance. Table 3 proves that the implemented hybrid model, C1-ARIMA, has higher performance than the ARIMA model. C1-ARIMA model, which was experimentally designed for Dataset #7, yields results of 2.106 and 1.714 with respect to RMSE and MAE values, respectively. Table 3 shows that, as opposed to the ARIMA model, ARIMA model with the clustering principle performed better the majority of the time. Table 3 shows that the suggested hybrid model, C2-GAS, works better than the ARIMA model. C2-GAS model that was empirically generated on Dataset #7 yields results of 2.016 and 1.715, respectively, with regard to RMSE and MAE values. Additionally, table 3 shows that with regard to the GAS model, the cluster-based GAS model has higher performance for the majority of the time. The hybrid models utilize clustering and perform better than the statistical models.

Conclusion & Future Works

Although recent models of wind forecasting, such as GAS and ARIMA, are more user-friendly, their application is severely limited owing to the data complexity. It is possible to improve model performance by reducing the complexity of the time series data without losing useful patterns. One such technique for best performance in generalization of our model is presented in this paper. Experiments have shown that the hybrid model utilizing clustering (C- GAS and C-ARIMA) outperforms the current GAS and ARIMA models. The research paper examines the trend features of time series data of wind employing the suggested hybrid technique on wind forecasting. It demonstrates that the trend aspect in time series data of wind has diverse forms. The results achieve accuracy provided the time series data is initially classified based on the trend component's form, and each cluster's model is then created utilizing the existing forecasting approach. The model is being created using twelve distinct datasets collected from the NREL site to improve generalization.

Future research directions could explore several areas to enhance the effectiveness and applicability of the proposed approach. One of the potential directions is to investigate the use of more sophisticated clustering algorithms, such as hierarchical clustering, density-based clustering (e.g., DBSCAN), or fuzzy clustering, to improve the accuracy and robustness of wind speed forecasting. The second future work is to explore hybrid models that combine clustering with other forecasting techniques, such as reinforcement learning, ensemble methods, or deep learning (e.g., LSTM networks), to

DOI: https://doi.org/10.52756/ijerr.2024.v42.004

improve predictive performance and capture complex temporal patterns. The third future work is to develop new feature engineering techniques to extract more relevant information from time series data, such as incorporating additional meteorological variables (e.g., humidity, temperature) or employing advanced statistical features. Next, future work includes assessing the scalability of the proposed method for large datasets and its applicability to real-time forecasting systems, which could involve optimizing computational efficiency and response times. Another work for the future is to investigate how the forecasting approach can be integrated with Internet of Things (IoT) devices and smart grid systems for real-time monitoring and management of wind energy resources. Examine how well the clustering approach handles seasonal variations and extreme weather events and develop strategies to enhance forecasting accuracy during such periods. Conduct comparative studies with other forecasting methods and approaches, such as statistical models (e.g., ARIMA), machine learning models (e.g., support vector machines), or hybrid approaches, to evaluate relative performance and advantages. Extend the research to different geographical regions and climates to assess the generalizability and adaptability of the clustering-based forecasting method. Explore advanced data preprocessing techniques to handle noisy or missing data and improve the quality of time series inputs for clustering and forecasting. Develop user-centric applications and tools that leverage the forecasting results for practical decision-making in sectors such as renewable energy management, agriculture, or disaster preparedness. These directions aim to build on the existing research by enhancing applicability, scalability, and accuracy, ultimately advancing the field of wind speed forecasting and its practical implementations.

Conflict of Interest

The authors declare no conflict of interest.

References

Azhar, A., & Huzaifa, H. (2024). Wind Mapping of Malaysia Using Ward's Clustering Method. *Energies, 17*, 1563. https://doi.org/10.3390/en17071563

Chih-Hung, H., Chien-Huei, L., Louis, Y. Y. L., Cruz, A. C., & Daim, T. (2024). Forecasting patenting areas with academic paper & patent data: A wind power energy case. *World Patent Information, 78*, 1-18. https://doi.org/10.1016/j.wpi.2024.102297

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics, 28*(5), 777–795. https://doi.org/10.1002/jae.1279

Dongre, B., & Pateriya, R. K. (2019). Power curve model classification to estimate wind turbine power output. *Wind Engineering, 43*(3), 213–224. https://doi.org/10.1177/0309524X19891671

Inniss, T. R. (2006). Seasonal clustering technique for time series data. *European Journal of Operational Research, 175*(1), 376–384. https://doi.org/10.1016/j.ejor.2005.03.049

