IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 17, 2021, accepted May 25, 2021, date of publication May 27, 2021, date of current version June 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084536

Short-Term Wind Speed Forecasting Based on Hybrid MODWT-ARIMA-Markov Model

MUHAMMAD UZAIR YOUSUF[1,2], IBRAHIM AL-BAHADLY[1], (Senior Member, IEEE), AND EBUBEKIR AVCI[1], (Member, IEEE)

[1]Department of Mechanical and Electrical Engineering, Massey University, Palmerston North 4442, New Zealand
[2]Department of Mechanical Engineering, NED University of Engineering and Technology, Karachi 75270, Pakistan

Corresponding author: Muhammad Uzair Yousuf (M.U.Yousuf@massey.ac.nz)

This work was supported in part by the Higher Education Commission (HEC) Pakistan Human Resource Development (HRD) Program HRD Initiative-MS Leading to Ph.D. Program of Faculty Development for UESTPs/UETs Phase-I under Grant 5-1/HRD/UESTPI(Batch VI)/6082/2019/HEC, and in part by the Department of Mechanical and Electrical Engineering, Massey University and Tilt Renewables Tararua Wind Farm Research Bursary.

ABSTRACT Markov chains (MC) are statistical models used to predict very short to short-term wind speed accurately. Such models are generally trained with a single moving window. However, wind speed time series do not possess an equal length of behavior for all horizons. Therefore, a single moving window can provide reasonable estimates but is not an optimal choice. In this study, a forecasting model is proposed that integrates MCs with an adjusting dynamic moving window. The model selects the optimal size of the window based on a similar approach to the leave-one-out method. The traditional model is further optimized by introducing a self-adaptive state categorization algorithm. Instead of synthetically generating time series, the modified model directly predicts one-step ahead wind speed. Initial results indicate that adjusting the moving window MC prediction model improved the forecasting performance of a single moving window approach by 50%. Based on preliminary findings, a novel hybrid model is proposed integrating maximal overlap discrete wavelet transform (MODWT) with auto-regressive integrated moving average (ARIMA) and adjusting moving window MC. It is evident from the literature that MC models are suitable for predicting residual sequences. However, MCs were not considered as a primary forecasting model for the decomposition-based hybrid approach in any wind forecasting studies. The improvement of the novel model is, on average, 55% for single deep learning models and 30% for decomposition-based hybrid models.

INDEX TERMS Wind speed, forecasting, markov chain, moving window, statistical, wavelets.

ABBREVIATIONS

ACF	Auto Correlation Function
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
ARIMA	Auto-Regressive Integrated Moving Average
BIC	Bayesian Information Criterion
CEEMDAN	Complete Ensemble Empirical Mode Decomposition with Adaptive Noise
CNN	Convolutional Neural Network
Db	Daubechies
DBN	Deep Belief Network
DT	Decision Tree
DTMC	Discrete-Time Markov Chain
EL	Ensemble Learning
EMD	Empirical Mode Decomposition
FL	Fuzzy Logic
FS	Feature Selection
GM	Grey Model
GSR	Gaussian Process Regression
I-MODWT	Inverse-Maximal Overlap Discrete Wavelet Transform
KELM	Kernel-based Extreme Learning Machine
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MC	Markov Chain
MHHOGWO	Mutation-Harris Hawks Optimization-Grey Wolf Optimizer
MODWT	Maximal Overlap Discrete Wavelet Transform
MSE	Mean Square Error

The associate editor coordinating the review of this manuscript and approving it for publication was Ton Duc Do.

PACF	Partial Auto Correlation Function
PCHIP	Piecewise Cubic Hermite Interpolating Polynomials
PE	Permutation Entropy
PRBF	Principal Component Analysis-Radial Basis Kernel Function
RMSE	Root Mean Square Error
SSAPSR	Singular Spectrum Analysis-Phase Space Reconstruction
SVR	Support Vector Regression
VMD	Variational Mode Decomposition
WT	Wavelet Transform

I. INTRODUCTION

Wind speed forecasting models can be classified from two perspectives: (a) time scale (b) applied methodology. Based on the time scale, the forecasting models can be further divided into four categories: very short-term (0-30 min), short-term (30 min - 6 hours), medium-term (6 hours - 1 day ahead), and long-term (>1 day ahead). These time scales are defined based on regularity situations, end-user requirements, and technical conditions. For instance, wind turbine regulation and real-time grid operation require a very short-term forecast. Similarly, economic load dispatch planning and load decisions involve short-term predictions. Correspondingly, decision making of unit commitment and the reserved requirement is based on medium-term wind speed forecasts. Likewise, maintenance scheduling and feasibility study of wind farms need long-term forecasting. Literature reveals that shorter time scales forecasts are more detailed and accurate compared to long-term forecasts. However, a limited time is available for the deployment of wind power generation [1].

Based on the applied methodology, wind speed forecasting models can be divided into five categories: i) Persistence, ii) Physical, iii) Statistical, iv) Artificial Intelligence/Machine Learning (AI/ML), and v) hybrid. In the persistence model, the immediate-future wind speed is considered the same as the present wind speed.

Physical models are based on orography and numerical weather prediction model. Such models generate predictions based on initial conditions to solve the complex numerical system. However, two significant problems limit the applicability of these models in very short to short term wind speed forecasting: (i) The extensive information on the characteristics of wind farms is not always possible [2] (ii) Higher spatial resolution and continuously updated environmental information requires significant computational time [3], [4].

Statistical and AI/ML models are based on the inner relationship among historical data and do not require physical insight. Both types of models show higher accuracy in very short to short-term forecasts. Commonly used statistical models include Kalman Filter [5], Markov Chain [6], Auto-Regressive Integrated Moving Average (ARIMA) [7], generalized additive model [8], and grey prediction models [9]. Similarly, the traditional AI/ML models include Artificial Neural Network (ANN) [10], Support Vector Regression (SVR) [11], and Fuzzy Logic (FL) [12]. Statistical models deal with linear conditions, whereas AI/ML model has stronger nonlinear estimation ability.

Besides traditional AI/ML models, deep learning and extreme learning machines are also commonly applied in wind speed forecasting. Notable architectures include Kernel Extreme Learning Machine (KELM) [13], [14], Long Short-Term Memory (LSTM) [15], [16], Echo State Network [17], Deep Belief Network (DBN) [18], [19], and Convolutional Neural Network (CNN) [20], [21].

Considering the intrinsic weaknesses of all models, it is difficult for a single forecasting model to adequately capture the complex relationships of wind speed time series [17], [22]. Therefore, a better approach is to use hybrid methods where every model utilizes its individual capability. Such models are further classified as weighted models, feature selection models, decomposition (or preprocessing) models, and error-processing (or postprocessing) models [23], [24].

In the weighted models, multiple forecasting methods are utilized to forecast wind speed simultaneously. Then, each individual model is assigned a weight coefficient based on prediction performance. The weights are either fixed or variable. However, the variable weight arrangement has better performance [25]. Optimization algorithms such as Flower Pollination Algorithm with Chaotic Local Search (CLSFPA) [26], Bat Algorithm (BA) [25], Chaos Particle Swarm Optimization (CPSO) [24], and Multi-Objective Grasshopper Optimization Algorithm (MOGOA) [27] have been used to optimize the weights of final combination based on model effectiveness [28].

Feature selection and optimization models improve the model performance by removing the redundant data. Different optimization algorithms are mentioned in the literature. Zhang *et al.* [29] used the Improved Genetic Algorithm (IGA) to optimize Fuzzy Neural Network (FNN). Li *et al.* [30] considered the Improved Dragonfly Algorithm (IDA) to optimize Support Vector Machine (SVM). Zhang *et al.* [31] applied improved Particle Swarm Optimization (IPSO) to optimize Long-Short Term Memory (LSTM). In literature, two basic frameworks of feature selection (Wrapper and Filter) are discussed. The details for both frameworks are available in [1], [2].

In error-processing models, the effect of residuals is considered to improve the model performance. Firstly, the errors are analyzed after the primary prediction model. Next, a secondary postprocessing model is applied to improve the initial forecast. Duan *et al.* [32] decomposed the forecasted error time series using ICEEMDAN and then applied the ARIMA model to predict the error sequence. Several other models, such as LSTM [33], Markov Chain [29], and ELM [34], are also reported in the literature as error correction models.

The widely used hybrid models are decomposition-based [2], and more than 100 research articles have focused on these models in recent years [35]. As Li *et al.* [36]

discussed, it is difficult for various prediction models to accurately understand the wind data pattern. Wind speed time series is a combination of multiple frequencies. Also, it possesses a highly chaotic, nonlinear, and intrinsic nature. Therefore, the motivation behind preprocessing models is to improve prediction accuracy by segregating the time series with different frequencies. Such models are based on the 'divide and conquer' strategy for nonstationary time series and show better performance than the conventional single models.

In the decomposition-based hybrid model, the nonstationary wind speed time series decomposes into several relatively stationary subseries. Then, a forecasting model is applied to each subseries to get several individual predictions. Aggregating all the individual predictions results in the final forecast [1]. Commonly used decomposition approaches are Wavelet Transformation [36], Empirical Mode Decomposition [37], Variational Mode Decomposition [17], and their variants [38]. The individual prediction models can either be statistical, artificial intelligence, or both – for instance, WT-ARIMA [7], EMD-PE-ANN [37], VMD-PRBF-ARMA-E [39], EEMD-PSF-ARIMA [38], WT-VMD-DLSTM-AT [40], EWT-BiDLSTM [41]. The decomposition-based hybrid models show better forecasting accuracy than conventional models.

In general, AI/ML models are given preference for subseries predictions. However, commonly used AI models have issues of premature convergence and overfitting. Also, the advanced learning methods tend to fall into local optima. Such problems are not associated with statistical models that make them suitable for very short to short-term wind speed forecasting. Also, the decomposition of wind speed time series can make the input data more stationary. Once the stationary condition is met, the statistical models would generate competitive predictions [42]. Therefore, trained by the preprocessed data, the statistical models can learn the nonlinear behavior of the wind [1].

One of the widely recognized statistical prediction models for wind energy is Markov Chains (MCs). Besides their simplicity, the advantage of MCs is their ability to model wind time dependence characteristics. MCs are based on the probability distribution, which shows that the wind speed at the following time step relies on the present wind state. Other commonly used statistical models cannot capture this probability dependence [43]. Table 1 summarized the motivation and limitation of studies implementing the MC model for wind energy forecasting. Although the traditional MC model showed better performance than other benchmark models [44], the method is computationally extensive with limited performance for a large range dataset. One primary reason is a poor state categorization method.

The conventional state categorization implemented in [6], [45]–[47] generates a huge size of a Transition Probability Matrix (TPM) that increases the computational complexities. Furthermore, the individual number of certain states could be much lower such that it shows near-zero probabilities in TPM. Therefore, efforts are made to improve the state categorization.

One way to overcome the issue of huge TPM is to reduce the number of states. Sahin and Sen [48] introduced interval boundaries as $\bar{V} \pm l\sigma$ where$\bar{V}$ is the mean, σ is standard deviation of available wind speed, and $l = 0, 1, 2, 3, \ldots$ until extreme in wind speed data. Yang *et al.* [49] also applied the same approach and defined states as $\bar{V} \pm l(0.4\sigma)$. This method significantly reduced the number of states. However, the forecasting accuracy is compromised. According to the analysis of Tang *et al.* [50], the wind speed generation with the uniform distribution assumption is worse in the latter case. The value of RMSE is increased from 0.007 (traditional method) to 0.010 (equal interval boundaries).

