ENHANCING STATISTICAL WIND SPEED FORECASTING MODELS

A THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

IN

ENGINEERING

AT MASSEY UNIVERSITY, MANAWATŪ CAMPUS, NEW ZEALAND.

MUHAMMAD UZAIR YOUSUF

2022

ABSTRACT

In recent years, wind speed forecasting models have seen significant development and growth. In particular, hybrid models have been emerging since the last decade. Hybrid models combine two or more techniques from several categories, with each model utilizing its distinct strengths. Mainly, data-driven models that include statistical and Artificial Intelligence/Machine Learning (AI/ML) models are deployed in hybrid models for shorter forecasting time horizons (< 6hrs). Literature studies show that machine learning models have gained enormous potential owing to their accuracy and robustness. On the other hand, only a handful of studies are available on the performance enhancement of statistical models, despite the fact that hybrid models are incomplete without statistical models. To address the knowledge gap, this thesis identified the shortcomings of traditional statistical models while enhancing prediction accuracy. Three statistical models are considered for analyses: Grey Model [GM(1,1)], Markov Chain, and Holt's Double Exponential Smoothing models. Initially, the problems that limit the forecasting models' applicability are highlighted. Such issues include negative wind speed predictions, failure of predetermined accuracy levels, non-optimal estimates, and additional computational cost with limited performance. To address these concerns, improved forecasting models are proposed considering wind speed data of Palmerston North, New Zealand. Several methodologies have been developed to improve the model performance and fulfill the necessary and sufficient conditions. These approaches include adjusting dynamic moving window, self-adaptive state categorization algorithm, a similar approach to the leave-one-out method, and mixed initialization method. Keeping in view the application of the hybrid methods, novel MODWT-ARIMA-Markov and AGO-HDES models are further proposed as secondary objectives. Also, a comprehensive analysis is presented by comparing sixteen models from three categories, each for four case studies, three rolling windows, and three forecasting horizons. Overall, the improved models showed higher accuracy than their counter traditional models. Finally, the future directions are highlighted that need subsequent research to improve forecasting performance further.

AUTHORS DECLARATION

The thesis is compiled according to Massey University's Guidelines for 'Doctoral Thesis with Publications'. This thesis is based on research that has either been published or is currently under review. Two articles are published in IEEE, two in Elsevier and one in Wiley journals. In accordance with IEEE's copyright policy, author and third parties, including funder websites, may post, share, and use the final published article without permission that are published under a Creative Commons Attribution License (CC BY), even for commercial purposes or to create derivative works. Similarly, as per Elsevier policy, theses and dissertations which contain embedded published journal articles as part of the formal submission can be hosted by the awarding institution with DOI links back to the formal publications on ScienceDirect. Further, as per Wiley's policy, an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. While the content is identical to the published versions, there may be stylistic differences between the enclosed work, and the published versions. Furthermore, some of the submitted chapters are relatively succinct, there is some repetition (particularly in the literature review of some chapters), and there are stylistic differences between the chapters themselves.

ACKNOWLEDGEMENTS

All praise belongs to Allah, and may He bestow His peace and blessings upon all His messengers, and His Final messenger Prophet Muhammad Peace be Upon Them all, until the day of Judgement.

I would like to express my deepest gratitude to my primary supervisor, Associate Professor Ibrahim Al-Bahadly, for his continuous guidance and kind support. I have learned a lot from you. This work would never have been completed without your persistent encouragement and great dedication.

I am equally thankful to my co-supervisor, Dr. Ebubekir Avci, for being an excellent mentor. Thank you a million for such a wonderful contribution.

I have been deeply indebted to my elder brother, Dr. Muhammad Umair, since my childhood. Thank you for motivating me toward higher studies, checking on my progress, praying for me, and continuing to support me. Thank you for everything you do. I am honoured to have you as my elder brother.

My gratitude also extends to the welcoming administration and IT support team at the School of Food and Advanced Technology, especially Glenda, Mark, Nick, and Tim. A big thank you also to Rehan, Haroon, Hamed, Asmat, Saad, and Waqar for offering advice and support. Many thanks to dear friends who have become my second family in New Zealand. Rafay Temuri, Riaz Rehman, Muneeb, Shahnawaz Ahmed, Syed Ahmer, Muhammad Shah, Salim, Saiful Bahri, and Ainuddin are a few of them from a long list. Your prayers and support meant a lot to me throughout this journey.

I gratefully acknowledge Higher Education Commission (HEC) Pakistan for HRDI- UESTP Scholarship. Further, the Tilt Renewables Tararua Wind Farm Research Bursary award, for two consecutive years, is not just an honour. It tells me that I am seen as having potential for the future of our society.

Most of all, I would like to say a heartfelt thank you to my beloved mother, Zainab, my dearest father, Muhammad Yousuf, and my lovely wife, Dr. Fatema. I cannot express my emotions and love for all of you in words. I am deeply grateful for all the sacrifices you made for me. Also, I am thankful to my eldest brother, Ovais, and my charming sister, Humaira. Thank you for your care and prays for my success. I am overjoyed to mention that I have been blessed with my first son, "Muhammad Ibrahim," at the end of this journey. Love you all.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION **1**
1.1. Research Motivation 1
1.2. Problem statement 2
1.3. Research Questions 2
1.4. Research Objectives 3
1.5. Scope and Limitations 3
1.6. Research Contributions 4
1.7. Related Publications 5
1.8. Thesis Organization and Outline 6
CHAPTER 2. CURRENT PERSPECTIVE ON THE ACCURACY OF DETERMINISTIC WIND SPEED AND POWER FORECASTING **9**
2.1.Introduction 10
2.1.1. Objectives and Motivations 11
2.2.Deterministic Wind Speed and Power Forecasting Classification 11
2.2.1. Input Data 12
2.2.2. Time-Scales 13
2.2.3. Output 13
2.2.4. Forecasting Model 13
2.2.4.1. Persistence Method 13
2.2.4.2. Physical Method 13
2.2.4.3. Statistical Method 14
2.2.4.4. Artificial Intelligence/ Machine Learning Methods 15
2.2.4.5. Hybrid Method 17
2.3.Methodology 18
2.3.1. Performance Indicators 18
2.3.2. Performance Analysis of Dataset 18
2.4.Conclusion 24
CHAPTER 3. A MODIFIED GM(1,1) MODEL TO ACCURATELY PREDICT WIND SPEED **28**
3.1.Introduction 29
3.1.1. Grey Prediction Models 30
3.1.2. Applications of grey prediction models in the field of energy forecasting 31
3.1.3. Limitations of previous studies 32

3.2.Materials and Methods..33
3.2.1. GM(1,1) model..34
3.2.2. Remnant GM(1,1) model 34
3.2.3. Singular phenomenon..34
3.2.4. Accuracy testing of GM(1,1) model 35
3.3.Results and Discussions...37
3.3.1. Site Specifications..37
3.3.2. One step ahead prediction based on traditional GM(1,1)........................37
3.3.3. One step ahead prediction based on the proposed methodology 37
3.3.4. Performance Indicators 39
3.3.5. Computational Time..40
3.3.6. Evaluation of robustness 41
3.4.Conclusions...42
CHAPTER 4. SHORT-TERM WIND SPEED FORECASTING BASED ON HYBRID MODWT-ARIMA-MARKOV MODEL...44
4.1.Introduction...46
4.1.1. Motivation and Contributions 49
4.2.Methodology..49
4.3.Experiments and Analysis...53
4.3.1. Data Set Description..54
4.3.2. Comparison of Single Moving Window with Adjusting Moving Window...........54
4.3.3. Discussions on MODWT-ARIMA-MARKOV Model..56
4.4.Conclusions...58
CHAPTER 5. WIND SPEED PREDICTION FOR SMALL SAMPLE DATASET USING HYBRID FIRST-ORDER ACCUMULATED GENERATING OPERATION-BASED DOUBLE EXPONENTIAL SMOOTHING MODEL...62
5.1.Introduction...63
5.1.1. Previous studies on exponential smoothing model 65
5.1.2. Limitations of previous studies 65
5.2. Materials and Methodology 68
5.2.1. Double Exponential Smoothing Model 68
5.3.Improved Double Exponential Smoothing Model...68
5.3.1. Improvement # 01: Optimal Smoothing constant 68
5.3.2. Improvement # 02: Initial Value calculation...69

5.3.3. Improvement # 03: First-order Accumulated Generating Operation 69
5.4.Results and Discussions 70
5.4.1. Site Description 70
5.4.2. Mixed Initialization Method 70
5.4.3. One-step ahead wind speed forecast 72
5.5.Conclusion 73
CHAPTER 6. STATISTICAL WIND SPEED FORECASTING MODELS FOR SMALL SAMPLE DATASETS: PROBLEMS, IMPROVEMENTS, AND PROSPECTS 77
6.1. Introduction 78
6.1.1. Objectives and Motivation 80
6.2. Materials and Methods 78
6.2.1. Case Studies 80
6.2.2.Traditional Statistical Models 81
6.2.2.1. Auto-Regressive Integrated Moving Average 81
6.2.2.2. Double Exponential Smoothing Model 81
6.2.2.3. Markov Chain Model 82
6.2.2.4. GM (1,1) 83
6.2.3. Problems associated with traditional models 85
6.2.4. Improved Models 86
6.2.4.1. Modified Double Exponential Smoothing Model (modDES) 86
6.2.4.2. Modified Markov Chain Model (modMC) 86
6.2.4.3. Modified Grey Model (modGM) 86
6.2.4.4. Weighted Fractional Grey Model (WFGM) 87
6.2.5. Machine Learning Models 87
6.2.5.1. Gaussian Process Regression 87
6.2.5.2. Decision Tree 87
6.2.5.3. Ensemble Learning 88
6.2.5.4. Support Vector Machine 88
6.3. Results and discussions 88
6.3.1. Qualitative Results 88
6.3.2. Quantitative results 89
6.3.3. Comparison of the best models 91
6.3.4. Future Directions 91
6.4.1. Unaddressed Issues of GM (1,1) 91

6.4.2. Prediction lag 92
6.4.3. Imputation Techniques 92
6.4. Conclusions 92
6.5. Data Availaibility 93
CHAPTER 7. CONCLUSIONS AND FUTURE DIRECTIONS 96
7.1. Conclusions 96
7.2. Future directions 97
APPENDICES 98

LIST OF FIGURES

CHAPTER 2

Figure 1. Cumulative installed wind power for 2001-2018 and forecast for 2019-2023 according to GWEC 10

Figure 2. Detailed classification of deterministic wind speed and power forecasting 12

Figure 3. Flowchart for the physical method 14

Figure 4. Flowchart for the statistical method 14

Figure 5. Overall Summary of the selected 28 papers based on forecasting classification. The percentages (%) indicate the frequency of data implemented for each subcategory out of 634 data entries 20

Figure 6. (a) Illustrative plot presenting the performance of the model reported in the selected papers, (b) 2-dimensional histogram presenting the qualitative picture of datasets 21

Figure 7. The performance of a) prediction models, b) time scale, in terms of averaged values. 22

Figure 8. Performance indicators (a) nMAE (b) nRMSE, for forecasting model, (c) nMAE (d) nRMSE, for time scale. 22

Figure 9. Statistical indicators with respect to the time horizon 23

Figure 10. Linear trend of statistical errors with respect to publication year 23

CHAPTER 3

Figure 1. The rolling mechanism for a window size of 5 30

Figure 2. One step ahead wind speed forecasting using traditional GM(1,1) model for window sizes varying from 4 to 11 32

Figure 3. One step ahead wind speed forecasting using traditional GM(1,1) model for varying weighting factor 33

Figure 4. The proposed modified GM(1,1) model 35

Figure 5. Accuracy testing for: (a) $\mathrm{N}=$ constant, $\alpha=$ variable; (b) $\mathrm{N}=$ variable, $\alpha=$ constant. The numbers within the graph show the total data points that achieve the accuracy level I. 37

Figure 6. Simulation results showing the combination of $(\mathrm{N},\alpha)$to achieve the highest prediction accuracy level 37

Figure 7. One step ahead prediction for the proposed methodology 38

Figure 8. Histograms of error for: (a) $\alpha = 0.9$, $\mathrm{N} =$ variable; (b) $\alpha =$ variable, $\mathrm{N} =$ variable 39
Figure 9. One step ahead wind speed forecasting using traditional GM(1,1) model for (a) $\mathrm{N} = 5, \alpha = 0.1$ (b) $\mathrm{N} = 11, \alpha = 0.9$ (c) $\mathrm{N} = 50, \alpha = 0.5$ 40
Figure 3.10. One step ahead wind speed forecasting using the proposed model and analysis of residuals for Mpika, Galicia, and Gwadar 41

CHAPTER 4

Figure 1. Selection of optimal window based on a similar approach to the leave-one-out method 50
Figure 2. Properties of a transition probability matrix with a transition graph of sequence discussed in the state categorization step 51
Figure 3. The flowchart of the proposed hybrid MODWT-ARIMA-Markov model 53
Figure 4. The detailed information of the experimental dataset 55
Figure 5. (a) Forecasting results of the proposed model (b) Frequency of optimal moving window 56
Figure 6. Effect of moving window size on statistical indicators 57
Figure 7. Decomposed signals of training dataset and forecasting results of the testing dataset 57
Figure 8. Forecasting results of the proposed model for a case study of Sotavento wind farm 58

CHAPTER 5

Figure 1. Global distribution of wind capacity (in MW) for the year 2020 64
Figure 2. Forecasting issues with the traditional Holt's Double exponential smoothing model for a rolling widow of 6 67
Figure 3. Proposed framework for AGO- HDES forecasting model 71
Figure 4. Effect of the initialization method on the performance of the proposed forecasting models for varying rolling window 73
Figure 5. One step ahead wind speed forecast for Markov Chain, GM (1,1), Traditional DES and AGO-HDES models 74
Figure 6. Scatter plot for measured and predicted wind speed 75

CHAPTER 6

Figure 1. Sites from New Zealand, Pakistan, Zambia, and Spain are selected as case studies 81
Figure 2. Taxonomy of the selected 16 models from three categories 83
Figure 3. Examples of erroneous predictions: exceeding and negative 84
Figure 4. The traditional value of $\alpha = 0.5$ is not always the best selection, and hence an optimal parameter is required 84
Figure 5. Effect of initialization method on DES model 85
Figure 6. Qualitative Performance of the forecasting models in terms of the average quality of forecasts per a) case study b) model type c) forecasting horizon d) Rolling window 87
Figure 7. Variations of errors in terms of boxplot 88
Figure 8. Taylor Diagram for the comparison of all 144 models (16 models × 3 rolling windows × 3 time-steps) 89
Figure 9. Failure of GM(1,1) at a rolling window of 50 92
Figure 10. Two main inconsistencies: Amplitude and Phase errors 93

APPENDICES

Figure A.1. Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows: Case study of Palmerston North. 99
Figure A.2. Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows: Case study of Gwadar 100
Figure A.3. Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows: Case study of Mpika 101
Figure A.4. Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows: Case study of Galicia 102
Figure A.5. DRC 16 - Chapter 2 103
Figure A.6. DRC 16 - Chapter 3 104
Figure A.7. DRC 16 - Chapter 4 105
Figure A.8. DRC 16 - Chapter 5 106
Figure A.9. DRC 16 - Chapter 6 107

LIST OF TABLES

CHAPTER 2

Table 1. Available global and regional NWP models......12
Table 2. Description of commercially available physical, statistical and combined models 15
Table 3. Description of the papers included in the study......19
Table 4. Performance of forecasting models according to time horizon and impact of input data on forecasting accuracy......23

CHAPTER 3

Table 1. Studies implementing GM(1,1) model for wind energy forecasting......31
Table 2. Developing coefficient and prediction length......34
Table 3. Levels of accuracy testing for GM(1,1)......35
Table 4. Summary of meteorological parameters at Palmerston North......36
Table 5. Forecasting Validity, individual accuracy level I, and forecasting performance of five cases of the proposed model with improvement ratios......38
Table 6. One step ahead wind speed prediction error for Palmerston North......39
Table 7. Average computational time per prediction for four models......40
Table 8. Site specifications of Mpika, Galicia and, Gwadar......40
Table 9. One step ahead wind speed prediction error for Mpika, Galicia, and Gwadar....42

CHAPTER 4

Table 1 .Studies Implementing Markov Chain Model for Wind Energy Forecasting......48
Table 2. Theoretical ACF and PACF for stationary time series......52
Table 3. Summary of Meteorological Parameters at Palmerston North, New Zealand.....54
Table 4. Comparative Analysis of Single and Adjusting Moving Window......55
Table 5. Comparative Analysis of Modified Markov Model with Six Other Statistical and Machine Learning Models......56
Table 6. Forecasting Performance of the Proposed Models......57
Table 7. Average Computational Time Per Prediction......58
Table 8. Comparison among SVR, KELM, LSTM, ConvLSTM, EMD-KELM, EMD-ConvLSTM, CEEMDAN-KELM, CEEMDAN-SSAPSR-KELM, and MODWT-ARIMA-Markov......58

CHAPTER 5

Table 1. Studies implementing the DES model for wind speed forecasting 66
Table 2. One step ahead wind speed forecast for the best possible combination 69
Table 3. MSE of the mixed initialization method for HDES...72
Table 4. MSE of the mixed initialization method for damped HDES 72
Table 5. MSE of the mixed initialization method for AGO- HDES..................................72
Table 6. Performance evaluation of the forecasting models..74
Table 7. Average computational time per prediction...75

CHAPTER 6

Table 1. Examples of studies that integrated statistical and machine learning algorithms 80
Table 2. Descriptive data of the selected case studies 82
Table 3. Effect of developing coefficient on prediction length 83
Table 4. Accuracy levels for posterior error detection measures.......................................84
Table 5. Shortcomings of considered statistical models and suggested improvements 86
Table 6. Choices of Mixed Initialization Combinations..86
Table 7. Ranking of models according to Global Performance Indicator (GPI) 90
Table 8. Comparison of modDES and SVM models for selected case studies 92

CHAPTER 1
INTRODUCTION

1.1. Research Motivation

Wind energy has been seen as a viable alternative to traditional fossil fuels in recent decades. According to the Statistical Review of World Energy 2021 [1], the highest contribution to renewable energy growth came from wind in 2020. Particularly, China, the USA, Germany, and India account for 38.46%, 16.06%, 8.48%, and 5.26%, respectively. Additionally, an impressive increase in new installations is observed for Colombia and Russia, with almost 28 and 9 times more capacity than the previous year. Other notable countries include Sri Lanka (1.96 times), Kazakhstan (1.71 times), Argentina (1.63 times), and Vietnam (1.6 times) for a remarkable increase in wind power from the preceding year [2]. Also, wind energy is considered an effective option in many countries' current supply and future energy policies. According to the New Zealand Wind Energy Association (NZWEA), at present, wind capacity is supplying nearly 6% of New Zealand's total electricity generation. NZWEA's vision is to increase this capacity by around 20% by 2035 [3]. Similarly, in Pakistan, wind power projects of 1,335 MW cumulative capacity are supplying electricity to National Grid, while 38% more capacity is under construction [4]. Furthermore, wind power generation has recently emerged in many countries. It includes the first wind farm of Saudi Arabia with a capacity of 400MW that started generating electricity in August 2021 [5]. In addition, the wind capacity is expected to expand further, as onshore wind power costs have dropped by around 40% in the last five years [1].