Johnpaul, C. I., Prasad, M. V. N. K., Nickolas, S., & Gangadharan G.R. (2020). Trendlets: A novel probabilistic representational structures for clustering the time series data. *Expert Systems with Applications, 145*, 113–119. https://doi.org/10.1016/j.eswa.2019.113119

Kavasseri, R. G., & Seetharaman, K. (2009). Day-ahead wind speed forecasting using f-ARIMA models. *Renewable Energy, 34*(5), 1388–1393. https://doi.org/10.1016/j.renene.2008.09.006

Kushwah, A. K., & Wadhvani, R. (2019). Performance monitoring of wind turbines using advanced statistical methods. *Sadhana, 44*, 163. https://doi.org/10.1007/s12046-019-1145-6

Kuznetsov, V., & Mohri, M. (2020). Discrepancy-based theory and algorithms for forecasting non-stationary time series. *Annals of Mathematics and Artificial Intelligence, 88*, 367–399. https://doi.org/10.1007/s10472-019-09683-1

Lim, B. Y., Wang, J. G., & Yao, Y. (2018). Time-series momentum in nearly 100 years of stock returns. *Journal of Banking and Finance, 97*, 283–296. https://doi.org/10.1016/j.jbankfin.2018.10.010

Lydia, M., Suresh, S. K., Immanuel Selvakumar, A., & Edwin, P. K. G. (2015). Wind resource estimation using wind speed and power curve models. *Renewable Energy, 83*, 425–434. https://doi.org/10.1016/j.renene.2015.04.045

Maatallah, O. A., Achuthan, A., & Janoyan, K. (2015). Recursive wind speed forecasting based on Hammerstein autoregressive model. *Applied Energy, 145*, 191–197. https://doi.org/10.1016/j.apenergy.2015.02.032

Morshedizadeh, M., Kordestani, M., Carriveau, R. Ting, D. S. K., & Saif, M. (2017). Improved power curve monitoring of wind turbines. *Wind Engineering, 41*(4), 260–271. https://doi.org/10.1177/0309524X17709730

National Renewable Energy Laboratory (NREL) (2007). Western dataset (Site-id 72509). Available at: https://www.nrel.gov/grid/western-wind-data.html

Thapar, V., Agnihotri, G., & Sethi, V. K. (2011). Critical analysis of methods for mathematical modelling of wind turbines. *Renewable Energy*, *36*(11), 3166–3177. https://doi.org/10.1016/j.renene.2011.03.016

Torres, J. L., Garcıa, A., De Blas, M., & Francisco, A. D. (2005). Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Solar Energy*, *79*(1), 65–77. https://doi.org/10.1016/j.solener.2004.09.013

Vilar, J. A., Lafuente-Rego, B., & D'Urso, P. (2018). Quantile autocovariances: A powerful tool for hard and soft partitional clustering of time series. *Fuzzy Sets and Systems*, *340*, 38–72. https://doi.org/10.1016/j.fss.2017.03.006

Wadhvani, R., & Shukla, S. (2018). Analysis of parametric and non-parametric regression techniques to model the wind turbine power curve. *Wind Engineering*, *43*(3), 225–232. https://doi.org/10.1177/0309524X18780398

Wang, X., Smith, K., & Hyndman, R. (2006). Characteristic-based clustering for time series data. *Data Mining and Knowledge Discovery*, *13*(3), 335–364. https://doi.org/10.1007/s10618-005-0039-x

Yang, D., Sharma, V., Ye, Z., Lim, L. I., Zhao, L., & Aryaputera, A. W. (2015). Forecasting of global horizontal irradiance by exponential smoothing, using decompositions. *Energy*, *81*, 111–119. https://doi.org/10.1016/j.energy.2014.11.082

Zhu, E., Zhang, Y., Wen, P., & Liu, F. (2019). Fast and stable clustering analysis based on grid-mapping K-means algorithm and new clustering validity index. *Neurocomputing*, *363*, 149–170. https://doi.org/10.1016/j.neucom.2019.07.048

How to cite this Article:

Aditya Dubey, Shirish Mohan Dubey, Jyoti Kumari, Pradeep Yadav and Gaurav Sharma (2024). Forecasting Wind Speed Using Clustering of Trend-Based Time Series Data. *International Journal of Experimental Research and Review*, *42*, 40-49.

DOI : https://doi.org/10.52756/ijerr.2024.v42.004