Another approach to improve state categorization is to use unequal intervals. Ettoumi *et al.* [51] defined the interval boundaries based on wind speed distribution as weak wind (0–3 m/s), mean wind (3–8 m/s), and strong wind (>8 m/s). Similarly, empirical quantiles are considered in [50] to construct the intervals. However, the unequal intervals method, defined in terms of the empirical probability distribution function, failed to improve the accuracy of the traditional method for a shorter horizon.

Other than the traditional model, the improper state categorization also compromised the accuracy of enhanced models such as nested MC [43], [52] and non-homogenous MC [53]. The accuracy improvements of nested MC are negligible for a higher number of states. Similarly, the state interval for non-homogenous MC is suggested as 0.5 m/s and 1 m/s, which is computationally extensive. Therefore, the discretization of the data into a proper number of states still needs improvements.

Another parameter that limits the forecasting accuracy of the traditional model is the size of modelling data (hereafter called window size). Carpinone *et al.* [47] analyzed the sliding window on a wind power time series and concluded that an optimal size of a sliding window is required to achieve a satisfying accuracy. Furthermore, a proper rolling window size also helps to prevent the MC from being stationary in time due to seasonality [44]. He *et al.* [54] constructed finite-state MC considering data of three hours and for each individual month. In this way, the diurnal non-stationarity and the seasonality of wind time series are accounted without complex models. The same conclusion is also inferred in [55], [56].

In [6], the optimal size of the sliding window is selected as 4320 based on prediction accuracy. Also, Yoder *et al.* [44] analyzed the window sizes of 45, 60, 90, 180 days and concluded that 180 day rolling window obtained the best estimates. In other cases, the optimal window size is selected as 4 [57], 100 [58], and 300 [59]. Although such models show good performance. However, a single moving window selection is not the best option for wind speed predictions. See, for example, Fig 5(a), where a window size of 4320 is only averaging the wind speed. To the best of the authors' knowledge, the concept of a dynamic moving

TABLE 1. Studies implementing Markov chain model for wind energy forecasting.

Models	Studies	Motivation	Limitation
First Order Markov Chain	[49, 52]	First-Order MC is simple and can model wind time dependence characteristics. The Markovian wind models showed significant improvement over the simple Monte Carlo approach without any temporal correlation.	Due to the memoryless property, the model only holds short-term autocorrelation.
Higher Order Markov Chain	[6, 46, 61-65]	The higher-order Markov chain considers multiple time step memories to improve the prediction accuracy.	These models are computationally demanding. For instance, a 3rd order MC with 25 states requires 15,625 state transition probabilities. Increasing order is the same as increasing features. It may cause overfitting as the model begins to fit nongeneralized attributes of the training data.
Nested Markov Chain	[44, 53]	One MC represents longer-term variation, whereas an additional MC captures high-frequency variation.	The model showed negligible improvements in terms of model accuracy with a higher number of states.
Non-homogenous Markov Chain	[54]	To include the effects of daily and seasonal variation.	The model limits its accuracy with large state intervals. The proposed model is computationally extensive as the suggested interval boundaries are in the range of 0.5 – 1 m/s.
State Categorization Method			
Interval of 0.5 – 1 m/s	[44, 46, 61, 66-69]	To increase the dimension of the state space for higher accuracy.	A large number of states generates a huge size of a Transition Probability Matrix (TPM) that increases the computational complexities. Furthermore, the individual number of certain states could be much lower such that it shows near-zero probabilities in TPM.
Equal Intervals	[49, 70]	To introduce the interval boundaries as some correlation to reduce the number of states.	It might be possible to select an inappropriate distribution. Thus, a big deviation is induced between the actual and generated time series. Hence, the forecasting accuracy is compromised.
Unequal Intervals	[51, 52, 63]	To construct state space with a small number of states with unequal intervals.	The unequal intervals method defined in terms of empirical probability distribution function failed to improve the accuracy of the traditional method for a shorter horizon.
Hybrid Model			
GM-Markov ANN-Markov Linear prediction-Markov LSSVM-Markov ARIMA-Markov CEEMD-IGA-FNN-Markov	[71] [68] [64] [70] [72] [30]	MC can accurately predict large random fluctuations. Therefore, MC is utilized as an error correction model.	None of the decomposition-based hybrid models analyzed MC as a primary model
Other			
Window Size	[6, 45, 48, 55-57, 64]	The optimal size of a sliding window is required to achieve satisfying accuracy. Furthermore, the rolling window of the most recent observation occurring immediately before the target forecast prevents the MC from being stationary in time due to seasonality.	Wind speed time series exhibit an intermittent nature. The historical data does not possess an equal length of behavior for all horizons. Thus, the prediction errors tend to increase while considering a single-window size.

window is not applied in any of the literature mentioned above.

A. MOTIVATION AND CONTRIBUTIONS

Based on the above discussions, three problems are identified with the traditional model:

1. State Categorization: The first problem is how to discretize the historical time series data. The state categorization is somewhat arbitrary and depends on the purpose [50]. If the range is divided into small interval boundaries, then a huge TPM is generated. In this case, the computational complexity might increase with a high number of near-zero probability states in TPM. Similarly, if the number of states is reduced with large interval boundaries, then the forecasting accuracy is compromised. Therefore, a balance is to be maintained between computational complexity and forecasting accuracy.
2. Window Size: The second problem is how much data required to capture the wind speed variation. Wind speed time series exhibit an intermittent nature. The historical data does not possess an equal length of behavior for all horizons. Thus, the prediction errors tend to increase while considering a single-window size. Therefore, a dynamic moving window is required that should not be arbitrarily chosen [7].
3. Decomposition-based hybrid model: Even though decomposition models are extensively studied for wind forecasting [35], [72], very few studies are based on decomposition-based MC models. It is evident from the literature that MC models are suitable for predicting residual sequences, such as LSSVM-Markov [69], ARIMA-Markov 72], and CEEMD-IGA-FNN-Markov [29]. However, MC models were not considered as a primary model in any of the literature mentioned above.

To address the first problem, a self-adaptive state categorization algorithm is developed. The algorithm itself decides the interval boundaries (equal and non-equal) using rounding and array functions. Such an adaptive method has not been discussed in any of the literature.

To address the second problem, a similar approach to the leave-one-out method is introduced to select the optimal size of the moving window at every time step. To the best of the authors' knowledge, the concept of a dynamic moving window is not applied in any of the literature mentioned above. This would also help in generalizing the model for any case study. We compared the accuracy of the proposed model with already developed models for the case study of the Sotavento Windfarm. The results proved the robustness of the proposed model.

As mentioned earlier, a single forecasting model cannot adequately capture the complex relationships of wind speed time series. In contrast, the hybrid model can utilize the individual capability of each model. Therefore, to address the third problem, a novel decomposition hybrid model is proposed based on the maximal overlap discrete wavelet transform (MODWT), autoregressive integrated moving average (ARIMA), and modified Markov Chain model.

The major contributions of this study are given as follows:

1. A single forecasting model is proposed that integrates the Markov process with an adjusting optimal moving window for very short to short-term wind speed forecasting. The optimal window size is selected at every time step based on historical data. A similar approach to the leave-one-out method is introduced to select the optimal size of the window. A comprehensive analysis is provided to compare single and adjusting moving window approaches.
2. The state-categorization of traditional MC is improved by introducing a self-adaptive algorithm to optimize the numbers of transition states. Also, the proposed model directly forecasts one step ahead wind speed instead of generating synthetic wind speed data [73], [53].
3. The performance of the modified Markov Chain model is compared with six other statistical, AI/ML, and deep learning models. These models include decision tree, ensemble learning, GSR, SVR, Grey model, and LSTM
4. Based on adjusting moving window Markov chains, a hybrid MODWT-ARIMA-Markov model is proposed for wind speed forecasting. In this model, the maximal overlap discrete wavelet transform (MODWT) decomposes the time series into several relatively stationary subseries. Next, autoregressive integrated moving average (ARIMA) and adjusting moving window Markov chains are applied to achieve individual predictions. Finally, Inverse MODWT is applied to get the final forecast.
5. The performance of the proposed hybrid model is compared with nine other AI/ML, deep learning, and hybrid models based on three evaluation metrics. These models include SVR, KELM, LSTM, ConvLSTM, EMD-KELM, EMD-ConvLSTM, CEEMDAN-KELM, CEEMDAN-SSAPSR-KELM, and CEEMDAN-SSAPSR-FS-KELM-MHHOGWO.

The remainder of this paper is organized as follows: Section II presents the methodology of the adjusting moving window Markov chain model. The optimizations are discussed in the fundamental knowledge of the traditional model. Also, MODWT-ARIMA-Markov is introduced in this section. Section III shows the experiments and analysis of the proposed forecasting system. The case study of Palmerston North, New Zealand, is used to compare the single and adjusting moving window approaches comprehensively. Next, the proposed MODWT-ARIMA-Markov model is compared with models available in the literature for the case study of the Sotavento wind farm, Spain. Finally, the conclusions are presented in Section IV.

II. METHODOLOGY

The current methodology integrates the adjusting moving window with the Discrete-Time Markov Chain (DTMC).

Instead of considering a single-window size throughout the time series proposed in [6], [44], [47], [54]–[56], [63], the window size is updated at every time step.

The proposed methodology is a two-step forecasting model. The size of the moving window is selected based on a similar approach to the leave-one-out method. The available sequence of n observation is divided into $(n-1)$ training and the last value as validation datasets. The observation of validation dataset (V_n) is forecasted based on training dataset observations $(V_T)_{T=1}^{n-1}$. For a window size of $N=1$, the prediction is based on only previous wind speed V_{n-1}. In this case, the modified model behaves as a persistence model. The predicted wind speed is stored in memory. Then, more previous training data is included for increasing window size $(V_{n-d})_{d=1}^{N}$. The maximum value of window size is $N_{max}=n-1$. The final prediction vector is comprised of $(n-1)$ values. As the observation V_t is available; therefore, the optimal window size is selected based on mean square error (MSE). MSE measures the average squared difference between the actual and the predicted value. The lower the MSE value, the better the model prediction. In this manner, the optimal window size is adjusted for every time step. The optimal window size is then used to predict one step ahead wind speed $\hat{V}_{t+1}$. The overall selection process is provided in Fig. 1.

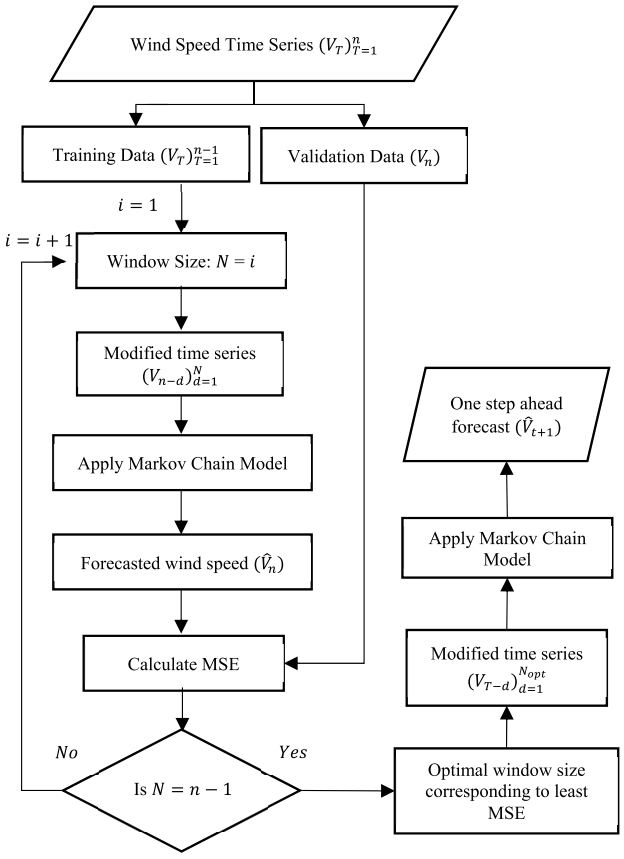


FIGURE 1. Selection of optimal window based on a similar approach to the leave-one-out method.