Although an economically competitive technology, wind energy has the drawback of intermittency, the stochastic behavior of wind raised various challenges to the power systems. Such problems include, but are not limited to, wind turbine regulation, real-time grid operation, and economic load dispatch planning. For example, if there is a power variation between the programmed and actual power output, the system operator must assign additional energy reserves. This rise in reserves would raise operational expenses, raising final energy prices [6]. Corresponding to this, wind speed forecasting on different timescales is a powerful tool that brings wind energy more compatible with power generation planning while reducing the need for extra reserve capacity. According to the literature studies, accurate wind speed and power forecasting are significant factors for a wide range of wind applications, from siting to integration [7]. Some of the standard applications discussed in the literature are siting and designing of a wind farm, grid integration and operations (dispatch planning, unit commitment decisions, farm regulations, maintenance scheduling, energy storage, reserve planning), stability (power stability, reducing breakdown probability) and revenue generation (tariff in the electricity market, electricity bidding, and trading).

1.2. Problem statement

Hybrid wind speed forecasting models have been emerging since the last decade. Such models combine two or more techniques from several categories, with each model utilizing its distinct strengths. Mainly, data-driven models are deployed in hybrid models for shorter forecasting time horizons (< 6hrs). There are two main categories of data-driven forecasting models: Statistical models and Artificial Intelligence/Machine Learning (AI/ML) models. In recent years, AI/ML techniques have gained much popularity in wind speed forecasting, occupying irreplaceable dominance. [8, 9]. Instead of the conventional programming of step-by-step coding instructions based on logic and if-then rules, machine learning models are based on algorithms capable of learning by trial and error and enhancing their performance over time [10]. As a result, the majority of current research is primarily focused on improving the performance of AI/ML models. However, hybrid models are incomplete without considering statistical models. Statistical models are not only easy to build and fast to calculate, but they are also robust and less prone to overfitting than more complex approaches. In that respect, statistical models are valuable not just on their own but also as part of hybrid models that incorporates more advanced methodologies [11]. Hence, accuracy improvements of statistical models are as significant as machine learning models. Therefore, the primary aim of this thesis is to address the problems of traditional statistical models while enhancing the prediction performance.

1.3. Research Questions

This thesis would help in answering the following research questions:

- What are the shortcomings of commonly used traditional statistical models for forecasting wind speed and how can they be addressed?
- How much performance is projected to improve using modified wind speed forecasting models?

1.4. Research Objectives

To accomplish the primary aim and to answer the research questions, we target to achieve the following objectives from this research.

- To identify the prediction problems of three commonly applied statistical models: Grey Model [GM(1,1)], Markov Chain, and Holt's Double Exponential Smoothing models, in terms of forecasting results acceptability
- To explore improved models while addressing the limitations of existing approaches.
- To compare the forecasting performance of traditional, improved, and machine learning models considering four case studies, three rolling windows, and three forecasting horizons.

1.5. Scope and Limitations

While considering the forecasting, one may note that universally best technique does not exist [12]. A model may improve the forecasting performance significantly for one region whereas marginally for another area. Similarly, a model that gives better results in one time horizon may worsen the outcomes in another time horizon. Therefore, to compare the effectivity of models and to present more general results, four case studies are considered for three forecasting horizons covering very short-term to short-term forecasts. The medium- and long-term forecast is not the scope of this study. Therefore, the Numerical Weather Prediction (NWP) input is not considered.

Data is essential to forecasting. For selected case studies, the raw data is retrieved from Energy Sector Management Assistance Program (https://energydata.info/), National Institute of Water and Atmospheric Research (https://cliflo.niwa.co.nz/), and Sotavento wind farm (https://www.sotaventogalicia.com/en/technical-area/real-time-data/historical/) open-access databases. However, the considered datasets do not limit the applicability of the proposed models. Therefore, the methodologies presented and explored in this thesis are general.

All the models are programmed on MATLAB using Intel i5, 1.70 GHz processor with quad-core, and 16GB RAM. The average computational time per prediction for forecasting models is provided in respective chapters. To address the scientific community's capacity to replicate and build on the research findings stated in this thesis, the author would be pleased to share the MATLAB code on individual request following the Massey University Policy Guide.

1.6. Research Contributions

The research contributions are summarized as follows.

Quantitative Review Analysis instead of qualitative review: Many review studies are conducted for wind speed and wind power forecasting. For instance: Foley et al. [13] presented a review of the physical and statistical models. Tascikaraoglu & Uzunoglu [14] and Xiao et al. [15] reviewed the hybrid wind forecasting models. Qian et al. [16] discussed the three structures of the decomposition model, whereas Bokde et al. [7] studied the Empirical Mode Decomposition (EMD) based hybrid models. Recently, Hui Liu, with other collaborators, reviewed data processing strategies [17] and intelligent predictors & auxiliary methods[8]. All these reviews are of significant importance in the field. However, all the above-mentioned review studies compared the cross-literatures qualitatively and not quantitatively. This is because, without a standard dataset, the comparisons of different models could only be qualitative. We compared the accuracy results of already published articles in dimensionless parameters to present an in-depth analysis. Such quantitative study has not been performed for wind energy forecasting in recent years. Further details of this work are covered in Chapter 2 of this thesis.

Improved Statistical models: As per the defined objectives, three statistical models: Grey Model [GM(1,1)], Markov Chain, and Holt's Double Exponential Smoothing models are extensively studied. The grey prediction has the advantage of requiring less information to comprehensively address the uncertainty of raw data. However, the traditional model may predict very high or even negative values of wind speeds. Considering these issues, a modified GM(1,1) model is proposed and tested for four case studies. Markov Chains are based on the probability distribution, which shows that the wind speed at the following time step relies on the present wind state. 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. Considering this issue, a modified MC model is proposed that integrates MCs with an adjusting dynamic moving window. Forecasts generated from the Double exponential smoothing (DES) model are a weighted average of past observations, with previous values assigned exponentially decreasing weights. Such models have been popular since their development; however, the forecasting performance decreases considerably for small datasets. Therefore, a modified Holt's Double Exponential Smoothing model is proposed for small samples. Further details of the modified models are provided in Chapters 3,4 and 5 of this thesis.

Novel Hybrid Models: In view of the state-of-the-art forecasting methods discussed in the review study, the hybrid models perform best for every time horizon. Other than enhancing the performance of statistical models, we also proposed two novel hybrid models: MODWT-ARIMA-Markov and AGO-DES, as secondary objectives. In the earlier model, maximal overlap discrete wavelet transform (MODWT) is combined with auto-regressive integrated moving average (ARIMA) and adjusting moving window Markov Chain models. In the latter case, the first-order accumulated generating operator is integrated with an improved double exponential smoothing model to enhance the performance. The performances of the proposed hybrid models are further compared with models available in the literature. The explanations are covered in Chapters 4 and 5 of this thesis.

Comprehensive Comparative Analysis. We also comprehensively discussed the problems, improvements, and prospects of statistical wind speed forecasting models for small sample datasets. We explored four case studies, sixteen models, three rolling windows, and three forecasting horizons, such that the final quality-controlled database contains 576 entries. The qualitative and quantitative results that would bring interesting insights are discussed in Chapter 6 of this thesis.

1.7. Related Publications

Below is the list of journal articles published as a result of this research.

- 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. DOI: https://doi.org/10.1109/ACCESS.2019.2951153
- M. U. Yousuf, I. Al-Bahadly, and E. Avci, "A modified GM (1, 1) model to accurately predict wind speed," *Sustainable Energy Technologies and Assessments*, vol. 43, p. 100905, 2021.DOI: https://doi.org/10.1016/j.seta.2020.100905
- M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Short-term wind speed forecasting based on hybrid MODWT-ARIMA-Markov model," *IEEE Access*, vol. 9, pp. 79695-79711, 2021. DOI: https://doi.org/10.1109/ACCESS.2021.3084536
- M. U. Yousuf, I. Al- Bahadly, and E. Avci, "Wind speed prediction for small sample dataset using hybrid first- order accumulated generating operation- based double exponential smoothing model," *Energy Science and Engineering*, vol. 10(3), pp. 726-739, 2022. DOI: https://doi.org/10.1002/ese3.1047

- M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Statistical wind speed forecasting models for small sample datasets: Problems, Improvements, and Prospects," *Energy Conversion and Management*, vol. 261, pp. 115658, 2022. DOI: https://doi.org/10.1016/j.enconman.2022.115658

1.8. Thesis Organization and Outline

The thesis is compiled according to Massey University's Guidelines for 'Doctoral Thesis with Publications'. The thesis is organized as follows.

Chapter 2. This chapter presents a comprehensive review and current perspective on the accuracy of deterministic wind speed and power forecasting models. The chapter first categorizes the wind forecasting methods into four broader classifications: input data, time scales, power output, and forecasting method. Following this, the performance of wind speed and power forecasting models is evaluated based on 634 accuracy tests reported in twenty-eight published articles covering fifty locations of ten countries. At last, the chapter is concluded by stating the highlights drawn from performance analysis.

Chapter 3. This chapter highlights the problems associated with the first-order grey model with one variable [GM(1,1)]. This is the first statistical model considered in this study for performance improvement. The limitations of previous studies are discussed in detail, including the reasons for the negative wind speed forecasts. Next, a comprehensive modified GM(1,1) model is proposed that integrates the traditional method with an optimal moving window and the adaptive weighting factor for very short-term wind speed forecasting. The effect of window size and weighting parameters is discussed to meet the accuracy levels. The necessary conditions of developing coefficient are considered in the proposed method that successfully solved previous models' major problems. The traditional model is modified by L' Hopital's rule and remnant model to fulfill the necessary and sufficient conditions. Lastly, the robustness of the model is discussed with three case studies.

Chapter 4. This chapter explains the enhancement of the Markov Chain (MC) model, the second considered model in this study. First, state categorization and window size problems are discussed in detail. Next, a forecasting model is proposed that integrates MC with an adjusting dynamic moving window. A similar approach to the leave-one-out method is suggested to select the optimal size of the rolling window. Further, a comprehensive analysis is provided to compare single and adjusting moving window approaches. Moreover, the state-categorization of traditional MC is improved by introducing a self-adaptive algorithm to optimize the number of transition states. Instead of synthetically generating time series, the modified model directly predicts one step ahead wind speed. Based on preliminary findings, a novel hybrid model is also proposed as a secondary objective combining maximal overlap discrete wavelet transform (MODWT) with auto-regressive integrated moving average (ARIMA) and adjusting moving window MC. Lastly, the performance of the proposed hybrid model is compared with other models available in the literature.

Chapter 5. Holt's Double Exponential Smoothing Model is analyzed as the third statistical model in this chapter. Firstly, the problems of traditional Holt's double exponential smoothing model are highlighted for small sample datasets. It also includes the reasons for negative predictions. Next, improvements are suggested to improve the model performance, including the mixed initialization method. Overall, eight choices of initial values from four categories are considered for the comparison. Finally, a novel hybrid first-order accumulated generating operator-based double exponential smoothing model is proposed and compared as a secondary objective.

Chapter 6. This comprehensive chapter addresses all three objectives and includes novel models proposed in chapters 3, 4, and 5 for comparison. Initially, the multi-step ahead wind speed forecasting performance of eight commonly used statistical approaches is evaluated using four different case studies and three rolling windows for $n - 100$ observations. The reasons for erroneous wind speed forecasts are discussed in detail. Next, four enhanced models (including three from chapters 3, 4, and 5) were considered while addressing the shortcomings of conventional methods. In addition, four machine learning models are also analyzed for comparison. With four case studies, sixteen models, three rolling windows, and three forecasting horizons, the final quality-controlled database contains 576 entries. The outcomes of the comparisons are discussed, explaining the higher prediction accuracy of improved models. Further issues that restrict the utilization of statistical forecasting models are also mentioned as potential future areas.

Chapter 7. In this chapter, the summary of the contributions made by the author is concluded. Also, the future directions are highlighted that need subsequent research to improve forecasting performance further.

References

[1] "Statistical Review of World Energy," 2021, Available: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2021-full-report.pdf.

[2] "Renewable Capacity Statistics," 2021, Available: https://www.irena.org/Statistics/Download-Data, Accessed on: 30th Jul. 2021.

[3] *Wind energy to be 20% of NZ generation by 2035*. Available: https://www.windenergy.org.nz/20-of-nz-generation-by-2030#:~:text=NZWEA%20vision%20is%20that%20wind,up%20from%206%25%20in%202017.

[4] *Current status of wind power projects*. Available: https://www.aedb.org/ae-technologies/wind-power/wind-current-status

[5] M. Niu, S. Sun, J. Wu, and Y. Zhang, "Short-term wind speed hybrid forecasting model based on bias correcting study and its application," *Mathematical Problems in Engineering*, vol. 2015, 2015.

[6] 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.

[7] 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, 2019.

[8] H. Liu, C. Chen, X. Lv, X. Wu, and M. Liu, "Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods," *Energy Conversion and Management*, vol. 195, pp. 328-345, 2019.

[9] R. Tawn and J. Browell, "A review of very short-term wind and solar power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 153, p. 111758, 2022.

[10] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PloS one*, vol. 13, no. 3, p. e0194889, 2018.

[11] F. Petropoulos *et al.*, "Forecasting: theory and practice," 2022.

[12] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914-938, 2016.

[13] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renewable Energy*, vol. 37, no. 1, pp. 1-8, 2012.

[14] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 243-254, 2014.

[15] L. Xiao, J. Wang, Y. Dong, and J. Wu, "Combined forecasting models for wind energy forecasting: A case study in China," *Renewable and Sustainable Energy Reviews*, vol. 44, pp. 271-288, 2015.

[16] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Applied energy*, vol. 235, pp. 939-953, 2019.

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

CHAPTER 2

CURRENT PERSPECTIVE ON THE ACCURACY OF DETERMINISTIC WIND SPEED AND POWER FORECASTING

This chapter contains content from the following article.

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. DOI: https://doi.org/10.1109/ACCESS.2019.2951153

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 20, 2019, accepted October 29, 2019, date of publication November 4, 2019, date of current version November 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2951153

Current Perspective on the Accuracy of Deterministic Wind Speed and Power Forecasting

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 Department of Mechanical and Electrical Engineering, Massey University, and 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'' Grant 5-1/HRD/UESTPI(Batch VI)/6082/2019/HEC.

ABSTRACT The intermittent nature of wind energy raised multiple challenges to the power systems and is the biggest challenge to declare wind energy a reliable source. One solution to overcome this problem is wind energy forecasting. A precise forecast can help to develop appropriate incentives and well-functioning electric markets. The paper presents a comprehensive review of existing research and current developments in deterministic wind speed and power forecasting. Firstly, we categorize wind forecasting methods into four broader classifications: input data, time-scales, power output, and forecasting method. Secondly, the performance of wind speed and power forecasting models is evaluated based on 634 accuracy tests reported in twenty-eight published articles covering fifty locations of ten countries. From the analysis, the most significant errors were witnessed for the physical models, whereas the hybrid models showed the best performance. Although, the physical models have a large normalized root mean square error values but have small volatility. The hybrid models perform best for every time horizon. However, the errors almost doubled at the medium-term forecast from its initial value. The statistical models showed better performance than artificial intelligence models only in the very short term forecast. Overall, we observed the increase in the performance of forecasting models during the last ten years such that the normalized mean absolute error and normalized root mean square error values reduced to about half the initial values.

INDEX TERMS Deterministic, wind speed, wind power, forecasting accuracy, normalized statistical indicators.

I. INTRODUCTION

In recent years, wind power is the most competitively priced technology in many markets. According to Global Wind Energy Council (GWEC) Annual Report 2018 [1], the cumulative wind power installed during 2001 to 2018 is 591 GW that is expected to reach 908 GW by the end of 2023 as shown in Fig. 1. Despite providing more than half of renewables growth [2], the intermittent nature of wind raised multiple challenges to the power systems and is the biggest challenge to declare wind energy a reliable source.

The challenges that raised to the power system due to the intermittent nature of wind includes planning and operational difficulties, quality of power, and standard of inter-connections. For example, the system operator needs to allocate additional energy reserves in case any power

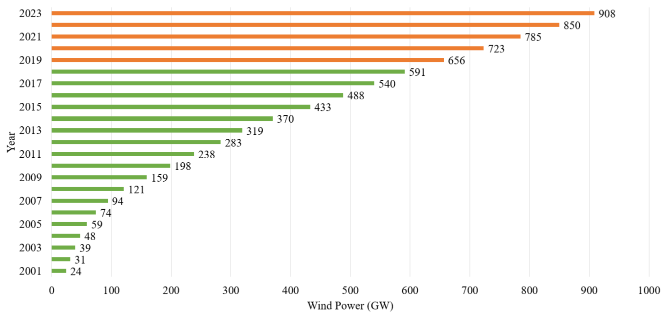


FIGURE 1. Cumulative installed wind power for 2001-2018 and forecast for 2019-2023 according to GWEC [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Ching-Ter Chang.

fluctuation occurs between programmed and actual power produced. This additional reserves would increase the operational costs, which subsequently increases the final energy prices [3]. Albadi and Saadany discussed a detailed review of wind power intermittency impacts on power systems [4].

One solution to overcome this problem is wind energy forecasting. A precise forecast would help to develop appropriate incentives and well-functioning hour-a-head or day-ahead electric markets [5]. Reliable forecasts help system operators to integrate wind energy into the grid with lesser complications. Literature suggests that accurate wind speed and power forecasting is a significant factor for various wind applications varying from siting till integration [6]. Even Liu *et al.* [7] predicted wind for railway warning systems for train protection. Some of the standard applications discussed in the literature are siting and designing of wind farm, grid integration and operations (dispatch planning, unit commitment decisions, farm regulations, maintenance scheduling, energy storage, reserve planning), stability (power stability, reducing breakdown probability) and revenue generation (tariff in the electricity market, electricity bidding and trading).

Forecasting is a way of predicting future events and is seen as a method of extrapolation. Forecasting process includes defining a problem, collecting data, analyzing data, selecting and fitting model to a set of data, validating the model using new data, model deployment, and performance evaluation [8]. Wind energy forecasting depends on cross-disciplinary approaches, including mathematics, statistics, meteorology, and power systems engineering [9]. There are two subcategories of wind speed and power forecasting: deterministic and probabilistic. Deterministic forecasting helps in evaluating point forecast for a specific time horizon while the probabilistic forecasting provides confidence intervals for the uncertainty of wind energy. As mentioned by Liu *et al.* [10], deterministic forecasting is the principal research direction of many scholars. Therefore, deterministic wind speed and power forecasting is the study focus of this review paper. Comprehensive reviews on probabilistic wind power forecasting are available in the literature [11], [12].

A. OBJECTIVES AND MOTIVATIONS

The motivation of this study is twofold. In recent years, various review papers are available on wind energy forecasting. Foley *et al.* [13] presented a review of the physical and statistical models. Tascikaraoglu and Uzunoglu [14] and Xiao *et al.* [15] reviewed the hybrid wind forecasting models based on weighted, decomposition (pre-processing), feature & optimization, and error processing (post-processing) approach. Qian *et al.* [16] further discussed the three structures of decomposition (pre-processing) approach, whereas Bokde *et al.* [6] reviewed the Empirical Mode Decomposition (EMD) based hybrid models. All these reviews are of significant importance in the field. However, in recent years, many models are updated. For example, Wind Power Prediction Tool (WPPT) was analysed as a statistical model earlier [13]; however, WindFor has replaced WPPT, which is the combination of advanced learning methods with a physical model [17]. Similarly, a comprehensive review of probabilistic wind power forecasting is presented [11], but no comprehensive study covering recent developments in deterministic forecasting methods has carried out.

Secondly, the papers published are limited to a specific site(s). The datasets are non-identical, and forecasting model is different in step size and location, which limits the comparison and applicability of the suggested model for the other regions. For comparative analysis and general conclusions, it is necessary to test the model performance for numerous case studies with diversified climatic regions. Some other review papers also mentioned the same perspective, but no detailed study was carried out.

In contrast with recent review papers, the major contribution is to present a comprehensive review of deterministic wind speed and power forecasting models from all the major perspectives. We explore the detailed classifications of wind speed and power forecasting and discussed the performance and limitations of forecasting models adopted in recent years. Also, we relate the performance and trend of recently developed forecasting models to present the performance of deterministic wind forecasting models for power generation. The motivation of this study is the review article of Blaga *et al.* [18], in which authors presented a detailed review of the performance evaluation of solar irradiance forecasting models based on available statistical indicators. After analyzing a large number of papers published between 2010-2019, twenty-eight papers ([57]–[58], [61], [72], [77]–[79], [82]–[83], [91], [97]–[114]) were shortlisted based on following criteria: the model performance is reported in terms of normalized mean absolute error (nMAE), and normalized root mean square error (nRMSE). In case the results are not presented in normalized values, then the statistical indicators of wind data must be listed in the paper. The normalized values help for inter-comparison analyses. After the selection of papers, we analysed a total number of 634 entries consisting of pair of nMAE and nRMSE. We investigated the study from three perspectives: forecasting models, time scale and performance trend over time. A brief description of shortlisted papers is enlisted in Table 3. First, we shall introduce the major classifications of the existing wind speed and power forecasting models in the next section. Later on, we shall analyse the performances of wind energy forecasting models, reported in shortlisted research articles, in section III to present inter-comparison analyses.

II. DETERMINISTIC WIND SPEED AND POWER FORECASTING CLASSIFICATION

In this section, we classify wind speed and power forecasting according to input data, time-scale, power output and forecasting method. Fig. 2 presents the overall classification of deterministic wind speed and power forecasting.

TABLE 1. Available global and regional NWP models.

Model	Developer	Model Type	Horizontal Resolution	Vertical Level	Forecast Length	Reference
ICON (ICOsahedral Nonhydrostatic)	DWD (Deutscher Wetterdienst)	Global	13 km	90	7 days	[22]
COSMO-DE (Consortium for small-scale modelling)		Regional	2.8 km	50	27 hr	[23]
IFS-HRES (Integrated Forecasting System - High RESolution)	ECMWF (European Centre for Medium-Range Forecast)	Global	9 km	137	10 days	[24]
SEAS5 (Integrated Forecasting System - SEASonal set V)		Global	36 km	91	13 months	[25]
Deterministic Global	UK Met Office	Global	10 km	70	6 days	[26]
Deterministic UK		Regional	1.5 km (inner) 4 km (outer)	70	120 hr (inner) 54 hr (outer)	[26]
GSM (Global Spectral Model)	JMA (Japan Meteorological Agency)	Global	20 km	100	84 hr	[27]
LFM (Local Forecast Model)		Regional	2 km	58	9 hr	[28]

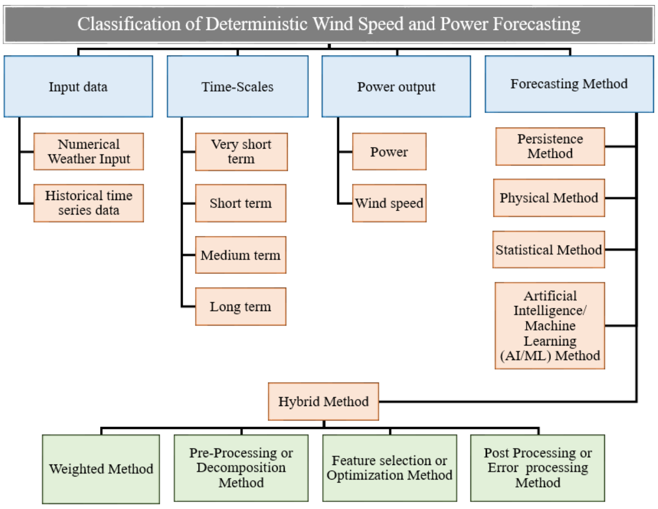


FIGURE 2. Detailed classification of deterministic wind speed and power forecasting.

A. INPUT DATA

There are two subclasses according to input data: Numerical Weather Prediction (NWP) input and historical time series data.

Meteorologists have developed NWP models to simulate the Earth's atmosphere to predict the weather. NWP model is a numerically approximate solution, based on equations associated with atmospheric processes and it changes. The primary equations are conservation of mass, conservation of energy, conservation of momentum, conservation of water and equation of state [19]. In NWP models, the atmosphere is divided into 3D cubes having a horizontal and vertical model resolution. The horizontal resolution presents orography, whereas vertical resolution presents weather phenomenon. The size of the resolution profoundly influenced the model. For example, a coarse resolution provides only limited details of valleys and height of mountains. The higher resolution provides better prediction but on a cost of more computational time. NWP models are available for both global and regional level. Table 1 presents a description of some of the available NWP models.

Most statistical methods use historical time speed data to correlate the wind speeds of the site. A mast is installed at the wind farm with at least one anemometer mount at the hub height to measure a minimum of six months data. Also, meteorological departments of countries and some global sources managed to record the data.

The primary benefit of NWP input is their applicability to predict long term horizon. The models using NWP data can provide forecasts for several days as well as several wind farms. Commercial procedures and software are available for NWP models. However, these models loss their applicability as the prediction horizon decreases, especially to predict very short term winds. One possible reason for this is the high wind variation that affects the model performance. Also, these models are complex to construct and need higher time to operate. In case of insufficient grid resolution, NWP models might contain systematic errors due to lack of handling sub-grid phenomena or physical parameterization.

In contrast to NWP data, time-series data requires lesser computational resource and time to model and operate. The traditional approach for long term forecasting is to use Measure–Correlate–Predict (MCP) approach. MCP approach takes into account the wind speed measurements at the wind farm and correlates with long term meteorological station data using a linear regression technique. However, several problems are associated with time-series data; the planning of meteorological mast, availability of suitable and calibrated weather station, and precise measurements from the meteorological station. Most importantly, the high cost associated with the weather stations resulting in a limited number of meteorological stations run in many countries.

B. TIME-SCALES

The time scale for a forecast depends on end-user requirements, technical conditions and regularity situations. The forecasting limits, according to time scales, are not well defined in the literature. However, keeping in view the literature, we divide the time horizon into four categories: Very short term forecasts (0-30 min), short term forecasts (30 min - 6 hours), medium-term forecasts (6 hours - 1 day ahead) and long term forecasts (>1 day ahead) [20], [21].

Very short term forecasts vary from few seconds to 30 minutes ahead. The major applications include wind turbine regulation and control strategies, electricity market clearing and real-time grid operation. These forecasts are possible based on time series data and do not require NWP data. Short term forecasts comprise of 30 minutes to 6 hours ahead. This category factored into economic load dispatch planning, operational security in the electric market and load decisions for increments. Online measurement data from the meteorological station, numerical weather prediction (NWP) or combination of both is used as input data, expecting that the weather condition will remain the same in short time horizon. However, the impact of NWP data is the least. Medium-term forecasts cover 6 hours to one day ahead and applied for decision making of unit commitment, reserved requirement and generator operation. NWP data is necessary for the medium-term forecast. Long term forecasts comprise of one day or more ahead. These forecasts use in maintenance scheduling, optimizing operational cost and feasibility study for designing a wind farm. Long term forecast necessarily requires NWP data for accurate estimation. In most of the literature, the performance of forecasting models is evaluated based on mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE). MAPE is a relative error and evaluates the ratio between residuals and actual values. A smaller error in the lower winds may have a smaller effect on MAPE may have a larger effect or a larger error in the higher winds MAE determines the difference between the actual and the estimated values. This performance evaluator is more robust to large errors. RMSE also evaluates the model dispersion but is very sensitive to the large errors due to the squared values.

The performance accuracy decreases as the time horizon increases. MAE for 40 min, 50 min and 1 hr ahead predictions was reported as 6.419 m/s, 7.085 m/s and 7.712 m/s [29]. Even in some cases, a lesser increase in forecast length results in a greater reduction in forecast accuracy. The MAPE for 10 s forecast length was 5.92 %, which increased to 7.81% for 20 s [30].

C. OUTPUT

There are two ways to get the forecast output. The first way is to forecast wind power generation directly from supervisory control and data acquisition (also termed as direct method). The second method is to forecast wind speed first, and then power curves are used to convert these forecasts into wind power as a next step (also termed as indirect method).

Kusiak *et al.* [31] applied the kNN model for both of the cases and concluded that the direct model offers better prediction performance than the indirect model. For the same dataset, MAE varied from 8.41% to 11.49% for direct method, whereas MAE varied from 9.67% to 12.72% for indirect method. Similarly, Renani *et al.* [32] also compared both direct and indirect approaches for a case study of a wind farm in Northern Iran. The analysis showed that errors increased by more than 100%. More specifically, the MAPE for 5 min, 15 min, 30 min and 60 min was 1.47%, 1.37%, 1.30% and 1.48% respectively in case of direct prediction which increased to 3.41%, 3.17%, 3.22% and 3.62% respectively in case of indirect prediction. The larger errors were due to the integration of two errors: one at wind speed prediction and second at wind power prediction. However, according to the argument of Zhu and Genton [33], the indirect method is a better approach than the direct method as the nearby wind farms with different wind turbines will experience the same wind speed. Therefore, it is better to convert the standard wind speed forecasts to their respective power curves, instead of performing individual wind farm power forecasts. Hong *et al.* [34] also supported the argument and discussed that wind power is dependent on multiple factors, including orography, wind speed, direction, and wake effects. Also, the rapid fluctuations and randomness in wind power data of a single wind farm will not guarantee to mine the wind regularity. Therefore, it is better to apply an indirect method which does not require correlation analysis of wind with other factors.

D. FORECASTING MODEL

1) PERSISTENCE METHOD

Persistence method (also termed as 'Naïve Predictor') is based on a high correlation between the present and immediate future wind speed. In this method, the wind speed at a time $(t + \Delta t)$ is assumed to be the same as was at the time (t), i.e.

$$v\,(t + \Delta t) = v(t). \tag{1}$$

The persistence method shows good accuracy when dealing with very short term forecasts. Wegley *et al.* [35] analysed three forecasting models: persistence, autoregressive (AR) and generalized equivalent Markov (GEM) on spring season data of Oklahoma City for 10, 30 and 60 minutes time interval and concluded that persistence method is superior in the 10-minute time interval. Authors suggested that AR and GEM can upgrade for further improvement, but persistence cannot. Also, the accuracy of the persistence method degrades rapidly as time increases. The method is served as a benchmark to compare improvement for newly developed forecast models [5], [36].

2) PHYSICAL METHOD

The physical method requires meteorological and other factors such as pressure, temperature, local surface roughness, obstacles, and wind turbines power curves for prediction.

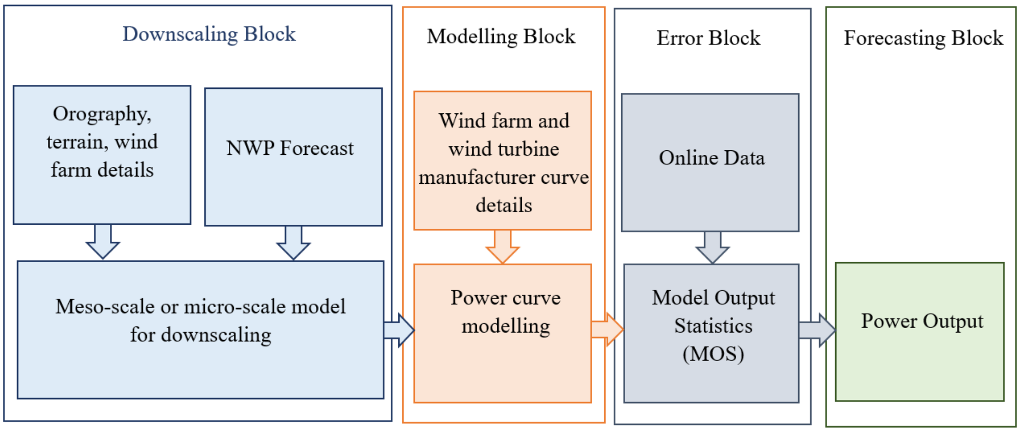


FIGURE 3. Flowchart for the physical method.

The physical methods are of two types: Diagnostic Model and Computational Fluid Dynamics (CFD) model. Diagnostic models [13] use parameterizations of boundary layer whereas CFD models simulate the wind flow fields dynamically. Diagnostic models are suitable for flow over flat terrain, whereas CFD models are appropriate for flow over complex terrain [37], [38].

The commercial methods for wind power forecasting use NWP wind forecasts as the input data and then carry out the necessary refinement of these output data (wind speed forecast) to the on-site conditions. The physical methods use a mesoscale or microscale model for the downscaling [39] for interpolating wind speed forecasts to the hub height of the wind farm. The attainable resolution and range of domain size differentiate the meso and micro model. The forecasted wind speed is used to estimate power. The easiest way is to utilize the manufacturer's power curve. Also, the Model Output Statistics (MOS) approach corrects the scaling errors. Fig. 3 illustrates the overall process.

Physical models incorporate orography, thus enables physical behavior understanding. These models generate regional and global forecasts using initial conditions to solve complex numerical systems. The historical data is of lesser importance in such models. However, to accurately predict the winds, it is necessary to have extensive information on surface roughness and characteristics of wind farms. Thus, these models need extensive efforts to set up. Table 2 provides details of some commercially available physical wind power forecasting models.