In contrast with the leave-one-out method, the procedure runs only once in each time step, and the last data point is always used for validation and selection.

The Markov Chain (MC) is a stochastic process that sequentially moves from one state to another in the state-space. State-space (S) is a set of values that a chain can take $\{s_t\}_{t\geq 0}$. MC consists of a state-space and a transition matrix. A transition matrix determines the probability of moving from one state to another. Forecasting through MC is summarized in the following five steps.

STEP 1) STATE CATEGORIZATION

The first step is to define states for the Markov chain process. The state categorization is rather arbitrary and depends on the purpose. If the range is divided into small interval boundaries, then a huge TPM is generated. In this case, the computational complexity might increase with a high number of near-zero probability states in TPM. Furthermore, the problem of overfitting may occur if more states are included than are supported by the input data. Similarly, if the number of states is reduced with large interval boundaries, then the state transition probability is significantly reduced. In this case, the forecasting accuracy is compromised. To identify a proper number of states, a self-adaptive algorithm is introduced in the present study. The algorithm itself decides the interval boundaries (equal and non-equal) using rounding and array functions based on historical time series at every time step. Such an adaptive method can generate only required states, thus avoid overfitting.

For state-space, the floating values of minimum and maximum wind speed is converted to an integer as $ceil\left\{(V_{t-d})_{d=1}^{N}\right\}$ whereas the remaining sequence as $floor\left\{(V_{t-d})_{d=1}^{N}\right\}$. Next, the state-space is constructed based on wind speed state boundaries as $0, ceil(V_{min}), sort(V_I, ascend), ceil(V_{max})$, with no repeating values. For example; if a sequence is $V=\{4.5, 3.9, 5.3, 5.9, 5.8, 5, 4.9, 5.5, 5.3, 5.9, 6.1, 5.5, 6.3, 6, 6, 5.3\}\ m/s$, then the state value for a floating number of minimum wind speed 3.9 m/s is taken as $ceil(V_{min})=4m/s$. The same procedure followed for maximum wind speed as well, i.e., for $\{6.1, 6.3\}\,m/s$ the value of the state is $ceil(V_{max})=7m/s$. Whereas, for the remaining sequence, $V_I=\{4.5, 4.9, 5.3, 5.5, 5.8, 5.9\}$, state value is taken as $floor(V_I)=\{4, 5\}\,m/s$. Based on these values, the wind speed state boundaries are set to 0, 4, 5, and 7 m/s, respectively. From the traditional method [43], [45], [60], [65]–[68], the state-space with an equal distance of 1 m/s would be 0, 1, 2, 3, 4, 5, 6, and 7 m/s respectively, which are much higher than the proposed method. In this manner, the state-space is updated every time step based on the time-series sequence of the optimal moving window.

STEP 2) STATE TRANSITION MATRIX

Next is to construct the Transition Probability Matrix (TPM). Let $\{X_t\}_{t\geq 0}$ be a sequence of discrete random numbers. As per

definition, the sequence $\{X_t\}_{t\geq 0}$ is an MC if it satisfies the following equality:

$$P\{X_{t+1}=j|X_t=i,X_{t-1}=i_{t-1},\ldots,X_o=i_o\} = P\{X_{t+1}=j|X_t=i\}=p_{ij}, \quad (1)$$

for all $t=1,2,3,\ldots$ and for all states i and j.

Equation (1) shows that only the most recent data of sequence affects what happens next. i.e., the probability of X_{t+1} depends only on X_t and not upon $X_{t-1},\ldots,X_1,X_o$. This is termed as Markov property. From the equation, MC is (time) homogenous as transition probabilities do not depend on time-shifting. The $s\times s$ matrix that describes the probability of the observed frequency of transition, i.e., jumps from one state to another state, is called a transition matrix (P). Thus for s possible states,

$$P=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s}\\ p_{21} & p_{22} & \cdots & p_{2s}\\ \vdots & \vdots & \ddots & \vdots\\ p_{s1} & p_{s2} & \cdots & p_{ss}\end{bmatrix}, \quad (2)$$

where $p_{ij}=\frac{m_{ij}}{\sum_{j=1}^{s} m_{ij}}$ and for all i and j.

The transition matrix has two properties. All the elements of P are non-negative, i.e., $0\leq p_{ij}\leq 1$, and the sum of every row is unity i.e. $\sum_{j=1}^{s} p_{ij}=1$. m_{ij} is the number of transitions from states i to j. The matrix defined in the equation is the first-order one-step transition matrix. A typical transition graph of P is shown in Fig. 2.

For k step, the transition probability matrix follows the following property,

$$P\{X_k=j|X_0=i\}=P\{X_{k+n}=j|X_n=i\}=p_{ij}^{k}. \quad (3)$$

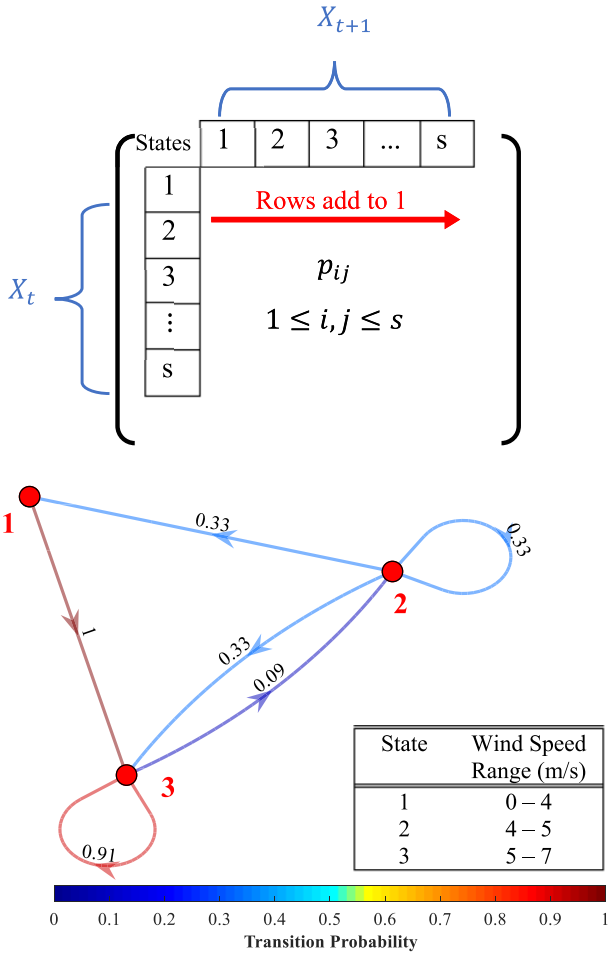


FIGURE 2. Properties of a transition probability matrix with a transition graph of sequence discussed in the state categorization step.

STEP 3) CONFIRMATION OF ERGODIC PROPERTIES

Once TPM is constructed, the third step is to confirm the ergodic properties of a Markov chain. An MC is ergodic if it is irreducible, and its states are aperiodic and positive recurrent [74]–[76].

1) DEFINITION

An MC is said to be irreducible if it has only one communicating class. Two states i and j are said to communicate if they are accessible from each other. Communication ($i\leftrightarrow j$) is an equivalence relation, i.e.

i. $i\leftrightarrow i$ for all $i\geq 0$ (reflexivity),
ii. If $i\leftrightarrow j$, then $j\leftrightarrow i$ (symmetry),
iii. If $i\leftrightarrow j$ and $j\leftrightarrow k$, then $i\leftrightarrow k$ (transitivity).

2) DEFINITION

A Markov chain that has a period (d) equals to one for every state is termed as aperiodic. Mathematically, a period d_i of state i is defined as,

$$d_i=gcd\left\{n\geq 1; p_{ii}^{(n)}>0\right\}, \quad (4)$$

if there is no $n\geq 1$ with $p_{ii}^{(n)}<0$.

3) DEFINITION

A state i is termed as recurrent if the process returns in state i with probability one. Mathematically, state i is recurrent if and only if,

$$\sum_{n=1}^{\infty} p_{ii}^{(n)}=\infty. \quad (5)$$

A recurrent state is positive if the return time has a finite expected value. For irreducible MC, if i is recurrent, then j is also recurrent.

Next is to construct a state transition probability vector. X_t is a random variable, and hence, it has a probability distribution. Consider a vector $\Pi_t=\{\pi_1,\pi_2,\ldots,\pi_s\}$denoting the probability distribution of the chain at time t. For the ergodic MC, the following assertions hold [75]:

4) DEFINITION

The limit

$$\pi_i=\lim_{n\to\infty} P\{X_n=i\}=\frac{1}{\mu_i}, \quad (6)$$

exists and is independent of the initial distribution. μ_i identifies the mean return time to state i [76].

5) DEFINITION

The stationary distribution is unique and satisfies a system of linear equations, i.e.

$$\sum_{i=1}^{s} \pi_i = 1 \text{ where } \pi_i \geq 0, \tag{7}$$

$$\pi_i = \sum_{j=1}^{s} \pi_j p_{ji}. \tag{8}$$

STEP 4) STATE TRANSITION PROBABILITY VECTOR

In the fourth step, the state transition probability of the future state is calculated based on the initial state probability vector (Π_0) and state transition matrix (P), as:

$$\Pi_1 = \Pi_0 \times P,$$
$$\Pi_2 = \Pi_1 \times P = \Pi_0 \times P^2.$$

Therefore, for k steps,

$$\Pi_k = \Pi_0 \times P^k. \tag{9}$$

All the elements of Π_0 is zero except the element corresponding to the current state at time t.

STEP 5) FORECASTING THE WIND SPEED

Finally, one step ahead wind speed ($\hat{V}_{t+1}$) is estimated based on state transition probability vector and mean wind speed values of the state ($\bar{S}$) as in equation (10) [6]:

$$\hat{V}_{t+1} = \sum_{i=1}^{s} \Pi_i \bar{S}_i. \tag{10}$$

In this study, a hybrid MODWT-ARIMA-Markov is proposed implementing the improvements suggested in the traditional model. The maximal overlap discrete wavelet transform (MODWT) decomposes the time series into stationary subseries in the proposed framework. Then autoregressive integrated moving average (ARIMA) and adjusting moving window Markov chains are applied to achieve individual one-step ahead forecasts.

MODWT is used for the multilevel decomposition of wind speed time series into subsequence signals in different frequency bands. In comparison with the Discrete Wavelet Transform (DWT), MODWT is highly redundant and non-orthonormal. In DWT, the sample size is required to be the multiple of 2^j where j is the level of decomposition. This condition is not required in MODWT. MODWT is defined for all sample sizes with no restrictions of an integer multiple of 2^j. The number of scaling and wavelet coefficients in DWT decrease by a factor two at each level of decomposition due to the decimation effect. This might introduce ambiguities in the time domain. On the other hand, MODWT addressed the issue of decimation and is known as undecimated WT. The down-sampling process can be avoided using MODWT, allowing the same number of scaling and wavelet coefficients as observation size [77], [78].