3) STATISTICAL METHOD

Statistical methods use time-series data to find out the relations generally by recursive techniques [14]. These models are easy and cheaper to build and provide precise predictions when dealing with short term forecasting. NWP input is optional for these models, as shown in Fig. 4. The accuracy of the statistical model degrades as time increases. These models are based on patterns and do not use any predefined mathematical model [5]. Statistical methods include

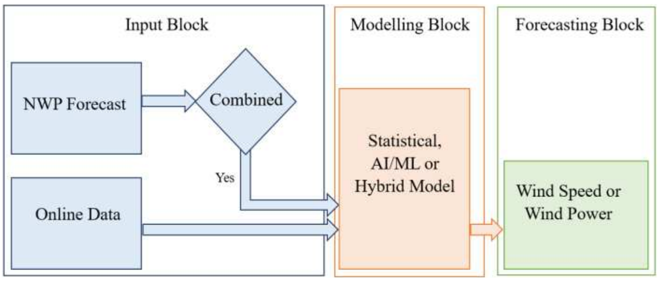


FIGURE 4. Flowchart for the statistical method.

autoregressive moving average (ARMA) [40], autoregressive integrated moving average (ARIMA) [41], fractional-ARIMA [42], seasonal-ARIMA [43], ARMA with exogenous input (ARMAX) [44], grey predictors [45], and exponential smoothing [46].

ARIMA models are the most commonly used statistical models. The general non-seasonal model structure form is ARIMA (p, d, q) where p is the order of autoregressive (AR) part, d is the degree of differencing taken to make time-series stationary, and q is the order of moving average (MA). The linear expression of ARIMA (p, d, q) is expressed in the form:

$$y_t = c + \left(\sum\nolimits_{i=1}^{p} \phi_i y_{t-i} + \sum\nolimits_{j=1}^{q} \theta_j \varepsilon_{t-j}\right), \tag{2}$$

where c is the constant term, ϕ_i is the coefficient of the i^{th} autoregressive parameter, θ_j is the coefficient of the j^{th} moving average parameter, y_{t-i} is the value at a time $(t-i)$, and ε_{t-j} is the error between the predicted value and actual value at $(t-j)$. ARIMA is a three-step iterative process. First, a tentative model is identified by analysing the time series data. Second, the unknown parameters are estimated. Third, the adequacy of the model is inspected through residual analysis. The residual analysis assists in performing the diagnostic checks or in specifying the potential improvements.

In comparison with physical models, statistical models do not require any real insight. Therefore, these models are easy to build and fast to calculate. However, for these models,

TABLE 2. Description of commercially available physical, statistical and combined models.

Prediction Model	Developer	Model Type	Description	Prediction Horizon	Ref
SOWIE (Simulation Model for the Operational Forecast of the Wind Energy Production in Europe)	EuroWind	Physical	SOWIE is a multi-model forecast that utilizes NWP data includes GFS, ECMWF, HIRLAM, and UKMET. It provides forecasts from single wind farm to regional level.	180 hrs	[47]
Previento	University of Oldenburg	Combined	It is the optimal combination of different NWP models input. This model uses weather services data from multiple services. Optimization is carried out using Kombibox procedure, i.e. lesser error predictions will be provided more weight.	15 days	[48]
WPMS (Wind Power Management System)	Fraunhofer Institute for Energy Economics and Energy System Technology	Statistical	WPMS is operational in three European countries. WPMS is based on Artificial Neural Network (KNN) and uses extrapolation algorithm to upscale the prediction of the wind farm to the regional level.	96 hrs	[49]
DNV GL	DNV GL	Statistical	Provide four services: Forecaster Now, Live, Plus and Solutions. The model works on auto-adaptive algorithms and uses SCADA data for MOS correction. Forecasts are available for a single plant as well as the market aggregate level.	15 days	[50]
EPREV	Prewind	Statistical	The model incorporates SCADA data and employs three statistical models: PCM, AR and NNAM. The model can forecast the wind power of a single wind turbine, wind farm or a region.	168 hrs	[51]
AleaWind	Alea Soft	Statistical	The model works on Neural Networks and Genetic Algorithms.	10 days	[52]
Scirocco	Aeolis Forecasting Services	Combined	The model is powered by NWP data and involves three adjustment schemes: Model Output Statistics, physical terrain and Model Chain Output Statistics. Calibration is done through error backpropagation. With ECMWF, the model can forecast long term winds.	15 days	[53]
WindFor (Former WPPT)	ENFOR	Combined	WindFor is a combination of advanced learning methods with a physical model. It can be initiated through time-series data or manufacturer's wind power curve. It can accurately predict short term data when integrated with the SCADA system.	15 days	[17]

the performance is highly dependent on the accuracy of the available data. Also, the lesser number of observations can limit the model performance. Furthermore, statistical models cannot deal with nonlinear conditions. Table 2 describes some commercially available statistical models.

Both the physical and statistical methods have performance limitations in different time horizon. Physical models predict a long-term wind precisely whereas the statistical models have high precision in short-term prediction. Therefore a combination of both will improve the performance of wind power forecasting. Physical models predict the long term trend, whereas statistical models improve the precision of local prediction. Table 2 provides details of some commercially available combine models.

4) ARTIFICIAL INTELLIGENCE/ MACHINE LEARNING METHODS

Artificial Intelligence/ Machine Learning (AI/ML) techniques are the most popular method for wind speed and power forecasting. These techniques train past data to find out the relationship between input and output wind-speeds. Common AI techniques include Artificial Neural Network (ANN) [54], Support Vector Machine (SVM) [55], and Adaptive Neuro-Fuzzy Inference System (ANFIS) [56]. The most commonly used AI model is ANN.

ANN is motivated by the way the human brain would solve the problem. The general form of ANN is a black-box approach and is used to handle non-linear data. A typical ANN has three layers: input layer (the original predictors), one or more hidden layer (set of constructed variables) and output layer (the responses).

Each variable in the layer is termed as a node. In the first step, a weight is used to measure the strength of each connection. The input nodes are multiplied by associative weights, and the net is summed up as in (3):

$$net = w_1x_1+w_2x_2+w_3x_3+\ldots+w_nx_n = \sum\nolimits_{i=1}^{n} w_ix_i. \quad (3)$$

Next, an activation (or transfer) function f is chosen to transform the net signal of each node i. Mathematically, the output form is:

$$output = f(net + b), \tag{4}$$

where b is the bias term, also called the activation threshold for the corresponding node. It is an offset value that regulates the signal and is same as the intercept term in the regression model. The activation function is typically a non-linear and the selection depends on the nature of the response variable. The commonly used activation function is logistic or hyperbolic tangent function.

SVM is based on the Statistical Learning Theory (SLT) and Structural Risk Minimization (SRM). SVM for data regression (SVR) maps the input data into a high-dimensional feature space through nonlinear kernel function and then generates a linear regression function in this hyperspace. The linear regression function is expressed as in (5):

$$f(x) = \sum_{i=1}^{n} w_i \varphi_i(x) + b, \tag{5}$$

where w is the associative weight, b is the bias term and $\varphi(x)$ is the mapping function that maps x into high dimensional feature space. The regression is then expressed by the optimization problem and is solved by the quadratic programming technique. Finally, the estimation function is obtained as in (6):

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^{n} (\alpha - \alpha^*) k(x_i, x) + b, \tag{6}$$

where α and α^* are the Lagrange multipliers and $k(x_i, x)$ is the kernel function. Different kernel functions are used in SVM models. The commonly used kernel function is a Radial Basis Function (RBF). There are other variants of SVR employed in wind speed and power forecasting including Least Square SVM (LSSVM) [57], Twin SVR (TSVR) [58] and Reduced SVM (RSVM) [59].

ANFIS is a class of adaptive multilayer feedforward network that integrates fuzzy logic principles and neural networks. It develops fuzzy rules with suitable membership functions to produce required inputs and outputs. The ANFIS model has five layers: fuzzification, rule evaluation, normalization, defuzzification and summation. Initially, the system designer sets the learning rules and membership functions based on expertise, and later ANFIS adjusts the rules and functions to minimize the output error index. Most commonly utilized membership function is bell-shaped.

Other than traditional machine learning models, extreme learning and deep learning is gaining much more attention in wind speed and power forecasting. These advanced learning model showed higher accuracy and can learn more complex nonlinear relations. Some notable architectures of deep learning utilized in wind speed and power forecasting includes Deep Belief Network (DBN) [60] and Long Short Term Memory (LSTM) [61].

Extreme Learning Machine (ELM) is a type of feedforward neural network with a single hidden layer. It has better generalization performance and higher convergence speed than the traditional neural network. In ELM, the input weights and hidden biases are generated randomly without iterative tuning. Therefore, the output weights between hidden and output layers are determined as finding the least square solution to the given linear system. Other variants of ELM utilized in wind speed and power forecasting include Hysteresis ELM (HELM) [61], Online Sequential ELM (OSELM) [62], Stacked ELM (SELM) [63], Regularized ELM (RELM) [64], and Weighted RELM (WRELM) [65]. Reference [66] discussed in detail the trends in ELM.

DBN is a multi-layered stochastic generative model, constructed by stacking multiple Restricted Boltzmann Machines (RBMs) [67]. RBM is an undirected bipartite graphical model in which visible observations (v) are connected to stochastic binary hidden units (h) using undirected weighted connections (w_{ij}) [68]. It is characterized by the energy function $E(v, h)$, defined as in (7):

$$E(v, h) = -\sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij} v_i h_j - \sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j \tag{7}$$

where a_i and b_j are the biases, and n and m are the numbers of neurons in the visible and hidden layers, respectively. The joint probability distribution of visible and hidden layers is expressed as in (8):

$$p(v, h) = \frac{e^{-E(v,h)}}{\sum_v \sum_h e^{-E(v,h)}} \tag{8}$$

For RBM, the individual activation probability of h_j or the conditional probability of v_i is expressed as:

$$P(h_j = 1 | v) = \frac{1}{1 + exp\left(-\sum_{i=1}^{n} w_{ij} v_i - b_j\right)} \tag{9}$$

$$P(v_i = 1 | h) = \frac{1}{1 + exp\left(-\sum_{j=1}^{m} w_{ij} h_j - a_i\right)} \tag{10}$$

The unknown parameters can be determined by training the model.

Deep Boltzmann Machine (DBM) is also based on stacked RBM and is applied in wind speed and power forecasting [69]. In contrast with DBN, all connections are undirected in DBM. In DBN, the top two layers have undirected connections, whereas the lower layers have directed connections [70].

LSTM is a variant of Recurrent Neural Network (RNN) and has a stable and excellent ability to solve long-term dependencies. In LSTM, the traditional node of the hidden layer is replaced by a memory cell (a core component of LSTM). The memory cell acts as an accumulator of state information. LSTM has three gates: input (write), output (read) and forget (reset), through which the state information is updated as: The information of incoming input will be accumulated to the cell if input gate is activated. The prior cell status will be forgotten if forget cell is activated. The latest cell output will be propagated to the final state if the output gate is activated.

In comparison with statistical models, AI/ML models have stronger nonlinear estimation ability. However, the problems associated with AI/ML models are slower convergence speed, overfitting, computational complexity, slow speed and generalization problems. Most commonly used AI/ML model is ANN that exhibits overtraining. When the training capacity is too large, it allowed too many iterations that caused over-training.

Both the statistical and AI/ML methods have limited applicability and thus limited prediction accuracy. Therefore, combining both statistical and AI/ML models have better prediction accuracy.

5) HYBRID METHOD

Hybrid forecasting methods take advantage of combining different forecasting methods to improve the performance of the final forecast. An individual model has limited performance in multiple situations. The hybrid model provides superiority as it utilizes capabilities of the individual model and therefore saves time with the better performance [14]. We used the same definition of hybrid models as discussed by Tascikaraoglu and Uzunoglu [14] and Xiao *et al.* [15]. The subclasses of hybrid methods include weighted method, preprocessing or decomposition method, feature selection or optimization method, and postprocessing or error processing method.

In the weighted method, a weight coefficient is assigned to each individual model forecast based on model effectiveness. There are two different arrangements of the weighted method hybrid model: Fixed weight and Variable Weight. The coefficient can be calculated in different ways. This may be weighted average [71], weighted median [71], or varied weights based on optimization algorithm [72]. According to a study conducted by Li *et al.* [72], the variable weight forecasts show better performance than fixed weight forecasts. The variable weight combination forecasting model can better adapt to changes in the sample, and match the weight of the sample points in the corresponding model. Multiple optimization algorithms are utilized to determine the optimal weights. Zhang *et al.* [73] applied CLSFPA (Flower Pollination Algorithm with Chaotic Local Search) to calculate optimum weights with NNCT (No Negative Constraint Theory) and compared the results of the combined model with other single prediction models (BPNN, RBFNN, ENN, GRNN, WNN and ARIMA) at four sites. Similarly, Li *et al.* [72] used BA (Bat Algorithm) with NCFM (Novel Combined Forecasting Model), Xiao *et al.* [15] applied CPSO (Chaos Particle Swarm Optimization) and GA (Genetic Algorithm) with NNCT whereas Okumus and Dilner [74] used LSM (Least Square Method) with FNN and ANFIS. All these models showed better performance than the individual prediction models. Instead of focusing on a single objective optimization algorithm, some researchers have focused on the multi-objective optimization algorithm. Niu and Wang [75] have applied MOGOA (Multi-Objective Grasshopper Optimization Algorithm) to calculate weight coefficient and compared the results with models based on CS (Cuckoo Search) algorithm and FA (Firefly Algorithm). Results showed that MOGOA based model performed very well for the considered five sites at 10 min, 20 min and 30 min prediction ahead followed by FA and CS.

Most of the hybrid models reported in the literature are decomposition-based approaches. In the decomposition method, pre-processing techniques are applied to decompose the non-stationary time series data into stationary subseries. Decomposition approaches widely reported in the literature including Wavelet Transform (WT) [58], Wavelet Packet Decomposition (WPD) [76], Empirical Mode Decomposition (EMD) [72], variants of EMD including Ensemble EMD (EEMD) [77], Fast EEMD (FEEMD) [64], Complementary EEMD (CEEMD) [78], Complete EEMD with Adaptive Noise (CEEMDAN) [73], Improved CEEMDAN (ICEEMDAN) [79], Intrinsic Time Scale Decomposition [80], Seasonal Adjustment Methods [81], Variational Mode Decomposition (VMD) [82], Optimized VMD (OVMD) [83], Empirical Wavelet Transform (EWT) [84], and Improved EWT (IEWT) [85]. There are two subtypes of decomposition: primary and secondary. In the primary arrangement, a decomposition model is used to decompose the non-stationary time series data into several stationary subseries at the time series and then a separate prediction model is on each subseries. In the latter arrangement, a secondary decomposition model is used to decompose further the most non-stationary subseries. The further method would be the same as the first arrangement. It is not necessary to use the same prediction model for the decomposed series. Han *et al.* [86] used Wavelet decomposition with ARMA and LSSVM to predict high and low-frequency subseries. Similarly, Zhang *et al.* [87] applied EEMD decomposition with SARIMA and ANFIS for periodic and nonlinear components modelling.

Feature selection and optimization technique are applied to remove the redundant data, thus improves the model performance. Several optimization algorithms are reported in the literature. Wang et al. used GA to optimize BP [88], Meng *et al.* used CSO (Crisscross Optimization) to optimize BP [89], Liu *et al.* used GA and MEA (Mind Evolutionary Algorithm) to optimize MLP (Multi-Layer Perceptron) [90], Kong *et al.* used PSO to optimize RSVM [59] and Osório *et al.* [91] used EPSO (Evolutionary PSO) to optimize ANFIS. Feature selection for unsupervised learning can further classify into two methods: wrapper and filter. The wrapper approach uses a search algorithm to rank the feature subset. This method requires a prediction model performance. The subset that shows the best prediction performance is selected as the final feature subset. The filter method uses arithmetic analysis and does not require prediction model performance. Therefore, filter methods are faster than wrapper methods, but the performance is worse than the wrapper method. This argument is supported by the study conducted by Carta et al. [92]. In this study, the authors analysed both Wrapper and Filter method for a case study of Spain. *CfsSubsetEval* is used as filter whereas *WrapperSubsetEval* is used as a wrapper method. Analysis of two years of data for five stations showed that the wrapper approach provided lower mean errors than the filter method in all of the cases. It was also concluded that the wrapper method is more significant when non-linear relation between features increases such as wind direction significance in complex terrain. This improved performance is achieved on a cost of higher computational time. On the other hand, if this relation is not of the higher-order, then the Filter method has the advantage of lower computational time without losing prediction accuracy. Some studies discussed combine method of wrapper and filter, thus take advantages of both methods. It uses filter algorithm information to accelerate wrapper algorithm convergence [82].

Studies based on post-processing techniques considered the influence of error factors on the performance of the model. The purpose is to analyse the errors after primary prediction model and then incorporate a post-processing model. The results of the post-processing model help to improve the initial forecast results in producing the final forecasts. Hao and Tian [93] discussed a two-stage forecasting model in which error factor is considered. In the first stage, VMD is used as the decomposition model and ELM optimized by MOGWO (Multi-Objective Grey Wolf Optimization) is used as a prediction model for the forecasting error. In a second stage, the nonlinear ensemble method is developed to integrate variational modes and forecast error predictors to get the final forecast. Comparisons are made between individual statistical and ANN models, a single decomposition model and the proposed model. The proposed model significantly increased the model accuracy.

References [14], [15], [20] provide details of sub-classes of hybrid models.

III. METHODOLOGY

A. PERFORMANCE INDICATORS

The forecast results are comprehensively evaluated based on several performance indicators. Jiang *et al.* [77] discussed three aspects of evaluation metrics: accuracy, stability and direction. MAE and RMSE are used to evaluate the accuracy, variance to measure the stability and direction-accuracy to estimate correctness. However, most of the studies inferred the results based on accuracy indicators only. The commonly used mean absolute error and root mean square error are defined as in (11) and (12), respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} \left|y_i - \hat{y}_i\right|, \tag{11}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(y_i - \hat{y}_i\right)^2}, \tag{12}$$

where N is the number of entries, y_i is the i^{th} measurement, and $\hat{y}_i$ is the i^{th} forecasted value. The average of measured values is denoted by $\bar{y}$. These performance indicators evaluate quantitative measures and have similar physical units as the dependent parameter.

Inner-comparison of results, published in terms of absolute values, is not possible due to non-identical datasets. For comparing the accuracy results, it is necessary to convert the absolute values in normalized values. There are different normalization techniques available. The most common reference quantity is the mean value (μ) as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_i. \tag{13}$$

Other reference quantities include present value (y_i), deviation from average ($|y_i - \bar{y}|$) and dynamic characteristics ($|y_i - y_{i-1}|$) as reported by Gensler *et al.* [94]. In this study, we select all the choices of normalization. The accuracy sets, where normalized values are not available, we use mean value as the reference quantity: nMAE $=$ MAE/μ and nRMSE $=$ RMSE/μ.

Vargas *et al.* [95] presented a systematic review for wind power generation based on citation network analysis (CNA). According to the study, physical models had aroused during the 90's whereas AI/ ML and hybrid models have been emerging since the last decade. Also, most studies used hourly data frequency and more than half studies implement wind speed as input and output variable. Almost two-thirds of the studies are related to China. After analysing 143 articles spread over 33 years (1985-2018), the authors concluded that wind energy studies started growing considerably after 2010. Keeping in view this analysis and the procedure adopted by Blaga *et al.* [18], the following criteria are set for the selection of paper:

1) The publishing year of the paper must be in between 2010-2019.
2) The model performance is reported in terms of normalized Mean Absolute Error (nMAE) and normalized Root Mean Square Error (nRMSE).
3) In case the results are not presented in normalized values, then the statistical indicators of wind data must be listed in the paper.

We shortlisted twenty-eight papers for the analysis. The final database comprises of 634 entries spread over 50 locations covering ten countries. The description of selected papers is given in Table 3, and the overall summary is presented in Fig. 5 [96]. The papers are listed according to the publication year.

B. PERFORMANCE ANALYSIS OF THE DATASET

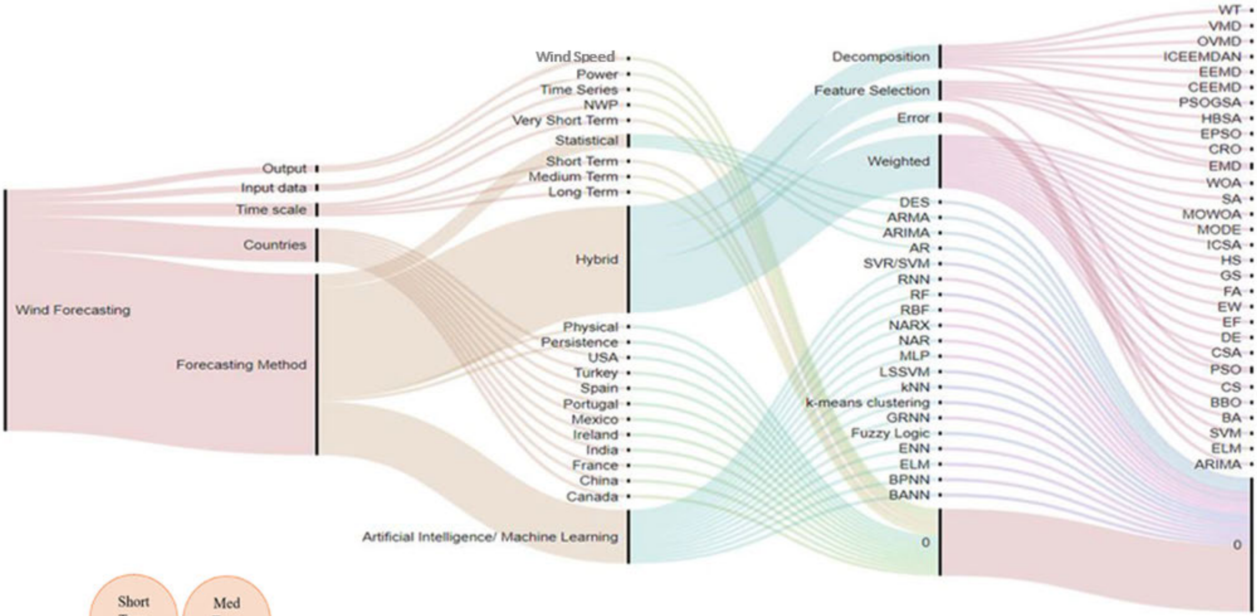
Fig. 6a summarizes the performance of prediction models in terms of averaged values. The statistical indicators are averaged over all entries of that specific paper. The 28 circles refer to 28 papers (as indexed in Table 3), the diameter of the circle reflects the number of data entries, and the color of the circle indicates the year of publication. For example, index 7 shows that the paper published in 2014 with 15 entries having averaged nMAE $= 13.31\%$ and averaged nRMSE $=$ 19.04%. Fig. 6a also shows the variability in data as some averaged values contain smaller nRMSE but larger nMAE and vice-versa. For example, index 19 and 20 both published in 2017 having almost the same number of data entries but contradictory averaged results. The averaged nMAE, and nRMSE values for index 19 is 3.19% and 5.33% respectively whereas averaged nMAE and nRMSE values for index 20 is

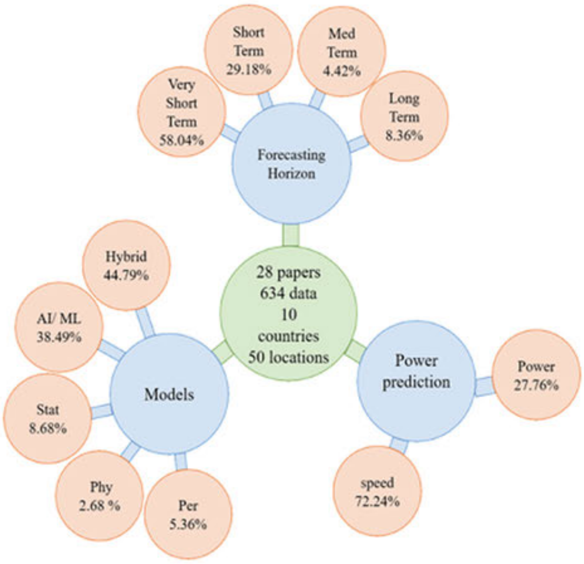
TABLE 3. Description of the papers included in the study.

Index	Year	Author	Ref	Input Data	Time scale	Forecasted variable	Forecasting Method	Country	Data Summary
1	2010	Cadenas and Rivera	[97]	TS	ST	wind speed	S, AI	Mexico	3 locations with 744, 672 and 720 observations respectively
2	2011	Blonbou	[98]	TS	VST	power	AI	France	2 locations with data ranging from July 2002 to March 2005
3	2011	An *et al.*	[99]	TS	VST	power	H	China	1 location with 470 points
4	2014	Wang *et al.*	[100]	TS	ST	wind speed	Per, AI, H	China	1 location with data ranging from January 1, 2010 to March 31, 2011
5	2014	Haque *et al.*	[101]	TS	MT, LT	power	Per, AI, H	USA	1 location with data from 2011
6	2014	Mandal *et al.*	[102]	TS	ST, MT	power	Per, AI, H	Canada	1 location with data ranging from January to December 2009
7	2014	Chen *et al.*	[103]	TS, NWP	MT	power	Per, S, AI, P	China	2 locations with data ranging from April 2010 to April 2013
8	2015	Liu *et al.*	[104]	TS	ST	wind speed	S, AI, H	China	1 location with data from July and August 2011
9	2015	Chitsaz *et al.*	[105]	TS	ST	power	Per, AI, H	Canada	1 location with data from 2012
10	2015	Osorio *et al.*	[91]	TS	ST	power	H	Portugal	1 location with data ranging from 2007-2008
11	2015	Wang *et al.*	[106]	TS	VST	wind speed	H, S, AI	China	8 locations with 1096 observations for four locations, 486 observations for one location, 551 observations for two locations, and 557 observations for the last one
12	2015	Ozkan and Karagoz	[107]	TS, NWP	LT	power	P, AI	Turkey	1 location with data ranging from June 2012 and the end of December 2012
13	2016	Zhang *et al.*	[108]	TS	ST	wind speed	AI, H	China	3 locations with data ranging from March 1, 2012 to March 31, 2012
14	2016	Abdoos	[82]	TS	VST, ST	power	AI, H	Spain, USA	1 location from each country with observations from 2014 (Spain) and 2006 (USA)
15	2016	Cadenas *et al.*	[109]	TS	ST	wind speed	Per, AI	Mexico	1 location with data ranging from May 1, 2006 to June 30, 2007
16	2016	Liang *et al.*	[110]	TS	ST	power	Per, S, AI, H	China	1 location with data ranging from April 19 to September 30, 2013
17	2017	Wang *et al.*	[78]	TS	VST	wind speed	Per, AI, H	China	1 location with 2880 observations
8	2017	Zhang *et al.*	[83]	TS	VST	wind speed	S, AI, H	China, Spain	2 locations from China with 600 observations and 1 location from Spain with data ranging from March 3 to March 12, 2017
19	2017	Liu *et al.*	[111]	TS	LT	power	AI	China	1 location with data from 2012
20	2017	Ranganayaki and Deepa	[112]	TS	ST	power	AI, H	India	1 location with data ranging from January 2010 to December 2014
21	2017	Feng *et al.*	[113]	TS	Short term	wind speed	Per, AI	USA	7 locations with data ranging from January 2015 to December 2015
22	2018	Lu *et al.*	[57]	TS	VST	wind speed	AI, H	China	1 location with data from 2016
23	2018	Li *et al.*	[72]	TS	VST	wind speed	H	China	1 location with 2304 observations
24	2018	Hu and Chen	[61]	TS	VST, ST	wind speed	S, AI	China	1 location with 720 observations
25	2018	Song *et al.*	[79]	TS	VST	wind speed	S, AI, H	China	1 location with data ranging from May 1, 2011 to May 18, 2011
26	2019	Aasim *et al.*	[114]	TS	VST	wind speed	Per, S, H	Ireland	1 location with data from December 2017
27	2019	Jiang *et al.*	[77]	TS	VST	wind speed	Per, S, AI, H	China	1 location with data ranging from 1 January to 20 January 2011
28	2019	Dhiman *et al.*	[58]	TS	VST, ST	wind speed	Hybrid	Spain, USA, India	1 location from Spain with data of October 2017, 4 locations of the USA with data ranging from January 1 to January 7, 2011and 1 location from India with data from January 2019

TS=Time Series, NWP= Numerical Weather Prediction, VST=Very Short Term, ST=Short Term, MT= Medium Term, LT= Long Term, Per=Persistence, P=Physical, S=Statistical, AI=Artificial Intelligence, H=Hybrid

9% and 7.18% respectively. Also, the number of entries is significant in some articles, whereas few of them shows only a handful of data. From descriptive statistics, the 95% confidence interval (CI) for the mean, for averaged nMAE and nRMSE varies between 6.73% to 10.07% and 8.276% to 12.550% respectively. Similarly, the 95% confidence





Acronyms: WT= Wavelet Transform, VMD=Variational Mode Decomposition, OVMD=Optimized Variational Mode Decomposition, ICEEMDAN= Improved Complete Ensemble Empirical Mode Decomposition Adaptive Noise, EEMD= Ensemble Empirical Mode Decomposition, CEEMD=Complimentary Ensemble Empirical Mode Decomposition, PSOGSA= Particle Swarm Optimization combined with Gravitational Search Algorithm, HBSA=Hybrid Backtracking Search Algorithm, EPSO=Evolutionary Particle Swarm Optimization, CRO=Coral Reef Optimization, EMD= Empirical Mode Decomposition, WOA=Whale Optimization Algorithm, SA=Simulated Annealing, MOWOA=Multi-Objective Whale Optimization Algorithm, MODE=Multi-Objective Differential algorithm, ICSA=Improved Clonal Selection Algorithm, HS=Harmony Search, GS= Gram-Schmidt, FA=Firefly Algorithm, EW=Equal Width, EF=Error Feedback scheme, DE=Differential Evolution, CSA= Clonal Selection Algorithm, PSO= Particle Swarm Optimization, CS=Cuckoo Search, BBO=Biogeography Based Optimization, BA=Bat Algorithm, SVM=Support Vector Machine, ELM= Extreme Learning Machine, ARIMA= Auto-Regressive Integrated Moving Average

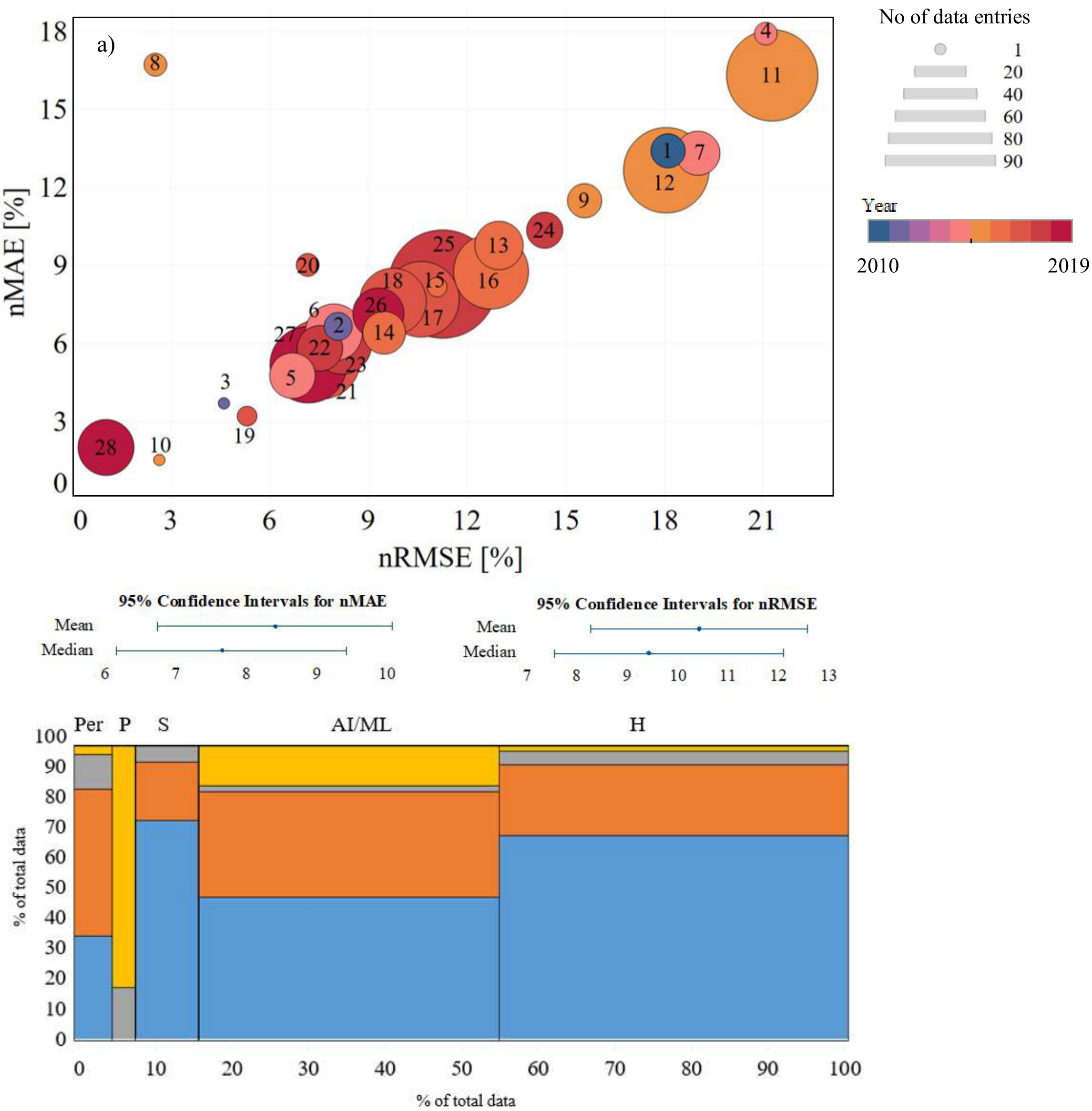
FIGURE 5. Overall Summary of the selected 28 papers based on forecasting classification. The percentages (%) indicate the frequency of data implemented for each sub category out of 634 data entries.

interval (CI) for the median, for nMAE and nRMSE varies between 6.14% to 9.41% and 7.548% to 12.085% respectively. Fig. 6b shows a two-dimensional histogram. The width and height of the bar show the relative size of datasets for forecasting horizon and predicting models respectively. For example, the Artificial Intelligence and Machine Learning (AI/ML) models comprised 38% of all data entries for which 48% is very short term, 36% is short term, 3% is the medium term, and 13% is long term.