ARIMA models (also known as Box-Jenkins models) are the most commonly used statistical models. The combination of ARIMA and Markov Chain models is proved to provide improved forecasting results [71]. ARIMA model captures the fluctuations, whereas randomness is predicted through Markov Chain. The linear expression for generalized non-seasonal model structure form of ARIMA (p, d, q) is given as:

$$y_t = c + \left(\sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}\right) \tag{11}$$

where p and q is the order of AR and MA part, d is the degree of differencing to form stationary time-series, and ϕ_i and θ_j are the coefficients of the i^{th} AR and j^{th} MA parameters, respectively. y_{t-i} represents the value at a time $(t-i)$, ε_{t-j} shows the error between the measured and predicted values at $(t-j)$, and c is the constant term. The prediction through ARIMA is a three-step iterative process, as described below:

The first step is to identify the model order. The prerequisite of this step is to check the stationarity of the time series. The simple way is to visualize the scatter plot. Another way to check the stationarity is the unit root test, such as the Dickey and Fuller test. If a time series is nonstationary, then the preliminary step is to convert the nonstationary time series into stationary by taking the backward $(B)d^{th}$ difference as $B^d V_t = V_{t-d}$. Once the stationarity of the time series can be presumed, then the sample Auto Correlation Function (ACF) and Partial ACF (PACF) should be obtained in the next step. The plots of ACF and PACF would help in identifying the model parameters. The behavior of theoretical ACF and PACF for stationary time series is given in Table 2.

TABLE 2. Theoretical ACF and PACF for stationary time series.

Model	ACF	PACF
MA(q)	Cuts off after lag q	Exponential decay and/or damped sinusoid
AR(p)	Exponential decay and/or damped sinusoid	Cuts off after lag p
ARMA (p, q)	Exponential decay and/or damped sinusoid	Exponential decay and/or damped sinusoid

Next, the parameters are estimated through the maximum likelihood method.

The third step is to check the model adequacy conditions to ensure the model performance. If the identified model is adequate, the residual values behave like a white noise process.

For the proposed MODWT-ARIMA-Markov model, the Markov chains are applied as a primary integrated model and not as a residual correction. The prediction model is summarized into four steps as follows:

STEP 1) DECOMPOSITION

The Daubechies wavelet (Db2) is used to obtain MODWT. Daubechies wavelet is mainly considered in the literature as it has a large vanishing point. The required minimum decomposition level is determined as [79]:

$$L = int\,[log\,(n)]\,,$$

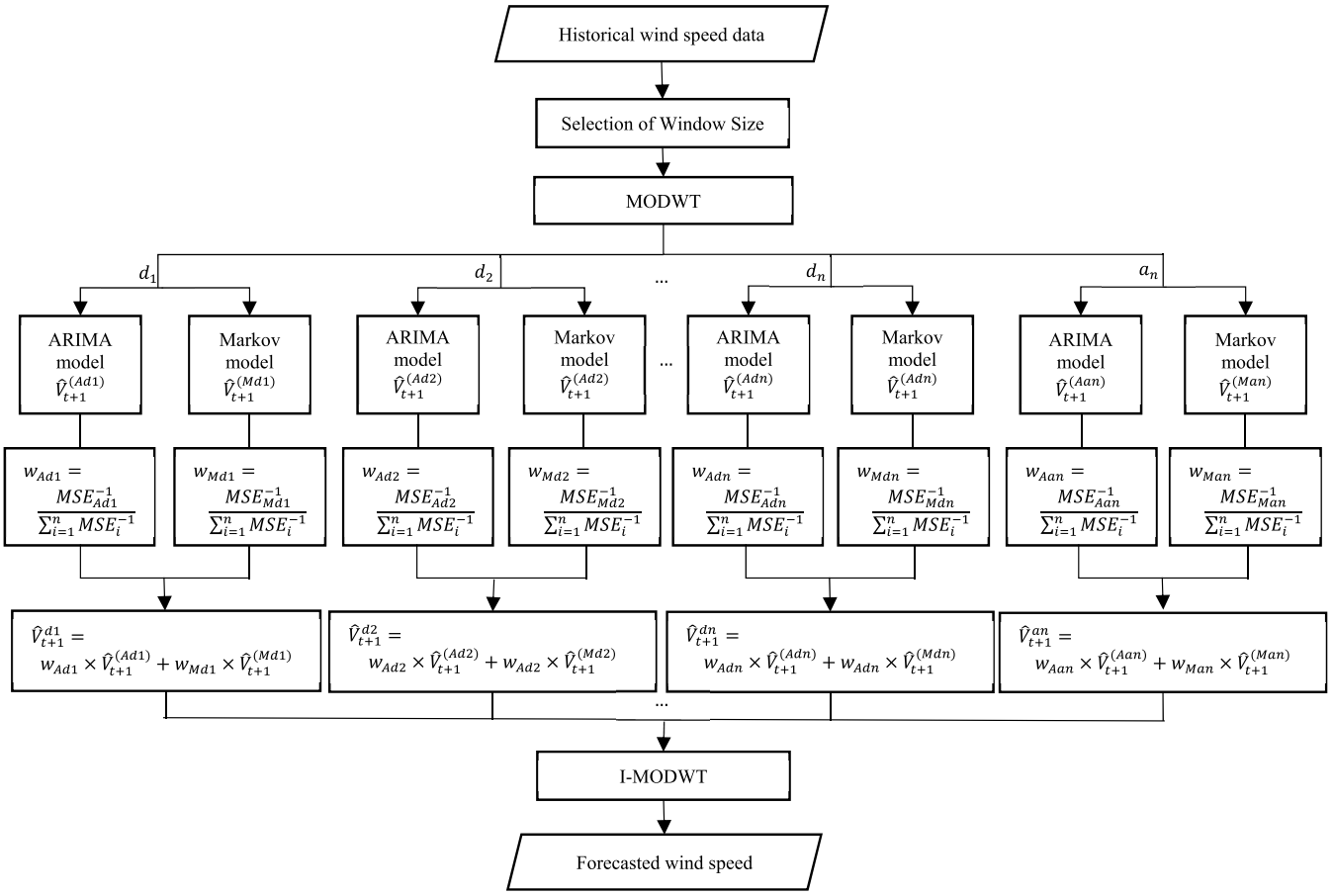


FIGURE 3. The flowchart of the proposed hybrid MODWT-ARIMA-Markov model.

where n is the number of observations. The detailed and approximate subseries are denoted as $d_1, d_2, \ldots, d_n$ and a_n respectively.

STEP 2) PREDICTION WITH THE ARIMA MODEL

The best ARIMA order for the detailed $(d_1, d_2, \ldots, d_n)$ and approximate subseries (a_n) are estimated separately. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are selected as deciding parameters. The predicted ARIMA series is denoted as $\hat{V}_{t+1}^{(Ad1)}, \hat{V}_{t+1}^{(Ad2)}, \ldots, \hat{V}_{t+1}^{(Adn)}, \hat{V}_{t+1}^{(Aan)}$.

STEP 3) PREDICTION WITH MODIFIED MARKOV CHAIN MODEL

Similar to the previous step, the Markov chain model is applied based on adjusting the moving window. If the signal lies between a single state, the time series is multiplied by the factor 10^P where P is $abs[ceil\{\log(V_{min})\} - 1]$ to increase its amplitude. The finalized time series is then divided by the same factor. Similarly, for the negative values of decomposed signals, the analysis is carried out for absolute data. The predicted Markov series is denoted as $\hat{V}_{t+1}^{(Md1)}, \hat{V}_{t+1}^{(Md2)}, \ldots, \hat{V}_{t+1}^{(Mdn)}, \hat{V}_{t+1}^{(Man)}$.

STEP 4) RECONSTRUCTION

The weights are calculated on an error-based combination method as $w_i = MSE_i^{-1} / \sum_{i=1}^{n} MSE_i^{-1}$. For example, the weight coefficients of d_1 are calculated as: $w_{Ad1} = MSE_{Ad1}^{-1} / \sum_{i=1}^{n} MSE_i^{-1}$ and $w_{Md1} = MSE_{Md1}^{-1} / \sum_{i=1}^{n} MSE_i^{-1}$.

Next, the individual forecast is evaluated as $\hat{V}_{t+1} = \sum_{i=1}^{P} w_i \hat{V}_{t+1}^{(i)}$. The weights are assumed to be unbiased and non-negative such that $\sum_{i=1}^{P} w_i = 1$. Finally, the predicted subseries are reconstructed through I-MODWT.

The overall methodology of the prediction model is presented in Fig. 3.

III. EXPERIMENTS AND ANALYSIS

In this section, one-step-ahead wind speed prediction based on single and adjusting moving window is analyzed based on wind speed data of Palmerston North, New Zealand. The website considered in [6] is updated, and the dataset is not available. Next, the proposed MODWT-ARIMA-Markov model is applied to the case studies available in the literature [80].

In this study, we are considering only wind speed as an input variable because of two reasons. Firstly, the

TABLE 3. Summary of meteorological parameters at Palmerston North, New Zealand.

Months	Meteorological parameters				
	$\bar{V}$ (m/s)	$\bar{T}$(°C)	$\bar{H}$(MJ/m^2/day)	$\bar{P}$(hPa)	$\overline{RH}$(%)
Jan	4.39	10.3	22.4	14.9	75.3
Feb	4.33	10.3	19.9	15.3	77.7
Mar	4.25	10.2	15.4	14	79.4
Apr	3.58	9.9	10.6	12.7	81.2
May	3.81	8.9	7	11.4	85.8
Jun	3.81	8.2	5.3	9.7	86.8
Jul	3.86	8.3	6.1	9.3	86.8
Aug	3.94	8.5	8.7	9.7	84.6
Sep	4.33	8.5	12.3	10.7	79.7
Oct	4.72	8.6	15.7	11.6	80.5
Nov	4.94	9.1	19.8	12.2	76.7
Dec	4.47	9.4	21.1	14	76

Acronyms

$\bar{V}$	Monthly average wind speeds (m/s)
$\bar{T}$	Average daily temperature range (°C)
$\bar{H}$	Monthly average daily global solar radiation (MJ/m^2/day)
$\bar{P}$	Monthly average 9 am vapor pressure (hPa)
$\overline{RH}$	Monthly average 9 am relative humidity (%)

meteorological parameters are not readily available for all sites. See [81] as an example. Also, sharing the SCADA data is against the confidentiality policy of many firms. Secondly, in this study, the proposed model is also compared with already developed algorithms. Therefore, the analyses of the same datasets present a fair comparison between developed and proposed models.

A. DATA SET DESCRIPTION

Wind speed data for Palmerston North is collected from the National Institute of Water and Atmospheric Research Ltd (NIWA) for the summer season, with a time interval of 10 minutes from December 1 2017 to February 28 2018 [82].

Palmerston North (latitude: 40.382° S, longitude: 175.609° E, height: 21m) has a huge amount of strong winds ($> 8.6\ m/s$) with west-northwest as dominant wind direction [9], [83]. From recorded data, the strong winds occurred 18% in autumn, 19% in winter, 26% in summer, and 37% in spring. For the selected site, the strong gust ($> 26\ m/s$) is infrequent. The other meteorological parameters are summarized in Table 3. In this study, the model performance is evaluated based on Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) and are given in equations (12) to (14).

$$MAE = \frac{1}{n}\sum_{z=1}^{n}\left|\hat{V}(z) - V(z)\right| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{z=1}^{n}\left(\hat{V}(z) - V(z)\right)^2} \tag{13}$$

$$MAPE = \frac{1}{n}\sum_{z=1}^{n}\frac{\left|V(z) - \hat{V}(z)\right|}{V(z)} \times 100\% \tag{14}$$

where $V(z)$ and $\hat{V}(z)$ are the actual and predicted values, and n is the number of observations. The lower values (closed to zero) are preferable for selected statistical indicators to ensure higher prediction accuracy. Firstly, the raw data is analyzed for any inaccuracies. Only 30 data points (out of 12,960) were identified as missing or outlier from the boxplot analysis and are imputed using Piecewise Cubic Hermite Interpolating Polynomials (PCHIP). The ratio of the training to test data is set at 3:1. Therefore, the training data includes 9,072 observations, whereas the testing data has 3,888 values. The finalized wind speed time series is shown in Fig. 4.