Visual inspection displays that the physical models were used only for medium and long term forecasts, whereas statistical models hardly used for long term forecasts. In general, most of the work was done for very short term and short term forecasts using AI/ML and hybrid models. The number of data entries for AI/ML and hybrid models are considerably more significant than the others. This number is due to the eligibility criteria for the selection of papers. The authors of these papers either provided the mean values of input data or the normalized values of the output data in their studies.

Fig. 7a displays model performance per forecasting method and Fig. 7b per time scale. Visual inspection of Fig. 7a indicates that the physical models have the largest errors, whereas hybrid models have the smallest one. Also, the performance of AI/ML models is better than statistical models. From Fig. 7b, it is as per expectations that the performance decreases as the time horizon increases.

If we compare the interdependence of both graphs, we analyse that the most substantial errors are witnessed for the physical models because these models are applied primarily for medium-term and long term forecasts. On the other hand, the hybrid models are applied mostly for the very short term, and short term forecasts and therefore, these models showed the best performance. Persistence model and AI/ML



b)

FIGURE 6. **a) Illustrative plot presenting the performance of the model reported in the selected papers, b) 2-dimensional histogram presenting the qualitative picture of datasets.**

models show close results. This unexpected output is due to the significant number of entries for AI/ML models. AI/ML models contain seven times more data than persistence models. Especially for long term forecast, AI/ML models contain four times more data entries than persistence models. Fig. 7a and 7b are generated based on averaged values which do not provide in-depth knowledge for the spread of error. Box and whiskers plot (Fig. 8) and standard deviation present the spread of errors more clearly.

In box and whisker plots, the lower and upper values of the box indicate the interquartile range (IQR) corresponds to 25th and 75th percentile, whereas the whiskers extend to 1.5 times the IQR. The standard deviation is a measure used to quantify the amount of variation or dispersion of a dataset and is defined as:

$$\sigma = \sqrt{\frac{1}{N} \sum\nolimits_{i=1}^{N} (y_i - \mu)^2}. \tag{14}$$

From Fig. 8a and 8b, the length of the whiskers are the largest for the persistence model ($IQR_{nMAE} = 6.6434\%$, $IQR_{nRMSE} = 8.1072\%$). It shows that the variation range of error is largest for the persistence model. In terms of median values, persistence is the same as AI/ML models. However, the spread of errors is 12% more in the case of nMAE and 13% more in the case of nRMSE for the persistence model. It proves our previous argument that despite having close results in Fig. 7a, AI/ML models perform better than persistence. The length of the whiskers for the physical model,

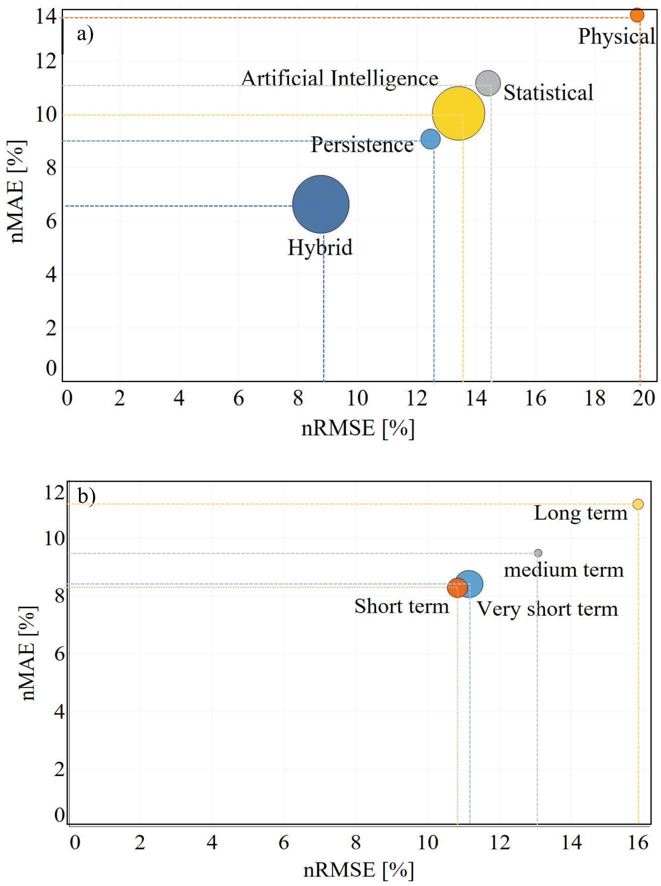


FIGURE 7. **The performance of a) prediction models, b) time scale, in terms of averaged values.**

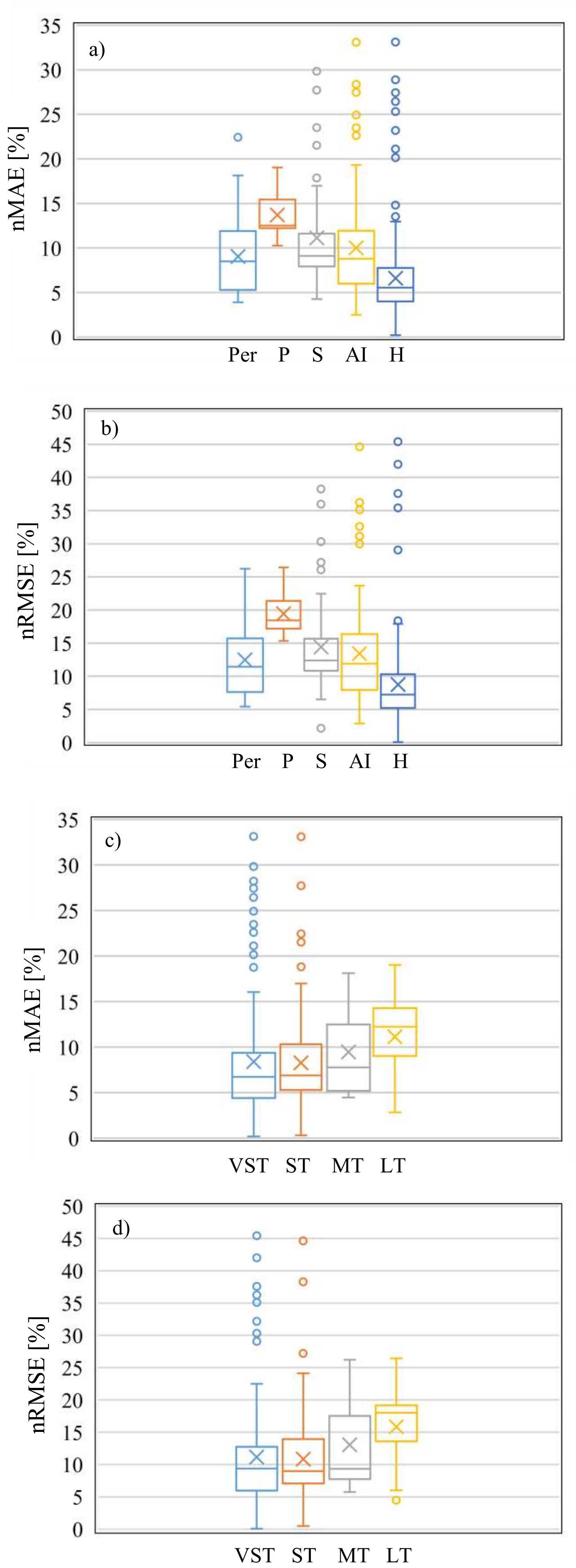


FIGURE 8. **Performance indicators (a) nMAE (b) nRMSE, for forecasting model, (c) nMAE (d) nRMSE, for time scale.**

statistical model the physical models are nearly 2.5 times, and statistical models are 1.8 times higher than the hybrid models, respectively. It shows that the best performance is of hybrid models. Although and hybrid model is almost the same, the median values of the physical models have a larger mean value of nRMSE ($\mu_{nRMSE} = 19.43\%$) but have small volatility ($\sigma_{nRMSE} = 3.29$). It shows that the model errors are systematic and not stochastic. Hence, post-processing the errors will further improve model accuracy. Out of all models, 10% outliers found for statistical models, 2.5% for AI/ML models and 4% for hybrid models. Mainly, the outliers are due to medium term and long term forecast. Similarly, from Fig. 8c and 8d, the whisker length shows that the accuracy decreases as the forecast horizon increases. In terms of nMAE median values, very short term forecast has a value of 6.73% that increases to 6.9% and 7.77% for short term and medium term forecast. Finally, the long term has the largest value of 12.24%. Based on previous analysis, here we again observe that the long term models have a larger mean value of nRMSE ($\mu_{nRMSE} = 15.86\%$) but small volatility ($\sigma_{nRMSE} = 5.74$). It again shows that the model errors are systematic and not stochastic. Thus, post-processing the errors will increase the model performance. In general, hybrid models outperform all the models.

TABLE 4. Performance of forecasting models according to time horizon and impact of input data on forecasting accuracy.

		Persistence	Physical	Statistical	Artificial Intelligence/ Machine Learning	Hybrid
Horizon	very short term	good	poor	good	good	good
	short term	average	average	good	good	good
	medium term	poor	good	average	average	good
	long term	poor	good	poor	poor	good
Input	NWP	not required	higher impact	lower impact	lower impact	higher impact in case of medium and long term forecast
	historical time series data	higher impact	lower impact	higher impact	higher impact	higher impact

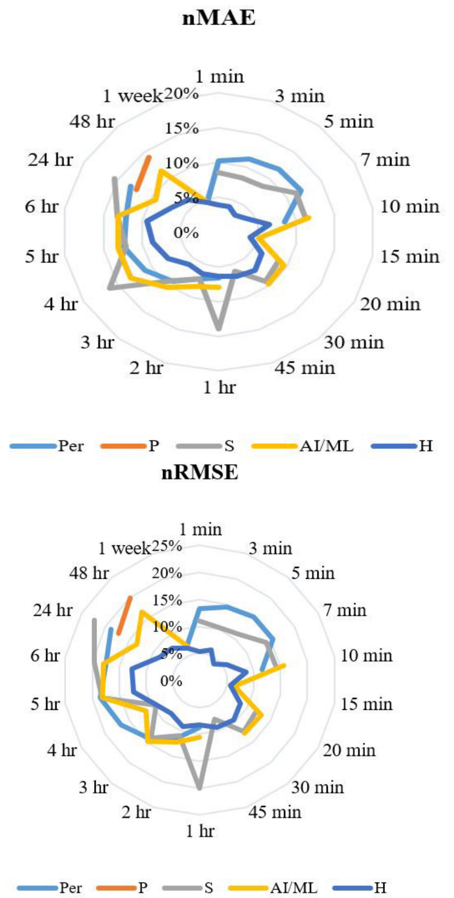


FIGURE 9. Statistical indicators with respect to the time horizon.

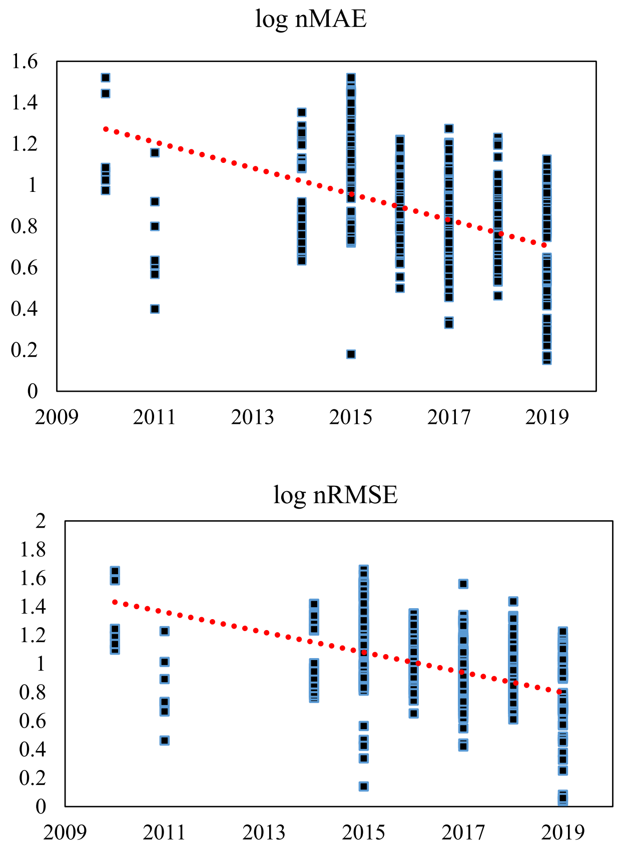


FIGURE 10. Linear trend of statistical errors with respect to publication year.

Fig. 9 indicates the dependence of statistical errors on the time horizon. Before interpreting the results, the reader may note that the number of entries is non-uniform in time horizon. The most significant subset is available for very short term forecasts. Also, the time steps are not equidistant. The available time horizons include 1 min, 3 min, 5 min, 7 min, 10 min, 15 min, 20 min, 30 min, 45 min, 1 hr, 2 hr, 3 hr, 4 hr, 5 hr, 6 hr, one day, two days and one week.

Persistence model performs the worst even in the very short term forecast. The performance of statistical models is excellent in the very short term such that it outperforms the AI/ML model (see 10 min). At 30 min, both of the models show almost the same results. From 1 hr onwards, the errors of statistical models start to increase, on average, by a factor 1.19. The data for physical models is only available for medium-term and long term forecast. The performance of the physical model outstrips persistence and statistical model, show the close result with AI/ML model and underperform the hybrid models. Overall, the hybrid models perform best for every time horizon. However, the errors almost doubled at the medium-term forecast from its initial value.

Among all discussions, it should be noted that model accuracy comparison between different techniques cannot be entirely justified until the same data set and the same level of effort is utilized. However, based on available studies, the performance of forecasting models according to time horizon and the impact of input data on forecasting accuracy is shown in Table 4. It is concluded that the best-suited models, among all, which work well for any time horizon and with any input data, are hybrid models. More specifically, the decomposition model with an optimization algorithm or error factor outperforms all the model when dealing with the short term forecast.

According to the analysis of Vargas *et al.* [59], wind power generation started growing substantially from the last decade. Therefore, it is a topic of interest to quantify the overall improvement of forecasting performance. Fig. 10 presents the linear trend of statistical errors with respect to publication year on a semi-logarithmic scale. The negative trend shows that the errors reduced every year. The slope (m) of regression line fitted to the $\log(nMAE)$ is $m = -0.069$ and for $\log(nRMSE)$ is $m = -0.0704$. It shows that the nMAE decreases with a factor of $\exp(-0.069) = 0.94$ every year. So, during the whole decade, nMAE drops to a factor of 0.533.

Similarly, the nRMSE drops with a factor of $\exp(-0.0704) = 0.93$ every year. So, during the whole decade, nRMSE drops to a factor of 0.49. It denotes that during the last ten years, the errors are reduced to about half the initial values.

IV. CONCLUSION

This paper reviewed the recent developments reported in the literature for deterministic wind speed and power forecasting. We classified and discussed the forecasting models as input data, time scale, power output, and forecasting methods. There is always a shortcoming of comparing different forecasting models. The studies available in the literature are not of identical datasets and authors have used data based on availability. Therefore, in each case study, the data is different in the forecasting horizon, model formation and location. The present study overcame this limitation and associated the performance and trend of recently developed forecasting models. Following conclusions are drawn from the detailed analysis:

- From descriptive statistics, the 95% CI for the mean, for averaged nMAE and nRMSE varies between 6.73% to 10.07% and 8.276% to 12.550% respectively. Similarly, the 95% CI for the median for nMAE and nRMSE varies between 6.14% to 9.41% and 7.548% to 12.085% respectively.
- In terms of median values, persistence is the same as AI/ML models. However, the spread of errors is more for the persistence model. Therefore, AI/ML models perform better than persistence, which is not expected from the illustrative plot.
- The most significant errors are witnessed for the physical models because these models are applied primarily for medium-term and long-term forecasts.
- The physical models have large nRMSE values but small volatility. Therefore, we conclude that the model errors are systematic and not stochastic. Hence, post-processing the errors will further improve model accuracy.
- Overall, the hybrid models perform best for every time horizon. However, the errors almost doubled at the medium-term forecast from its initial value.
- Based on the available dataset, the performance increased during the last ten years. On average, the errors are reduced to about half the initial values.

Improving the performance of wind forecasts is still a challenge for the researchers. A detailed study is required in this context to cover datasets from different climatic zones to analyse the performance of different forecasting models. The comprehensive review presented in this paper would help professionals and researchers to improve forecasting accuracy and to come up with more precise wind energy forecasts.

REFERENCES

[1] (2018). *Annual Global Wind Report*. Accessed: Sep. 3, 2019. [Online]. Available: https://gwec.net/wp-content/uploads/2019/04/GWEC-Global-Wind-Report-2018.pdf

[2] 67th ed. (2018). *BP Statistical Review of World Energy*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2018-full-report.pdf

[3] A. Fabbri, T. G. S. Roman, J. R. Abbad, and V. H. M. Quezada, "Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market," *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1440–1446, Aug. 2005.

[4] M. H. Albadi and E. F. El-Saadany, "Overview of wind power intermittency impacts on power systems," *Electr. Power Syst. Res.*, vol. 80, no. 6, pp. 627–632, 2010.

[5] S. S. Soman, H. Zareipour, O. Malik, and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," in *Proc. North Amer. Power Symp.*, Sep. 2010, pp. 1–8.

[6] 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, 2019.

[7] H. Liu, H.-Q. Tian, and Y.-F. Liu, "An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system," *J. Wind Eng. Ind. Aerodyn.*, vol. 141, pp. 27–38, Jun. 2015.

[8] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to Time Series Analysis and Forecasting*. Hoboken, NJ, USA: Wiley, 2008.