B. COMPARISON OF SINGLE MOVING WINDOW WITH ADJUSTING MOVING WINDOW

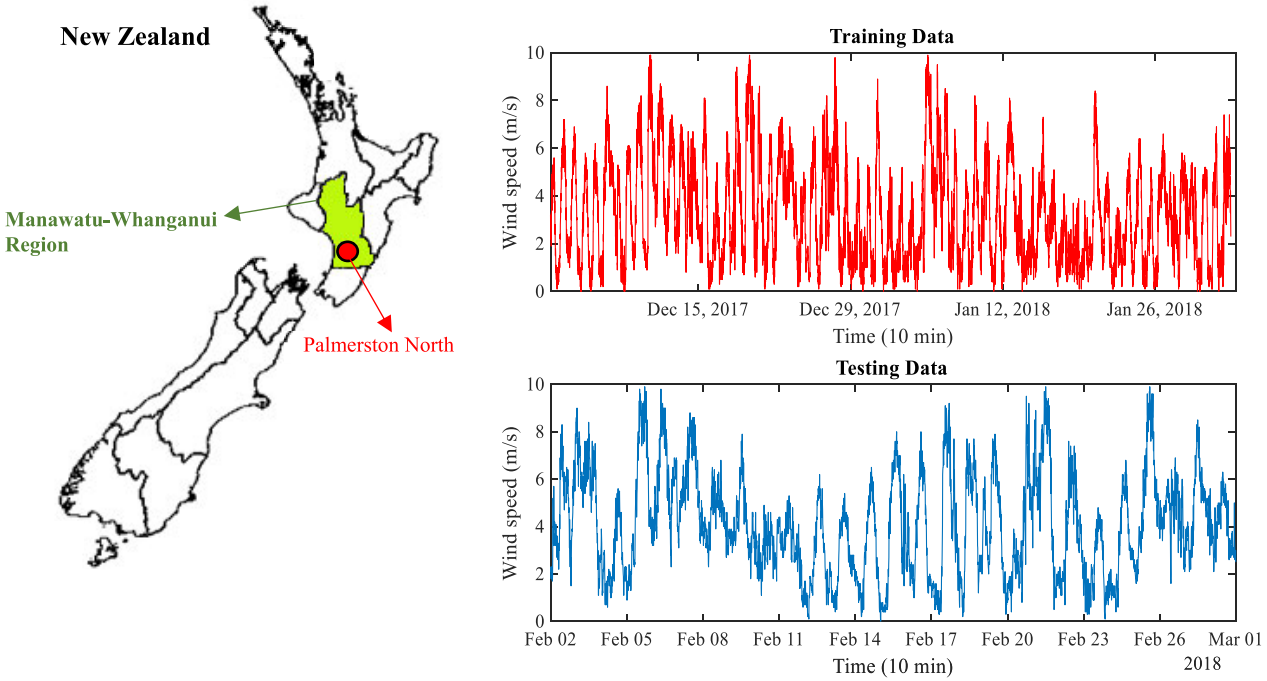
Fig. 5(a) shows one step ahead of wind speed forecasting with the testing dataset. The prediction results show excellent agreement with the measured data with $R = 0.94$. Fig. 5(b) shows that 81.3% of the data requires a maximum of a prior week's observation. Also, more than half of the values require only 100 values (∼17 hours). These are much lesser data points, as considered by Carpinone *et al.* [6]. When a large moving window is considered, as shown in Fig. 5(a) for N = 4320, i.e., 30 days, the model averages the wind speed that eventually decreases the model performance instead of improving model accuracy. Thus, the recent observations are not much influenced by distant data, and a choice of smaller window size yields more accurate predictions. This analysis is also justified in defining a relatively small number of states to avoid very low probabilities for certain states.

Another alternative is to select a smaller single moving window. Based on Fig. 5(b), a moving window varying from 100 to 1000 is selected, and the statistical results are presented in Fig. 6. After removing erroneous predicted values, if any, the results of statistical indicators are plotted against window size. For a single moving window, the curve behavior shows that an increasing window improves the model performance reaching an optimum value and then decreases. Based on the results, the best-suited window is 200 with the least MAE and RMSE values of 1.3262 m/s and 1.6557 m/s, respectively.

However, in contrast with the modified model, the single optimal moving window still has lower performance, as shown in Fig. 6. The maximum correlation coefficient ($R = 0.61$) shows that the prediction results are not in good agreement with the measured data for any case of a single moving window. Also, from Table 4, if a single and updating moving window is compared for N=200, then the proposed model is relatively 54% better than a single moving window.

Among all, the proposed model with window size equivalent to the length of the training dataset is the best with the least MAE and RMSE values of 0.4875 m/s and 0.6501 m/s, respectively.

It is observed from the analysis of Fig. 5(b) that the larger training dataset will result in better model performance. It might be confusing that a larger input dataset is required as approximately 6.5% of the results are based on 30 to 60 days observations; however, this is not the case. From Table 4, it is observed that the absolute difference in MAE and RMSE between adjusting the moving window based on one week (∼1000 data points) and two months (∼9000 data points)



Dataset	No. of observations	Statistical Information	
		Mean (m/s)	Standard Deviation (m/s)
All samples	12,960	3.5171	2.1490
Training	9,072	3.2658	2.0996
Testing	3,888	4.1036	2.1487

FIGURE 4. The detailed information of the experimental dataset.

TABLE 4. Comparative analysis of single and adjusting moving window.

Max Window Size (N_{max})	Single moving Window		Adjusting moving Window	
	MAE (m/s)	RMSE (m/s)	MAE (m/s)	RMSE (m/s)
100	1.3717	1.6762	0.5714	0.7755
200	1.3262	1.6557	0.5089	0.6763
300	1.5896	1.9595	0.5022	0.6673
400	1.7076	2.0756	0.5001	0.6626
500	1.6882	2.0388	0.4995	0.6625
600	1.7157	2.0758	0.4988	0.6617
700	1.7727	2.1424	0.4984	0.6615
800	1.7486	2.1253	0.4974	0.6604
900	1.7634	2.1448	0.4960	0.6588
1000	1.8005	2.1978	0.4953	0.6581
length of the training dataset	--	--	0.4875	0.6501

observations is 0.0085 m/s (1.62%) and 0.0088 m/s (1.22%), respectively. This shows that prior week information with 10 min resolution would be enough to ensure better performance. Therefore, the seasonality of the wind could not affect the model performance.

Furthermore, as seen from Fig. 5(b), around 68% of the forecasting results are based on only 100 data points. Out of which, 48% of the predictions required a maximum of 20 historical values. For such a short interval, the non-homogenous MC model might only increase the model complexity. Therefore, a dynamic moving window is a better approach for short intervals.

The modified Markov model is further compared with six other statistical and machine learning models for three window sizes. Table 5 shows that the size of training data highly influenced the performance of machine learning algorithms. For example, the MAE of the GSR model increased from 0.437m/s to 0.615m/s ($\sim$ 41%) when N varied from 9072 to 1000. The same problem occurred for every machine learning model. In comparison, MAE of modified MC model only increased 1.64% i.e. from 0.487m/s to 0.495m/s for the same case. As a matter of fact, a machine learning model results in a different solution whenever it is trained. Models trained on the same input can give different outputs. It is due to the different initial weights and bias values. Therefore, the black box models need to retrain several times. In contrast, the statistical models do not exhibit such an issue.

Besides the excellent model performance and less complexity, the first-order Markov Chain model is also helpful to avoid overfitting issues. From the analysis of [84], overfitting occurs in higher-order chains. The higher-order MC models begin to fit attributes of the training data that are not general. Hence, the modified model has no problem of overfitting.

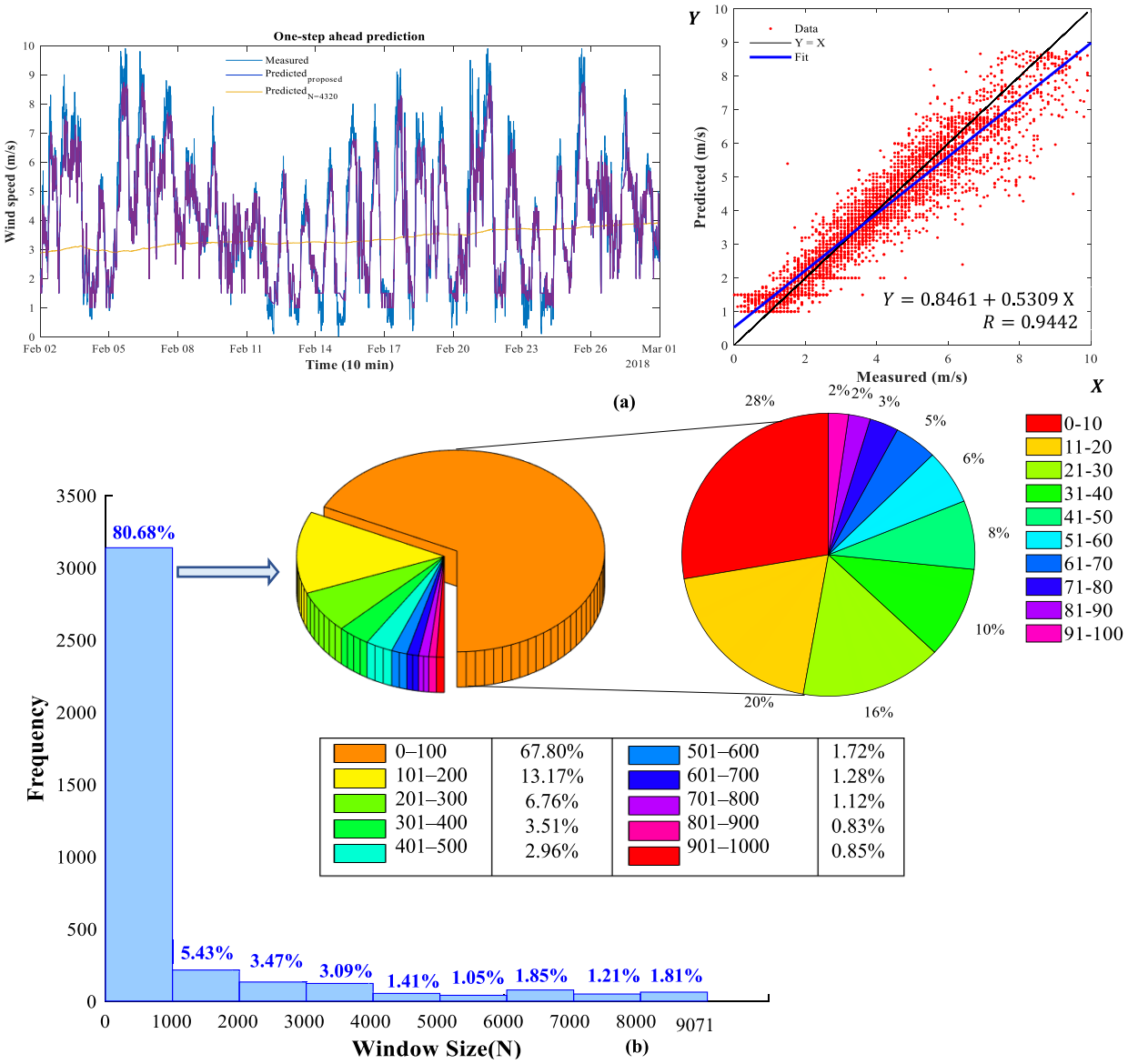


FIGURE 5. (a) Forecasting results of the proposed model (b) Frequency of optimal moving window.