[9] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets: Operational Problems*. New York, NY, USA: Springer, 2013. Accessed: Sep. 3, 2019. [Online]. Available: https://link.springer.com/content/pdf/10.1007%2F978-1-4614-9411-9.pdf, doi: 10.1007/978-1-4614-9411-9.

[10] H. Liu, C. Chen, X. Lv, X. Wu, and M. Liu, "Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods," *Energy Convers. Manage.*, vol. 195, pp. 328–345, Sep. 2019.

[11] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, Apr. 2014.

[12] J. Yan, Y. Liu, S. Han, Y. Wang, and S. Feng, "Reviews on uncertainty analysis of wind power forecasting," *Renew. Sustain. Energy Rev.*, vol. 52, pp. 1322–1330, Dec. 2015.

[13] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh, "Current methods and advances in forecasting of wind power generation," *Renew. Energy*, vol. 37, no. 1, pp. 1–8, 2012.

[14] 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.

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

[16] 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.

[17] *Windfor*. Accessed: Sep. 3, 2019. [Online]. Available: https://enfor.dk/services/windfor/

[18] R. Blaga, A. Sabadus, N. Stefu, C. Dughir, M. Paulescu, and V. Badescu, "A current perspective on the accuracy of incoming solar energy forecasting," *Progr. Energy Combustion Sci.*, vol. 70, pp. 119–144, Jan. 2019.

[19] G. J. Haltiner, *Numerical Weather Prediction*. New York, NY, USA: Wiley, 1971.

[20] 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.

[21] J. Jung and R. P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renew. Sustain. Energy Rev.*, vol. 31, pp. 762–777, Mar. 2014.

[22] *ICOsahedral Nonhydrostatic*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.dwd.de/DE/forschung/wettervorhersage/num_modellierung/01_num_vorhersagemodelle/icon_beschreibung.html?nn=512942

[23] *Consortium for Small-Scale Modelling*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.dwd.de/DE/forschung/wettervorhersage/num_modellierung/01_num_vorhersagemodelle/regionalmodell_cosmo_de.html?nn=512942

[24] *Integrated Forecasting System—High RESolution (IFS-HRES)*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.ecmwf.int/en/forecasts/datasets/set-i

[25] *SEASonal Set V*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.ecmwf.int/en/forecasts/datasets/set-v

[26] *Deterministic Global & Deterministic UK*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.metoffice.gov.uk/research/modelling-?systems/unified-model/weather-forecasting

[27] *Global Spectral Model*. Accessed: Sep. 3, 2019. [Online]. Available: http://www.jma.go.jp/jma/en/NMHS/table/spec_GSM.pdf

[28] *Local Forecast Model*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.jma.go.jp/jma/jma-eng/jma-center/nwp/specifications/specifications_LFM.pdf

[29] A. Kusiak, H. Zheng, and Z. Song, "Short-term prediction of wind farm power: A data mining approach," *IEEE Trans. Energy Convers.*, vol. 24, no. 1, pp. 125–136, Mar. 2009.

[30] A. Kusiak and Z. Zheng, "Short-horizon prediction of wind power: A data-driven approach," *IEEE Trans. Energy Convers.*, vol. 25, no. 4, pp. 1112–1122, Dec. 2010.

[31] A. Kusiak, H. Zheng, and Z. Song, "Wind farm power prediction: A data-mining approach," *Wind Energy, Int. J. Progr. Appl. Wind Power Convers. Technol.*, vol. 12, no. 3, pp. 275–293, 2009.

[32] E. T. Renani, M. F. M. Elias, and N. A. Rahim, "Using data-driven approach for wind power prediction: A comparative study," *Energy Convers. Manage.*, vol. 118, pp. 193–203, Jun. 2016.

[33] X. Zhu and M. G. Genton, "Short-term wind speed forecasting for power system operations," *Int. Stat. Rev.*, vol. 80, no. 1, pp. 2–23, 2012.

[34] D. Y. Hong, T. Y. Ji, L. L. Zhang, M. S. Li, and Q. H. Wu, "An indirect short-term wind power forecast approach with multi-variable inputs," in *Proc. IEEE Innov. Smart Grid Technol.-Asia (ISGT-Asia)*, Nov./Dec. 2016, pp. 793–798.

[35] H. L. Wegley, M. R. Kosorok, and W. J. Formica, "Subhourly wind forecasting techniques for wind turbine operations," Pacific Northwest Lab., Richland, WA, USA, Tech. Rep. PNL-4894, 1984.

[36] Y.-K. Wu and J.-S. Hong, "A literature review of wind forecasting technology in the world," in *Proc. IEEE Lausanne Power Tech*, Jul. 2007, pp. 504–509.

[37] Y. Wang, Y. Liu, L. Li, D. Infield, and S. Han, "Short-term wind power forecasting based on clustering pre-calculated CFD method," *Energies*, vol. 11, no. 4, pp. 1–9, 2018.

[38] F. Castellani, D. Astolfi, M. Mana, M. Burlando, C. Meißner, and E. Piccioni, "Wind power forecasting techniques in complex terrain: ANN vs. ANN-CFD hybrid approach," *J. Phys., Conf. Ser.*, vol. 753, no. 8, 2016, Art. no. 082002.

[39] G. Kariniotakis, P. Pinson, N. Siebert, G. Giebel, and R. Barthelmie, "The state of the art in short-term prediction of wind power—From an offshore perspective," in *Proc. SeaTechWeek*, Brest, France, Oct. 2004.

[40] J. Wang, Q. Zhou, and X. Zhang, "Wind power forecasting based on time series ARMA model," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 199, no. 2, 2018, Art. no. 022015.

[41] S.-K. Sim, P. Maass, and P. G. Lind, "Wind speed modeling by nested ARIMA processes," *Energies*, vol. 12, no. 1, p. 69, 2019.

[42] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renew. Energy*, vol. 34, no. 5, pp. 1388–1393, 2009.

[43] J. L. T. García, E. C. Calderón, G. G. Ávalos, E. R. Heras, and A. M. Tshikala, "Forecast of daily output energy of wind turbine using sARIMA and nonlinear autoregressive models," *Adv. Mech. Eng.*, vol. 11, no. 2, 2019, Art. no. 1687814018813464.

[44] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. P. Kumar, "Linear and non-linear autoregressive models for short-term wind speed forecasting," *Energy Convers. Manage.*, vol. 112, pp. 115–124, Mar. 2016.

[45] N. Shi, S.-Q. Zhou, X.-H. Zhu, X.-W. Su, and X.-Y. Zhao, "Wind speed forecasting based on grey predictor and genetic neural network models," in *Proc. 2nd Int. Conf. Meas., Inf. Control*, vol. 2, Aug. 2013, pp. 1479–1482.

[46] T. Jónsson, P. Pinson, H. A. Nielsen, and H. Madsen, "Exponential smoothing approaches for prediction in real-time electricity markets," *Energies*, vol. 7, no. 6, pp. 3710–3732, Jun. 2014.

[47] *Simulation Model for the Operational Forecast of the Wind Energy Production in Europe*. Accessed: Sep. 3, 2019. [Online]. Available: http://www.eurowind.info/en/services/forecasts-and-actual-data/#c84

[48] Previento. Accessed: Sep. 3, 2019. [Online]. Available: https://www.energymeteo.com/products/power_forecasts/wind-solar-power-forecasts.php#Windsolar

[49] *Wind Power Management System*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.iee.fraunhofer.de/de/projekte/suche/laufende/wind_power_managementsystem.html

[50] *Det Norske Veritas—Germanischer Lloyd*. Accessed: Sep. 3, 2019. [Online]. Available: https://www.dnvgl.com/services/forecaster-introduction-3848

[51] EPREV. Accessed: Sep. 3, 2019. [Online]. Available: https://www.prewind.eu/services

[52] AleaWind. Accessed: Sep. 3, 2019. [Online]. Available: https://aleasoft.com/wind-energy-forecasting/

[53] Scirocco. Accessed: Sep. 3, 2019. [Online]. Available: http://www.windknowhow.com/upload/scirocco_brochure_english.pdf

[54] A. K. Yadav and H. Malik, "Short-term wind speed forecasting for power generation in Hamirpur, Himachal Pradesh, India, using artificial neural networks," in *Applications of Artificial Intelligence Techniques in Engineering*. Singapore: Springer, 2019, pp. 263–271.

[55] A. Khosravi, L. Machado, and R. O. Nunes, "Time-series prediction of wind speed using machine learning algorithms: A case study Osorio wind farm, Brazil," *Appl. Energy*, vol. 224, pp. 550–566, Aug. 2018.

[56] A. Khosravi, R. N. N. Koury, L. Machado, and J. J. G. Pabon, "Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system," *Sustain. Energy Technol. Assessments*, vol. 25, pp. 146–160, Feb. 2018.

[57] P. Lu, L. Ye, B. Sun, C. Zhang, Y. Zhao, and J. Teng, "A new hybrid prediction method of ultra-short-term wind power forecasting based on EEMD-PE and LSSVM optimized by the GSA," *Energies*, vol. 11, no. 4, p. 697, 2018.

[58] H. S. Dhiman, D. Deb, and J. M. Guerrero, "Hybrid machine intelligent SVR variants for wind forecasting and ramp events," *Renew. Sustain. Energy Rev.*, vol. 108, pp. 369–379, Jul. 2019.

[59] X. Kong, X. Liu, R. Shi, and K. Y. Lee, "Wind speed prediction using reduced support vector machines with feature selection," *Neurocomputing*, vol. 169, pp. 449–456, Dec. 2015.

[60] K. Wang, X. Qi, H. Liu, and J. Song, "Deep belief network based *k*-means cluster approach for short-term wind power forecasting," *Energy*, vol. 165, pp. 840–852, Dec. 2018.

[61] Y.-L. Hu and L. Chen, "A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm," *Energy Convers. Manage.*, vol. 173, pp. 123–142, Oct. 2018.

[62] D. Zhang, X. Peng, K. Pan, and Y. Liu, "A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine," *Energy Convers. Manage.*, vol. 180, pp. 338–357, Jan. 2019.

[63] X. Luo, J. Sun, L. Wang, W. Wang, W. Zhao, J. Wu, J.-H. Wang, and Z. Zhang, "Short-term wind speed forecasting via stacked extreme learning machine with generalized correntropy," *IEEE Trans. Ind. Informat.*, vol. 14, no. 11, pp. 4963–4971, Nov. 2018.

[64] W. Sun and M. Liu, "Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China," *Energy Convers. Manage.*, vol. 114, pp. 197–208, Apr. 2016.

[65] J. Wang, Y. Wang, and Y. Li, "A novel hybrid strategy using three-phase feature extraction and a weighted regularized extreme learning machine for multi-step ahead wind speed prediction," *Energies*, vol. 11, no. 2, p. 321, 2018.

[66] G. Huang, G.-B. Huang, S. Song, and K. You, "Trends in extreme learning machines: A review," *Neural Netw.*, vol. 61, pp. 32–48, Jan. 2015.

[67] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[68] A.-R. Mohamed, G. Dahl, and G. Hinton, "Deep belief networks for phone recognition," in *Proc. Nips Workshop Deep Learn. Speech Recognit. Rel. Appl.*, Vancouver, BC, Canada, 2009, vol. 1, no. 9, p. 39.

[69] M. Santhosh, C. Venkaiah, and D. M. V. Kumar, "Short-term wind speed forecasting approach using ensemble empirical mode decomposition and deep Boltzmann machine," *Sustain. Energy, Grids Netw.*, vol. 19, Sep. 2019, Art. no. 100242.

[70] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.

[71] D. K. Barrow and S. F. Crone, "A comparison of AdaBoost algorithms for time series forecast combination," *Int. J. Forecasting*, vol. 32, no. 4, pp. 1103–1119, 2016.

[72] 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.

[73] 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.

[74] I. Okumus and A. Dinler, "Current status of wind energy forecasting and a hybrid method for hourly predictions," *Energy Convers. Manage.*, vol. 123, pp. 362–371, Sep. 2016.

[75] 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.

[76] H. Liu, X. Mi, and Y. Li, "Comparison of two new intelligent wind speed forecasting approaches based on wavelet packet decomposition, complete ensemble empirical mode decomposition with adaptive noise and artificial neural networks," *Energy Convers. Manage.*, vol. 155, pp. 188–200, Jan. 2018.

[77] P. Jiang, H. Yang, and J. Heng, "A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting," *Appl. Energy*, vol. 235, pp. 786–801, Feb. 2019.

[78] J. Wang, P. Du, T. Niu, and W. Yang, "A novel hybrid system based on a new proposed algorithm—Multi-objective whale optimization algorithm for wind speed forecasting," *Appl. Energy*, vol. 208, pp. 344–360, Dec. 2017.

[79] J. Song, J. Wang, and H. Lu, "A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting," *Appl. Energy*, vol. 215, pp. 643–658, Apr. 2018.

[80] L. L. Zhang, M. S. Li, T. Y. Ji, and Q. H. Wu, "Short-term wind power prediction based on intrinsic time-scale decomposition and LS-SVM," in *Proc. IEEE Innov. Smart Grid Technol.-Asia (ISGT-Asia)*, Nov./Dec. 2016, pp. 41–45.

[81] W. Zhang, J. Wang, J. Wang, Z. Zhao, and M. Tian, "Short-term wind speed forecasting based on a hybrid model," *Appl. Soft Comput.*, vol. 13, no. 7, pp. 3225–3233, 2013.

[82] A. A. Abdoos, "A new intelligent method based on combination of VMD and ELM for short term wind power forecasting," *Neurocomputing*, vol. 203, pp. 111–120, Aug. 2016.

[83] C. Zhang, J. Zhou, C. Li, W. Fu, and T. Peng, "A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting," *Energy Convers. Manage.*, vol. 143, pp. 360–376, Jul. 2017.

[84] H. Liu, X.-W. Mi, and Y.-F. Li, "Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and elman neural network," *Energy Convers. Manage.*, vol. 156, pp. 498–514, Jan. 2018.

[85] L. Xiang, Z. Deng, and A. Hu, "Forecasting short-term wind speed based on IEWT-LSSVM model optimized by bird swarm algorithm," *IEEE Access*, vol. 7, pp. 59333–59345, 2019.

[86] X. Han, X. Zhang, F. Chen, Z. Song, and C. Wang, "Short-term wind speed prediction method based on time series combined with LS-SVM," in *Proc. 31st Chin. Control Conf.*, Jul. 2012, pp. 7593–7597.

[87] J. Zhang, Y. Wei, Z.-F. Tan, W. Ke, and W. Tian, "A hybrid method for short-term wind speed forecasting," *Sustainability*, vol. 9, no. 4, p. 596, 2017.

[88] S. Wang, N. Zhang, L. Wu, and Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method," *Renew. Energy*, vol. 94, pp. 629–636, Aug. 2016.

[89] A. Meng, J. Ge, H. Yin, and S. Chen, "Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm," *Energy Convers. Manage.*, vol. 114, pp. 75–88, Apr. 2016.

[90] H. Liu, H. Tian, X. Liang, and Y. Li, "New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, mind evolutionary algorithm and artificial neural networks," *Renew. Energy*, vol. 83, pp. 1066–1075, Nov. 2015.

[91] G. Osório, J. C. O. Matias, and J. Catalão, "Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information," *Renew. Energy*, vol. 75, pp. 301–307, Mar. 2015.

[92] J. A. Carta, P. Cabrera, J. M. Matías, and F. Castellano, "Comparison of feature selection methods using ANNs in MCP-wind speed methods. A case study," *Appl. Energy*, vol. 158, pp. 490–507, Nov. 2015.

[93] 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.

[94] A. Gensler, B. Sick, and S. Vogt, "A review of deterministic error scores and normalization techniques for power forecasting algorithms," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–9.

[95] S. A. Vargas, G. R. T. Esteves, P. M. Maçaira, B. Q. Bastos, F. L. C. Oliveira, and R. C. Souza, "Wind power generation: A review and a research agenda," *J. Cleaner Prod.*, vol. 218, pp. 850–870, May 2019.

[96] M. Mauri, T. Elli, G. Caviglia, G. Uboldi, and M. Azzi, "RAWGraphs: A visualisation platform to create open outputs," in *Proc. 12th Biannual Conf. Italian SIGCHI Chapter*, 2017, p. 28.

[97] E. Cadenas and W. Rivera, "Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA–ANN model," *Renew. Energy*, vol. 35, no. 12, pp. 2732–2738, 2010.

[98] R. Blonbou, "Very short-term wind power forecasting with neural networks and adaptive Bayesian learning," *Renew. Energy*, vol. 36, no. 3, pp. 1118–1124, 2011.

[99] X. An, D. Jiang, C. Liu, and M. Zhao, "Wind farm power prediction based on wavelet decomposition and chaotic time series," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11280–11285, 2011.

[100] J. Wang, W. Zhang, Y. Li, J. Wang, and Z. Dang, "Forecasting wind speed using empirical mode decomposition and Elman neural network," *Appl. Soft Comput.*, vol. 23, pp. 452–459, Oct. 2014.

[101] A. U. Haque, M. H. Nehrir, and P. Mandal, "A hybrid intelligent model for deterministic and quantile regression approach for probabilistic wind power forecasting," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1663–1672, Jul. 2014.

[102] P. Mandal, H. Zareipour, and W. D. Rosehart, "Forecasting aggregated wind power production of multiple wind farms using hybrid wavelet-PSO-NNs," *Int. J. Energy Res.*, vol. 38, no. 13, pp. 1654–1666, 2014.

[103] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind power forecasts using Gaussian processes and numerical weather prediction," *IEEE Trans. Power Syst.*, vol. 29, no. 2, pp. 656–665, Mar. 2014.