TABLE 5. Comparative analysis of modified markov model with six other statistical and machine learning models.

	N=200		N=1000		N=9072 (length of training dataset)	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Decision Tree	0.652	0.945	0.647	0.980	0.440	0.591
Ensemble Learning	0.643	0.956	0.636	0.953	0.439	0.591
GSR	0.659	1.033	0.615	0.926	0.437	0.587
SVR	0.699	1.209	0.772	1.369	0.439	0.591
GM	1.566	1.950	1.735	2.108	1.855	2.312
LSTM	1.967	2.420	0.821	1.125	0.427	0.571
Modified MC	0.509	0.676	0.495	0.658	0.487	0.650

In general,

1. Identifying a single moving window size is not a suitable choice. A constant moving window size might show promising results for a particular time series fragment but does not capture the overall behavior. Therefore, the adjusting moving window is superior to a single moving window approach.
2. Both equal and unequal intervals of states have disadvantages. Therefore, the proposed state categorization algorithm is generating the interval boundaries based on time series fragments. Also, the state space is modifying at every time step.
3. A training dataset of < 200 data points achieves reasonable estimates, whereas a dataset of 200 – 1000 has excellent forecasting performance. In such a case, modified MC is superior to machine learning models.

C. DISCUSSIONS ON MODWT-ARIMA-MARKOV MODEL

Based on the previous analysis, the maximum window size is set to 1000. Therefore, the minimum level of decomposition is evaluated as $int\,[log\,(1000)] = 3$. Fig. 7 shows

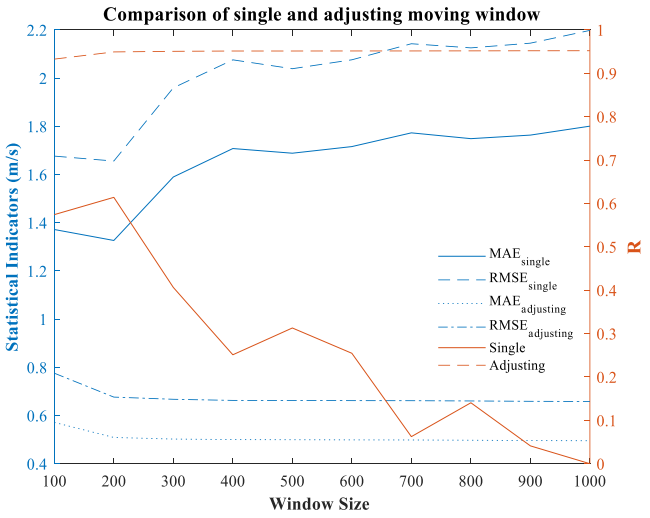


FIGURE 6. Effect of moving window size on statistical indicators.

decomposed wind speed with level 3 for the training dataset and forecasting results of the proposed model for the testing dataset.

TABLE 6. Forecasting performance of the proposed models.

Models	MAE	RMSE	R
Modified MC	0.495	0.658	0.9442
MODWT-Markov	0.439	0.583	0.9625
MODWT-ARIMA-Markov	0.224	0.297	0.9905

Table 6 shows the statistical results of the novel hybrid model. A typical wind speed time series contains slow-moving variations representing the long-term trend. Also, the time series has high-frequency variation in a small-time duration due to turbulence and gust [7]. Therefore, multilevel decomposition of wind speed time series helps in extracting subsequence signals for different frequency bands. In this case, the MODWT-ARIMA-Markov model is advantageous over the modified Markov model. The performance metrics show that the proposed MODWT-ARIMA-Markov model provided the best accuracy utilizing the individual capability of both models. The performances of MODWT-Markov and MODWT-ARIMA-Markov are about 11% and 55% more than the modified MC model, respectively.

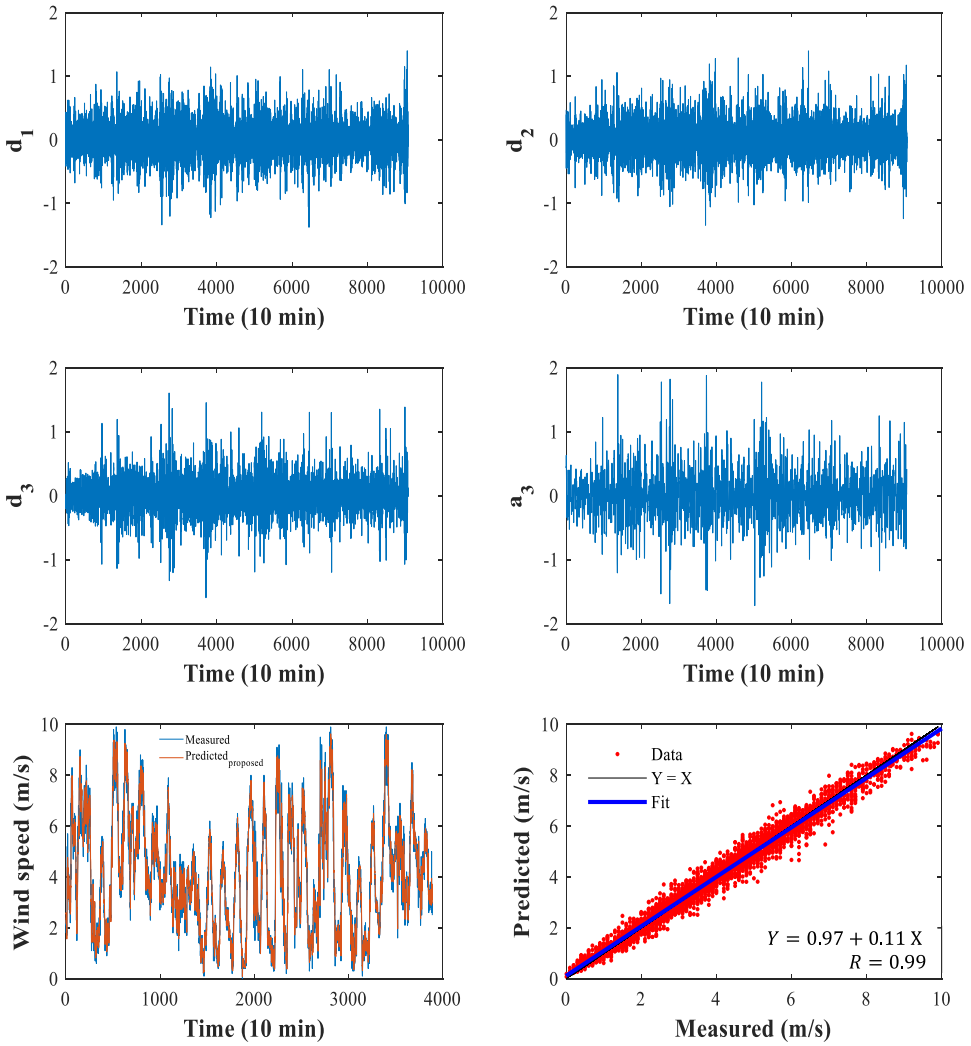


FIGURE 7. Decomposed signals of training dataset and forecasting results of the testing dataset.

TABLE 7. Average computational time per prediction.

Model	Window Size (N)	Computational time (sec)
Modified Markov Chain	200	1.40
	1000	7.82
MODWT-ARIMA-Markov	200	22.10
	1000	115.23

Table 7 shows the average computational time per prediction. Every simulation was run ten times, and the average value of these simulations was considered. The models were programmed on MATLAB using Intel i5, 1.70 GHz processor with quad-core, and 16GB RAM. It should be noted that the computation complexity of MODWT is higher than DWT. DWT requires $O(N)$ multiplications whereas MODWT can be computed in $O(N\ log_2 N)$ multiplications. However, the computational burden of MODWT is equivalent to that of a fast Fourier transform algorithm and is quite acceptable [77]. It is observed from Table 7 that the average computational time of the modified MC model is much lower than the MODWT-ARIMA-Markov model. Similarly, the computational time of the hybrid model is lower than the considered time resolution of 10 min. Therefore, the hybrid model is practically applicable for very short to short-term wind speed forecasts.

The accuracy of the predictions and the adequate computational time show that the forecasting model can be successfully applied to end-user requirements. It includes, but is not limited to, electricity market clearing, operational security in the electric market, real-time grid operation, and load decisions for increments. According to the NIWA wind study report [85], 10 min to hourly wind speed data is an important parameter to model the performance of wind farms and their impact on the national electricity grid.

Further, the model's robustness is validated by the case study of the Sotavento wind farm (43.359° N, 62.346° E) [80]. The hourly data [81] from January 18 2019 to February 17 2019 is considered to compare the performance of the novel model with SVR, KELM, LSTM, ConvLSTM, EMD-KELM, EMD-ConvLSTM, CEEMDAN-KELM, and CEEMDAN-SSAPSR-KELM. The training data includes 400 observations, whereas the testing data has 344 values. The final predicted hourly wind speed time series is shown in Fig. 8, whereas the comparison of statistical results for all considered models is presented in Table 8.

The forecasting results of the MODWT-ARIMA-Markov model showed significant improvements to the relevant contrast models. Considering the training dataset, the order of ARIMA (p, d, q) is selected as (4,1,0) and (5,0,0) for approximate and detailed subseries, respectively. The metrics MAE, RMSE, and MAPE obtained by the proposed model are 0.3163 m/s, 0.4465 m/s, and 7.1829 m/s. The improvement in MAPE is, on average, 55% for single deep learning models and 30% for decomposition-based hybrid models. Although the CEEMDAN-SSAPSR-FS-KELM-MHHOGWO model is

TABLE 8. Comparison among SVR, KELM, LSTM, ConvLSTM, EMD-KELM, EMD-ConvLSTM, CEEMDAN-KELM, CEEMDAN-SSAPSR-KELM, and MODWT-ARIMA-Markov.

Models	MAE	RMSE	MAPE
SVR	0.9909	1.2616	26.1772
KELM	0.9431	1.2084	24.3529
LSTM	0.9598	1.2534	25.8565
ConvLSTM	0.9313	1.1922	23.8726
EMD-KELM	0.5802	0.7748	13.4504
EMD-ConvLSTM	0.4579	0.6424	9.4576
CEEMDAN-KELM	0.6887	0.8087	17.5013
CEEMDAN-SSAPSR-KELM	0.6936	0.7722	17.27
CEEMDAN-SSAPSR-FS-KELM-MHHOGWO	0.4138	0.6212	8.7565
MODWT-ARIMA-Markov	0.4713	0.6285	10.6791

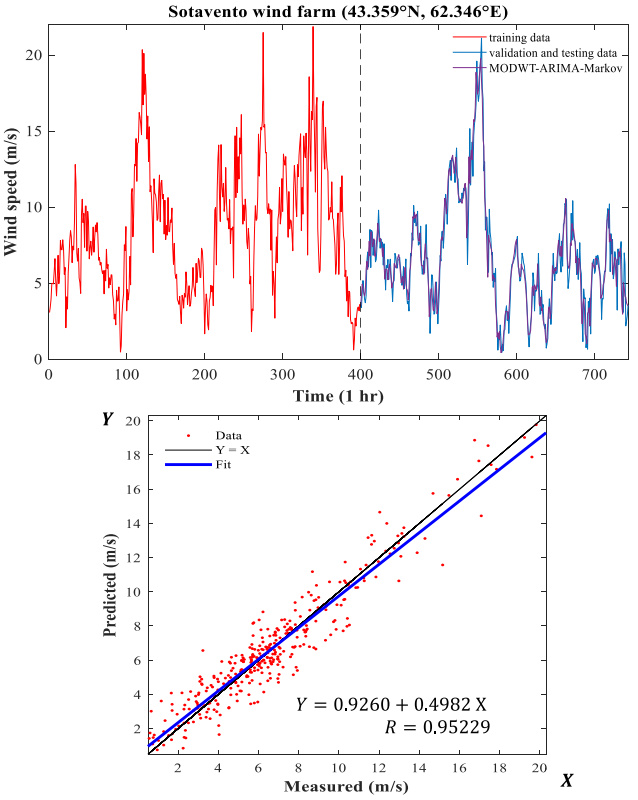


FIGURE 8. Forecasting results of the proposed model for a case study of Sotavento wind farm.

superior to the proposed model in terms of MAPE. However, RMSE values show that both models are applied for long-term planning. The former model is better due to multiple optimization algorithms applied to the model. Therefore, the performance of the MODWT-ARIMA-Markov model will further improve if the feature selection is applied. However, it will increase the complexity of the model. Currently, the novel model is superior to counter artificial intelligence prediction models when lesser historical data is available.

IV. CONCLUSION

This study presents a novel MODWT-ARIMA-Markov forecasting model integrating the maximal overlap discrete wavelet transform (MODWT) with auto-regressive integrated

moving average (ARIMA) and adjusting moving window Markov chains. The novel hybrid model is used to forecast very short to short term-term wind speeds. The following conclusions can be drawn from the present work:

- A single window moving window integrated with MC can provide a good estimation but is not always optimal. A better choice is an adjusting moving window because the wind speed time series do not possess an equal length of behavior for all horizons.
- Compared with a single-window size, the adjusting moving window shows 50% higher forecasting performance. Also, a training dataset of 200 to 1000 observations would be enough to ensure excellent performance.
- The self-adaptive algorithm for state categorization itself decides the interval boundaries (equal and non-equal) using rounding and array functions.
- The novel MODWT-ARIMA-Markov model outperformed considered statistical, AI/ML, deep learning, and hybrid models. The prediction results are in excellent agreement with the measured data for both cases.

ACKNOWLEDGMENT

The authors would like to thank the National Institute of Water and Atmospheric Research Ltd. (NIWA), New Zealand, and Sotavento Galicia, S.A., for providing wind speed data.

REFERENCES

[1] H. Liu and C. Chen, "Data processing strategies in wind energy forecasting models and applications: A comprehensive review," *Appl. Energy*, vol. 249, pp. 392–408, Sep. 2019.

[2] M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Current perspective on the accuracy of deterministic wind speed and power forecasting," *IEEE Access*, vol. 7, pp. 159547–159564, 2019.

[3] Z. Liu, P. Jiang, L. Zhang, and X. Niu, "A combined forecasting model for time series: Application to short-term wind speed forecasting," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114137.

[4] W. Zhang, L. Zhang, J. Wang, and X. Niu, "Hybrid system based on a multi-objective optimization and kernel approximation for multi-scale wind speed forecasting," *Appl. Energy*, vol. 277, Nov. 2020, Art. no. 115561.

[5] D. Song, J. Yang, M. Dong, and Y. H. Joo, "Kalman filter-based wind speed estimation for wind turbine control," *Int. J. Control, Autom. Syst.*, vol. 15, no. 3, pp. 1089–1096, Jun. 2017.

[6] A. Carpinone, M. Giorgio, R. Langella, and A. Testa, "Markov chain modeling for very-short-term wind power forecasting," *Electr. Power Syst. Res.*, vol. 122, pp. 152–158, May 2015.

[7] Aasim, S. N. Singh, and A. Mohapatra, "Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting," *Renew. Energy*, vol. 136, pp. 758–768, Jun. 2019.

[8] L. O. Daniel, C. Sigauke, C. Chibaya, and R. Mbuvha, "Short-term wind speed forecasting using statistical and machine learning methods," *Algorithms*, vol. 13, no. 6, p. 132, May 2020.

[9] M. U. Yousuf, I. Al-Bahadly, and E. Avci, "A modified GM(1,1) model to accurately predict wind speed," *Sustain. Energy Technol. Assessments*, vol. 43, Feb. 2021, Art. no. 100905.

[10] V. N. Sewdien, R. Preece, J. L. R. Torres, E. Rakhshani, and M. van der Meijden, "Assessment of critical parameters for artificial neural networks based short-term wind generation forecasting," *Renew. Energy*, vol. 161, pp. 878–892, Dec. 2020.

[11] L.-L. Li, Z.-Y. Cen, M.-L. Tseng, Q. Shen, and M. H. Ali, "Improving short-term wind power prediction using hybrid improved cuckoo search arithmetic—Support vector regression machine," *J. Cleaner Prod.*, vol. 279, Jan. 2021, Art. no. 123739.

[12] P. A. Adedeji, S. Akinlabi, N. Madushele, and O. O. Olatunji, "Wind turbine power output very short-term forecast: A comparative study of data clustering techniques in a PSO-ANFIS model," *J. Cleaner Prod.*, vol. 254, May 2020, Art. no. 120135.

[13] W. Fu, K. Zhang, K. Wang, B. Wen, P. Fang, and F. Zou, "A hybrid approach for multi-step wind speed forecasting based on two-layer decomposition, improved hybrid DE-HHO optimization and KELM," *Renew. Energy*, vol. 164, pp. 211–229, Feb. 2021.

[14] L. Xiao, W. Shao, F. Jin, and Z. Wu, "A self-adaptive kernel extreme learning machine for short-term wind speed forecasting," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106917.

[15] B. Gu, T. Zhang, H. Meng, and J. Zhang, "Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation," *Renew. Energy*, vol. 164, pp. 687–708, Feb. 2021.

[16] A. Altan, S. Karasu, and E. Zio, "A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer," *Appl. Soft Comput.*, vol. 100, Mar. 2021, Art. no. 106996.

[17] H. Hu, L. Wang, and R. Tao, "Wind speed forecasting based on variational mode decomposition and improved echo state network," *Renew. Energy*, vol. 164, pp. 729–751, Feb. 2021.

[18] X. Yan, Y. Liu, Y. Xu, and M. Jia, "Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition," *Energy Convers. Manage.*, vol. 225, Dec. 2020, Art. no. 113456.

[19] K.-P. Lin, P.-F. Pai, and Y.-J. Ting, "Deep belief networks with genetic algorithms in forecasting wind speed," *IEEE Access*, vol. 7, pp. 99244–99253, 2019.

[20] X. Zhao, N. Jiang, J. Liu, D. Yu, and J. Chang, "Short-term average wind speed and turbulent standard deviation forecasts based on one-dimensional convolutional neural network and the integrate method for probabilistic framework," *Energy Convers. Manage.*, vol. 203, Jan. 2020, Art. no. 112239.

[21] Q. Wu, F. Guan, C. Lv, and Y. Huang, "Ultra-short-term multi-step wind power forecasting based on CNN-LSTM," *IET Renew. Power Gener.*, vol. 15, no. 5, pp. 1019–1029, Apr. 2021.

[22] Y. Chen, Z. He, Z. Shang, C. Li, L. Li, and M. Xu, "A novel combined model based on echo state network for multi-step ahead wind speed forecasting: A case study of NREL," *Energy Convers. Manage.*, vol. 179, pp. 13–29, Jan. 2019.

[23] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renew. Sustain. Energy Rev.*, vol. 34, pp. 243–254, Jun. 2014.

[24] L. Xiao, J. Wang, Y. Dong, and J. Wu, "Combined forecasting models for wind energy forecasting: A case study in China," *Renew. Sustain. Energy Rev.*, vol. 44, pp. 271–288, Apr. 2015.

[25] H. Li, J. Wang, H. Lu, and Z. Guo, "Research and application of a combined model based on variable weight for short term wind speed forecasting," *Renew. Energy*, vol. 116, pp. 669–684, Feb. 2018.

[26] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, and X. Fan, "A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting," *Energy Convers. Manage.*, vol. 136, pp. 439–451, Mar. 2017.

[27] X. Niu and J. Wang, "A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting," *Appl. Energy*, vol. 241, pp. 519–539, May 2019.

[28] M.-R. Chen, G.-Q. Zeng, K.-D. Lu, and J. Weng, "A two-layer nonlinear combination method for short-term wind speed prediction based on ELM, ENN, and LSTM," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6997–7010, Aug. 2019.

[29] Y. Zhang, J. Han, G. Pan, Y. Xu, and F. Wang, "A multi-stage predicting methodology based on data decomposition and error correction for ultra-short-term wind energy prediction," *J. Cleaner Prod.*, vol. 292, Apr. 2021, Art. no. 125981.

[30] L.-L. Li, X. Zhao, M.-L. Tseng, and R. R. Tan, "Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm," *J. Cleaner Prod.*, vol. 242, Jan. 2020, Art. no. 118447.

[31] Y. Zhang, R. Li, and J. Zhang, "Optimization scheme of wind energy prediction based on artificial intelligence," *Environ. Sci. Pollut. Res.*, early access, Mar. 25, 2021, doi: 10.1007/s11356-021-13516-2.

[32] J. Duan, H. Zuo, Y. Bai, J. Duan, M. Chang, and B. Chen, "Short-term wind speed forecasting using recurrent neural networks with error correction," *Energy*, vol. 217, Feb. 2021, Art. no. 119397.

[33] W. Xu, P. Liu, L. Cheng, Y. Zhou, Q. Xia, Y. Gong, and Y. Liu, "Multi-step wind speed prediction by combining a WRF simulation and an error correction strategy," *Renew. Energy*, vol. 163, pp. 772–782, Jan. 2021.
[34] Y. Hao and C. Tian, "A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting," *Appl. Energy*, vol. 238, pp. 368–383, Mar. 2019.
[35] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Appl. Energy*, vol. 235, pp. 939–953, Feb. 2019.
[36] L.-L. Li, Y.-B. Chang, M.-L. Tseng, J.-Q. Liu, and M. K. Lim, "Wind power prediction using a novel model on wavelet decomposition-support vector machines-improved atomic search algorithm," *J. Cleaner Prod.*, vol. 270, Oct. 2020, Art. no. 121817.
[37] J. Ruiz-Aguilar, I. Turias, J. González-Enrique, D. Urda, and D. Elizondo, "A permutation entropy-based EMD-ANN forecasting ensemble approach for wind speed prediction," *Neural Comput. Appl.*, vol. 33, no. 7, pp. 2369–2391, 2020.
[38] N. Bokde, A. Feijóo, N. Al-Ansari, S. Tao, and Z. M. Yaseen, "The hybridization of ensemble empirical mode decomposition with forecasting models: Application of short-term wind speed and power modeling," *Energies*, vol. 13, no. 7, p. 1666, Apr. 2020.
[39] Y. Zhang, Y. Zhao, C. Kong, and B. Chen, "A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic," *Energy Convers. Manage.*, vol. 203, Jan. 2020, Art. no. 112254.
[40] X. Liao, Z. Liu, and W. Deng, "Short-term wind speed multistep combined forecasting model based on two-stage decomposition and LSTM," *Wind Energy*, early access, Feb. 1, 2021, doi: 10.1002/we.2613.
[41] K. U. Jaseena and B. C. Kovoor, "Decomposition-based hybrid wind speed forecasting model using deep bidirectional LSTM networks," *Energy Convers. Manage.*, vol. 234, Apr. 2021, Art. no. 113944.
[42] Z. Sun and M. Zhao, "Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis," *IEEE Access*, vol. 8, pp. 134422–134434, 2020.
[43] F. Tagliaferri, B. P. Hayes, I. M. Viola, and S. Z. Djokić, "Wind modelling with nested Markov chains," *J. Wind Eng. Ind. Aerodyn.*, vol. 157, pp. 118–124, Oct. 2016.
[44] M. Yoder, A. S. Hering, W. C. Navidi, and K. Larson, "Short-term forecasting of categorical changes in wind power with Markov chain models," *Wind Energy*, vol. 17, no. 9, pp. 1425–1439, 2014.
[45] A. Shamshad, M. Bawadi, W. Wanhussin, T. Majid, and S. Sanusi, "First and second order Markov chain models for synthetic generation of wind speed time series," *Energy*, vol. 30, no. 5, pp. 693–708, Apr. 2005.
[46] H. Nfaoui, H. Essiarab, and A. A. M. Sayigh, "A stochastic Markov chain model for simulating wind speed time series at Tangiers, Morocco," *Renew. Energy*, vol. 29, no. 8, pp. 1407–1418, Jul. 2004.
[47] A. Carpinon, R. Langella, A. Testa, and M. Giorgio, "Very short-term probabilistic wind power forecasting based on Markov chain models," in *Proc. IEEE 11th Int. Conf. Probabilistic Methods Appl. Power Syst.*, Jun. 2010, pp. 107–112.
[48] A. D. Sahin and Z. Sen, "First-order Markov chain approach to wind speed modelling," *J. Wind Eng. Ind. Aerodyn.*, vol. 89, nos. 3–4, pp. 263–269, Mar. 2001.
[49] H. Yang, Y. Li, L. Lu, and R. Qi, "First order multivariate Markov chain model for generating annual weather data for hong kong," *Energy Buildings*, vol. 43, no. 9, pp. 2371–2377, Sep. 2011.
[50] J. Tang, A. Brouste, and K. L. Tsui, "Some improvements of wind speed Markov chain modeling," *Renew. Energy*, vol. 81, pp. 52–56, Sep. 2015.
[51] F. Y. Ettoumi, H. Sauvageot, and A.-E.-H. Adane, "Statistical bivariate modelling of wind using first-order Markov chain and weibull distribution," *Renew. Energy*, vol. 28, no. 11, pp. 1787–1802, Sep. 2003.
[52] S. Z. Djokic, B. P. Hayes, R. Langella, and A. Testa, "Modelling of wind energy resources and wind farm power outputs using nested Markov chain approach," in *Proc. Int. Conf. Clean Electr. Power (ICCEP)*, Jun. 2015, pp. 241–246.
[53] K. Xie, Q. Liao, H.-M. Tai, and B. Hu, "Non-homogeneous Markov wind speed time series model considering daily and seasonal variation characteristics," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1281–1290, Jul. 2017.
[54] M. He, L. Yang, J. Zhang, and V. Vittal, "A spatio-temporal analysis approach for short-term forecast of wind farm generation," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1611–1622, Jul. 2014.
[55] S. Murugesan, J. Zhang, and V. Vittal, "Finite state Markov chain model for wind generation forecast: A data-driven spatiotemporal approach," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Jan. 2012, pp. 1–8.
[56] L. Yang, M. He, J. Zhang, and V. Vittal, "Support-vector-machine-enhanced Markov model for short-term wind power forecast," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 791–799, Jul. 2015.
[57] A. Wilinski, "Time series modeling and forecasting based on a Markov chain with changing transition matrices," *Expert Syst. Appl.*, vol. 133, pp. 163–172, Nov. 2019.
[58] A. Gellert, A. Florea, U. Fiore, F. Palmieri, and P. Zanetti, "A study on forecasting electricity production and consumption in smart cities and factories," *Int. J. Inf. Manage.*, vol. 49, pp. 546–556, Dec. 2019.
[59] M. F. Niri, T. Q. Dinh, T. F. Yu, J. Marco, and T. M. N. Bui, "State of power prediction for lithium-ion batteries in electric vehicles via wavelet-Markov load analysis," *IEEE Trans. Intell. Transp. Syst.*, early access, Oct. 20, 2020, doi: 10.1109/TITS.2020.3028024.
[60] T. R. Ayodele, A. S. O. Ogunjuyigbe, R. O. Olarewaju, and J. L. Munda, "Comparative assessment of wind speed predictive capability of first-and second-order Markov chain at different time horizons for wind power application," *Energy Eng.*, vol. 116, no. 3, pp. 54–80, Mar. 2019.
[61] K. Brokish and J. Kirtley, "Pitfalls of modeling wind power using Markov chains," in *Proc. IEEE/PES Power Syst. Conf. Exposit.*, Mar. 2009, pp. 1–6.
[62] G. Papaefthymiou and B. Klockl, "MCMC for wind power simulation," *IEEE Trans. Energy Convers.*, vol. 23, no. 1, pp. 234–240, Mar. 2008.
[63] S. A. P. Kani, G. H. Riahy, and D. Mazhari, "An innovative hybrid algorithm for very short-term wind speed prediction using linear prediction and Markov chain approach," *Int. J. Green Energy*, vol. 8, no. 2, pp. 147–162, Mar. 2011.
[64] T. R. Ayodele, R. Olarewaju, and J. L. Munda, "Comparison of different wind speed prediction models for wind power application," in *Proc. Southern Afr. Universities Power Eng. Conf./Robotics Mechatronics/Pattern Recognit. Assoc. South Afr. (SAUPEC/RobMech/PRASA)*, Jan. 2019, pp. 223–228.
[65] F. Hocaoglu, O. Gerek, and M. Kurban, "The effect of Markov chain state size for synthetic wind speed generation," in *Proc. 10th Int. Conf. Probablistic Methods Appl. Power Syst.*, May 2008, pp. 1–4.
[66] Z. Song, X. Geng, A. Kusiak, and C. Xu, "Mining Markov chain transition matrix from wind speed time series data," *Expert Syst. Appl.*, vol. 38, no. 8, pp. 10229–10239, Aug. 2011.
[67] S. A. P. Kani and M. M. Ardehali, "Very short-term wind speed prediction: A new artificial neural network–Markov chain model," *Energy Convers. Manage.*, vol. 52, no. 1, pp. 738–745, Jan. 2011.
[68] G. D'Amico, F. Petroni, and F. Prattico, "Wind speed and energy forecasting at different time scales: A nonparametric approach," *Phys. A, Stat. Mech. Appl.*, vol. 406, pp. 59–66, Jul. 2014.
[69] Y. Wang, J. Wang, and X. Wei, "A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: A case study of wind farms in northwest China," *Energy*, vol. 91, pp. 556–572, Nov. 2015.
[70] S. Chen, L. Ye, G. Zhang, C. Zeng, S. Dong, and C. Dai, "Short-term wind power prediction based on combined grey-Markov model," in *Proc. Int. Conf. Adv. Power Syst. Autom. Protection*, Oct. 2011, pp. 1705–1711.
[71] L. Lijuan, L. Hongliang, W. Jun, and B. Hai, "A novel model for wind power forecasting based on Markov residual correction," in *Proc. IREC 6th Int. Renew. Energy Congr.*, Mar. 2015, pp. 1–5.
[72] N. Bokde, A. Feijóo, D. Villanueva, and K. Kulat, "A review on hybrid empirical mode decomposition models for wind speed and wind power prediction," *Energies*, vol. 12, no. 2, p. 254, Jan. 2019.
[73] G. D'Amico, F. Petroni, and F. Prattico, "First and second order semi-Markov chains for wind speed modeling," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 5, pp. 1194–1201, Mar. 2013.
[74] P. Brémaud, *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. New York, NY, USA: Springer-Verlag, 2013.
[75] L. Lakatos, L. Szeidl, and M. Telek, *Introduction to Queueing Systems With Telecommunication Applications*. Cham, Switzerland: Springer, 2019.
[76] N. N. Zakaria, M. Othman, R. Sokkalingam, H. Daud, L. Abdullah, and E. A. Kadir, "Markov chain model development for forecasting air pollution index of Miri, Sarawak," *Sustainability*, vol. 11, no. 19, p. 5190, Sep. 2019.
[77] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[78] V. Alarcon-Aquino and J. Barria, "Change detection in time series using the maximal overlap discrete wavelet transform," *Latin Amer. Appl. Res.*, vol. 39, no. 2, pp. 145–152, 2009.

[79] R. Prasad, R. C. Deo, Y. Li, and T. Maraseni, "Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone murray darling basin region using IIS and MODWT algorithm," *Atmos. Res.*, vol. 197, pp. 42–63, Nov. 2017.
[80] W. Fu, K. Wang, J. Tan, and K. Zhang, "A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 205, Feb. 2020, Art. no. 112461.
[81] *Historical-Real Time Data*. Sotavento. Accessed: May 25, 2019. [Online]. Available: http://www.sotaventogalicia.com/en/technical-area/real-time-data/historical/
[82] *The National Climate Database*. National Institute of Water and Atmospheric Research (NIWA). Accessed: May 25, 2019. [Online]. Available: https://cliflo.niwa.co.nz/
[83] P. R. Chappell, *The Climate and Weather of Manawatu-Wanganui*. Auckland, New Zealand: NIWA, Taihoro Nukurangi, 2015.
[84] C. McComb, J. Cagan, and K. Kotovsky, "Capturing human sequence-learning abilities in configuration design tasks through Markov chains," *J. Mech. Des.*, vol. 139, no. 9, pp. 1–38, Sep. 2017.
[85] (2009). *Multi-Year Ten-Minute Synthetic Wind Speed Time-Series for 15 Actual or Proposed New Zealand Wind Farms*. [Online]. Available: https://niwa.co.nz/sites/niwa.co.nz/files/import/attachments/NIWA-Synthetic-Wind-Study-report.pdf

MUHAMMAD UZAIR YOUSUF received the B.E. and M.E. degrees in mechanical engineering from the NED University of Engineering and Technology, Karachi, Pakistan, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree in mechanical and electrical engineering with Massey University. His current research interests include wind energy forecasting and solar energy modeling.

IBRAHIM AL-BAHADLY (Senior Member, IEEE) received the B.Sc. (Eng.) degree from the Baghdad University of Technology, in 1987, and the M.Sc. and Ph.D. degrees from Nottingham University, in 1990 and 1994, respectively, all in electrical and electronic engineering.

From 1994 to 1996, he was a Research Associate with the Electric Drives and Machines Group, Newcastle University, Newcastle upon Tyne, U.K. Since 1996, he has been with Massey University, where he is currently an Associate Professor in electrical and electronic engineering. His research interests include power electronic applications, variable speed machines and drives, renewable energy systems, instrumentation, robotics, and automation.

EBUBEKIR AVCI (Member, IEEE) received the Ph.D. degree in robotics from Osaka University, Japan, in 2013. He was a Postdoctoral Researcher in parallel mechanism robots with the Toyota Technological Institute, Japan, from 2013 to 2014. He was a Research Associate with the Hamlyn Centre for Robotic Surgery, Imperial College London, U.K., from 2014 to 2015. He is currently a Senior Lecturer in mechatronics with Massey University.