[104] D. Liu, J. Wang, and H. Wang, "Short-term wind speed forecasting based on spectral clustering and optimised echo state networks," *Renew. Energy*, vol. 78, pp. 599–608, Jun. 2015.

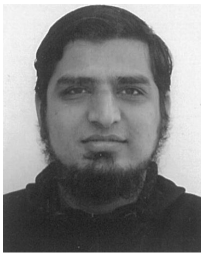
[105] H. Chitsaz, N. Amjady, and H. Zareipour, "Wind power forecast using wavelet neural network trained by improved clonal selection algorithm," *Energy Convers. Manage.*, vol. 89, pp. 588–598, Jan. 2015.

[106] 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.

[107] M. Ozkan and P. Karagoz, "A novel wind power forecast model: Statistical hybrid wind power forecast technique (SHWIP)," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 375–387, Apr. 2015.

[108] C. Zhang, H. Wei, J. Zhao, T. Liu, T. Zhu, and K. Zhang, "Short-term wind speed forecasting using empirical mode decomposition and feature selection," *Renew. Energy*, vol. 96, pp. 727–737, Oct. 2016.

[109] E. Cadenas, W. Rivera, R. Campos-Amezcua, and C. Heard, "Wind speed prediction using a univariate ARIMA model and a multivariate NARX model," *Energies*, vol. 9, no. 2, p. 109, 2016.
[110] Z. Liang, J. Liang, C. Wang, X. Dong, and X. Miao, "Short-term wind power combined forecasting based on error forecast correction," *Energy Convers. Manage.*, vol. 119, pp. 215–226, Jul. 2016.
[111] J. Liu, X. Wang, and Y. Lu, "A novel hybrid methodology for short-term wind power forecasting based on adaptive neuro-fuzzy inference system," *Renew. Energy*, vol. 103, pp. 620–629, Apr. 2017.
[112] V. Ranganayaki and S. N. Deepa, "SVM based neuro fuzzy model for short term wind power forecasting," *Nat. Acad. Sci. Lett.*, vol. 40, no. 2, pp. 131–134, 2017.
[113] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, "A data-driven multi-model methodology with deep feature selection for short-term wind forecasting," *Appl. Energy*, vol. 190, no. 15, pp. 1245–1257, Mar. 2017.
[114] 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. [Online]. Available: https://www.scopus.com/authid/detail.uri?authorId=57202388377&eid=2-s2.0-85048151311



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 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, University of 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 received the Ph.D. degree in robotics from Osaka University, Japan, in 2013. He was a Postdoctoral Researcher with the Toyota Technological Institute, Japan, in the field of parallel mechanism robots, from 2013 to 2014. Then, 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 Lecturer in mechatronics with Massey University.

CHAPTER 3

A MODIFIED GM (1, 1) MODEL TO ACCURATELY PREDICT WIND SPEED

This chapter contains content from the following article.

M. U. Yousuf, I. Al-Bahadly, and E. Avci, "A modified GM (1, 1) model to accurately predict wind speed," *Sustainable Energy Technologies and Assessments*, vol. 43, p. 100905, 2021.
DOI: https://doi.org/10.1016/j.seta.2020.100905

Contents lists available at ScienceDirect

Sustainable Energy Technologies and Assessments

journal homepage: www.elsevier.com/locate/seta

A modified GM(1,1) model to accurately predict wind speed

Muhammad Uzair Yousuf[a,b,*], Ibrahim Al-Bahadly[a], Ebubekir Avci[a]

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

A R T I C L E I N F O

Keywords:
Wind speed
Forecasting
Grey prediction model
Remnant
Developing coefficient

A B S T R A C T

Grey prediction models are suitable for uncertain systems and are recognized as a versatile wind energy forecasting technique. However, the traditional model has a disadvantage of seldom failure of necessary conditions. The first-order grey model with one variable [GM(1,1)] predicts the negative values of wind speeds, which is physically impossible. Also, forecasting results of the traditional model demonstrated that approximately 5% of the predicted values failed to achieve a predetermined accuracy level. In this study, a comprehensive modified GM(1,1) model is proposed considering wind speed data of Palmerston North, New Zealand. Remnant model and L' Hopital's rule are incorporated to overcome the issues of the traditional method. Results showed that the modified GM(1,1) model improved the forecasting validity of the traditional model by 98% while the individual accuracy level by 86%. Also, the forecasting performance of the new model is 9% higher than the traditional model. The robustness is further demonstrated by applying the model to three case studies. Overall, the modified model has excellent index of agreement with no negative wind speed predictions.

Introduction

Electricity generation from renewable sources has favorable economics relative to conventional thermal options. It is considered a significant option in the current supply and future energy policies of many countries. Currently, more than 80% of electricity in New Zealand generates from renewable sources (mainly hydro, geothermal, and wind) [1]. It is a strong directive of the National Policy Statement for Renewable Electricity Generation (NPS-REG) 2016, which states that the government is targeting 90% of electricity generation from renewables by 2025 [2]. New Zealand has a world-class wind resource. According to the New Zealand Wind Energy Association (NZWEA), the country has 17 operating wind farms. These wind farms currently have a combined installed capacity of 690 MW. The average annual output of these farms ranges from 100 kW at Southbridge to 161 MW at the Tararua wind farm. They supply around 6% of New Zealand's total electricity generation [3].

Despite an economically competitive and one of the fastest-growing technology, wind energy has a disadvantage of intermittent nature. The stochastic behavior of wind raised various challenges to the power systems. Such problems include, but are not limited to, planning and operational difficulties, unit commitment decisions, quality of power, the stability of power systems, and standard of inter-connections [4].

Forecasting is one of the possible solutions to overcome these challenges. Over the past few decades, numerous forecasting methods have been developed to predict wind speed accurately. These methods are divided into five categories: Persistence, Physical, Statistical, Artificial Intelligence/Machine Learning (AI/ML), and Hybrid. In the persistence model, the future wind speed is considered the same as the current time. Such a model is used for the very short-term forecast (0–30 min); however, the model cannot upgrade for further improvement.

Physical models are based on orography and require comprehensive information on the characteristics of a wind farm. These models predict medium-term (6 h-1 day ahead) and long-term wind forecast (>1 day ahead) accurately. Physical models generate global and regional predictions using initial conditions to solve complex numerical systems. However, physical models are computationally extensive to set up, require a long time to run, and are not suitable for very short-term and short-term forecasts (30 min–6 h) [5].

Statistical models are based on time series data and are therefore easy to build and fast to calculate. In contrast with physical models, statistical methods do not require extensive information to set up and are more appropriate for very-short-term and short-term forecasts. Commonly used statistical models are Auto-Regressive Integrated Moving Average (ARIMA) [6], Markov chain [7], and grey prediction models [8]. However, the forecasting performance is highly dependent on the time-series data. Also, these models rarely deal with nonlinear

* Corresponding author at: Department of Mechanical and Electrical Engineering, Massey University, Palmerston North 4442, New Zealand.
E-mail addresses: M.U.Yousuf@massey.ac.nz (M.U. Yousuf), I.H.AlBahadly@massey.ac.nz (I. Al-Bahadly), E.Avci@massey.ac.nz (E. Avci).

https://doi.org/10.1016/j.seta.2020.100905
Received 28 July 2020; Received in revised form 15 October 2020; Accepted 6 November 2020
Available online 23 November 2020

Abbreviations

1-AGO	First-order Accumulated Generating Operation
ANN	Artificial Neural Network
ARIMA	Auto-Regressive Integrated Moving Average
CTAGO	Cycle Truncation Accumulated Generating Operation
DGGM	Data Grouping Grey Model
ELM	Extreme Learning Machine
FAGM	Fractional Accumulated Grey Model
FANGBM	Nonlinear grey Bernoulli model with fractional order accumulation
FGM	Fractional Order Grey Model
FOTP	Full-Order Time Power
GA	Genetic Algorithm
GCA	Generalized regression neural network based on Cluster Analysis
GM	Grey Model
GM(1,1)	first-order grey model with one variable
GSA	Gravitational Search Algorithm
HEA	Hybrid Evolutionary Adaptive
HFGM	Heuristic Fuzzy time series Grey Model
HIM	Hybrid Intelligent Method
HP	Hodrick-Prescott
IoA	Index of Agreement
LSSVM	Least Square SVM
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
NRF	New Reference Forecasting
NWP	Numerical Weather Prediction
PCHIP	Piecewise Cubic Hermite Interpolating Polynomials
PFAGM	Power Driven Fractional Accumulated Grey Model
PSO	Particle Swarm Optimization
RGM	Rolling Grey Model
RMSE	Root Mean Square Error
SA	Simulated Annealing
SAIGM	Self-Adaptive Intelligence Grey Model
SFLA	Shuffled Frog Leaping Algorithm
SGM	Seasonal Grey Model
SVM	Support Vector Machine
WGMIPSO	Wavelet decomposed Grey Model improved by Particle Swarm Optimization
WOA	Whale Optimization Algorithm

behavior.

AI/ML techniques have stronger nonlinear estimation ability and are commonly used to predict wind speed. These include Artificial Neural Network (ANN) [9], Support Vector Machines (SVM) [10], and Fuzzy Logic (FL) [11]. AI/ML models can implicitly construct nonlinear and highly complex mapping relationships [5]. However, commonly used AI models have issues of low efficiency, overfitting, dimensionality issues, and premature convergence [12]. Other than traditional ML models, deep learning and extreme learning is gaining more consideration. These models showed higher accuracy and can learn more complex nonlinear relations [4]. Commonly used deep learning models include Deep Belief Network (DBN), Long Short-Term Memory (LSTM) [13], Echo State Network (ESN) [14], and Convolutional Neural Network (CNN) [15]. Although advanced learning methods improve the forecasting accuracy, however, the standard method tends to fall into local optima based on the predetermination of input hidden weights and hidden biases [16].

A single forecasting model has limited performance in multiple situations. Therefore, hybrid models are proposed as it utilizes the capabilities of the individual model. Hybrid models are further divided into four categories 1) preprocessing or decomposition method, 2) weighted method, 3) feature selection or optimization method, and 4) post-processing or error processing method. A preprocessing model decomposes nonstationary time series into stationary subseries, and then a separate prediction model is applied in each subseries. For the weighted method, each individual model is assigned a weight coefficient based on model effectiveness. Feature selection and optimization techniques are applied to remove redundant data. In post-processing techniques, a secondary model is applied to analyze the errors. Detailed reviews of these methods are available in the literature [4,17–19].

Grey prediction models

Grey prediction models are statistical methods that are widely recognized as a versatile forecasting model in research fields of medicine [20,21], engineering [22,23], and social sciences [24,25]. The grey prediction model is focused on the grey system established by Deng [26]. Compared with other statistical forecasting models, the grey prediction has an advantage that it only requires a small sample size, i.e., a minimum of four data points, to address the uncertainty of raw data comprehensively. Other than applying GM(1,1) model to practical applications, recent studies are also focusing on optimizing the model further. Besides combining GM(1,1) model with other techniques, Liu et al. [27] divided such research dimensions into eight different areas: (1) nature and characteristics of traditional model (2) selecting initial value (3) optimizing model parameters (4) recreating background value (5) optimizing model (6) improving simulation accuracy of discrete GM (1,1) model (7) modelling for non-equidistant sequence (8) application bound of models.

Integrating the rolling mechanism [28] with the traditional GM(1,1) model is also considered to enhance forecasting performance. A rolling

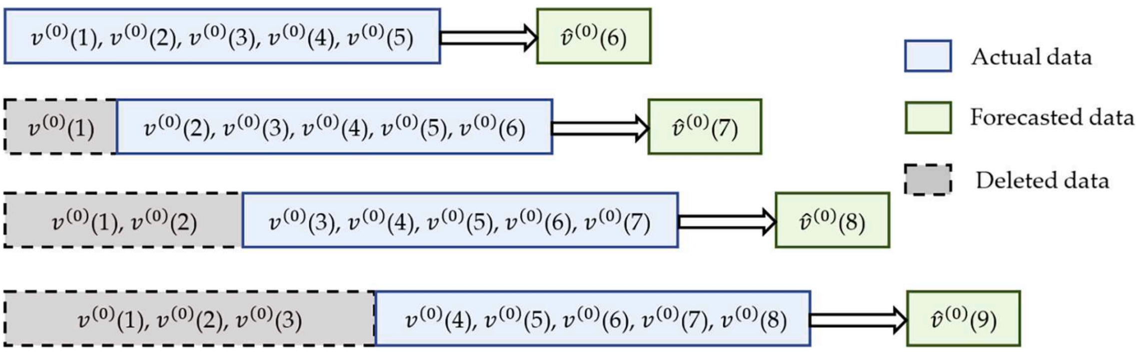


Fig. 1. The rolling mechanism for a window size of 5.

Table 1
Studies implementing GM(1,1) model for wind energy forecasting.

Reference	Time Horizon	Model Type	Proposed Model	Size of analyzed Dataset	Resolution of data	Predicted parameter	Period of forecasting	Superiority over	Modification
El Fouly et al. [8]	Short Term	Statistical	GM(1,1)	50	1 hr	Wind Speed Wind Power	One step ahead	Persistence	–
El Fouly et al. [45]	Short Term	Statistical	RGM(1,1) With adaptive alpha	50	1 hr	Wind Speed Wind Power	One step ahead	Persistence GM(1,1)	Adaptive alpha N = 4
	Short Term	Statistical	Averaged Grey Model	50	1 hr	Wind Speed Wind Power	One step ahead	Persistence GM(1,1)	Combination of traditional and Rolling model with adaptive alpha N = 4
An et al. [40]	Short Term	Hybrid	wavelet decomposition, chaotic time series, and GM (1, 1) model	470	10 min	Wind Power	One step ahead	Non-decomposed time series	GM(1,1) is used for the non-chaos part of decomposed time series N = 10
Chengxian and Shuqin [53]	Long Term	Statistical	revised parameter GM(1,1)	12	Monthly average	Wind Speed	4 step ahead	GM(1,1)	Weighting factor = $\frac{1}{a}-\frac{1}{e^a-1}$
Guo et al. [46]	Short Term	Hybrid	WGMIPSO	1451	5 min	Wind Speed	One step ahead	GM(1,1) GMIPSO	Particle Swarm Optimization is used to improve the grey model for every decomposed time series N = 5
Wu and Gao [38]	Medium Term	Hybrid	SVM-GM	168	1 hr	Wind Power	3 step ahead	Persistence NRF GM(1,1) ANN LSSVM-GSA HEA GCA HIM	GM(1,1) established with a combination of SVM and NWP optimized by PSO N = 10
Qolipour et al. [39]	Medium Term	Hybrid	Grey-ELM	1038	1 hr	Wind Speed	24 hr change	ELM NN-SFLA NN-GA NN-SA	GM (1,1) used to convert data into grey numbers and data reprocessed by ELM
Zhang et al. [30]	Short Term	Hybrid	PSO-SVR and Grey Combination Model	576	10 min	Wind Speed Wind Power	One step ahead	GM(1,1) ARIMA FGM (1,1)	$\alpha=\left\{0,\frac{3}{4},\frac{1}{2},\frac{1}{4},1\right\}$ Optimal N = 6
Zhang et al. [48]	–	Hybrid	WOA-PFAGM	12	Yearly average	energy consumption	3 years	GM(1,1) FAGM (1,1)	optimizes the grey input of the FAGM(1,1) with an exponential term of time
Wu et al. [49]	–	Hybrid	FANGBM	7	Yearly average	energy consumption	3 years	NGM FAGM(1,1) NGBM(1,1) ARIMA VM	Generalised model for GM(1,1), FAGM, the Verhulst model, and the NGBM(1,1) model
Qian and Wang [47]	–	Hybrid	improved seasonal model based on HP filter	28	3 monthly average	energy consumption	2 years	DGGM(1,1) TAGO-SGM (1,1)	Separately analyze the trend and seasonal fluctuation based on HP filter

mechanism considered only the latest data while deleting the older one, as shown in Fig. 1. Wu et al. [29] presented that the solution of the GM (1,1) model will be more stable if the small sample size (hereafter called window size *N*) is used. Zhang et al. [30] applied a window size from 4 to 8 for a case study of a wind farm in China and found out the optimal window size is 6. Similarly, Dejamkhooy et al. [31], Zhou et al. [32], and Sahin [33] selected the optimal window size of 5.

Applications of grey prediction models in the field of wind energy forecasting

Wind energy exhibits a stochastic nature. The current observation is more dependent on the latest values rather than the former one. Including more previous data decreases the forecasting performance. Therefore, grey forecasting is an effective method to predict wind speed data. The widely used grey prediction model is a first-order grey model with one variable – GM(1,1). Many researchers applied a grey forecasting model as an individual or hybrid model to predict wind speed, wind power, and wind energy consumption [30].

As an individual model, El-Fouly et al. [8] applied GM(1,1) model for a case study of a 600 kW wind turbine with 50 data points. The model showed improvements up to 12% of the persistence model. Yang et al. [34] employed GM(1,1) model to forecast wind power in China for nine years. The forecasting results depicted that the model predictions are consistent with the high GDP growth rate. A hybrid grey prediction model further improves the forecasting performance. The traditional model is combined with ARIMA [35], Markov [36], Nonlinear Auto-Regressive neural network (NARnet) [35], Elman NN [37], SVM [38], ELM [39], wavelets [40], Empirical Mode Decomposition (EMD) [41], and Ensemble Empirical Mode Decomposition (EEMD) [42]. Wu and Gao [38] considered GM(1,1) model combined with SVM for a wind farm case study in Denver, USA. The proposed model outperformed the

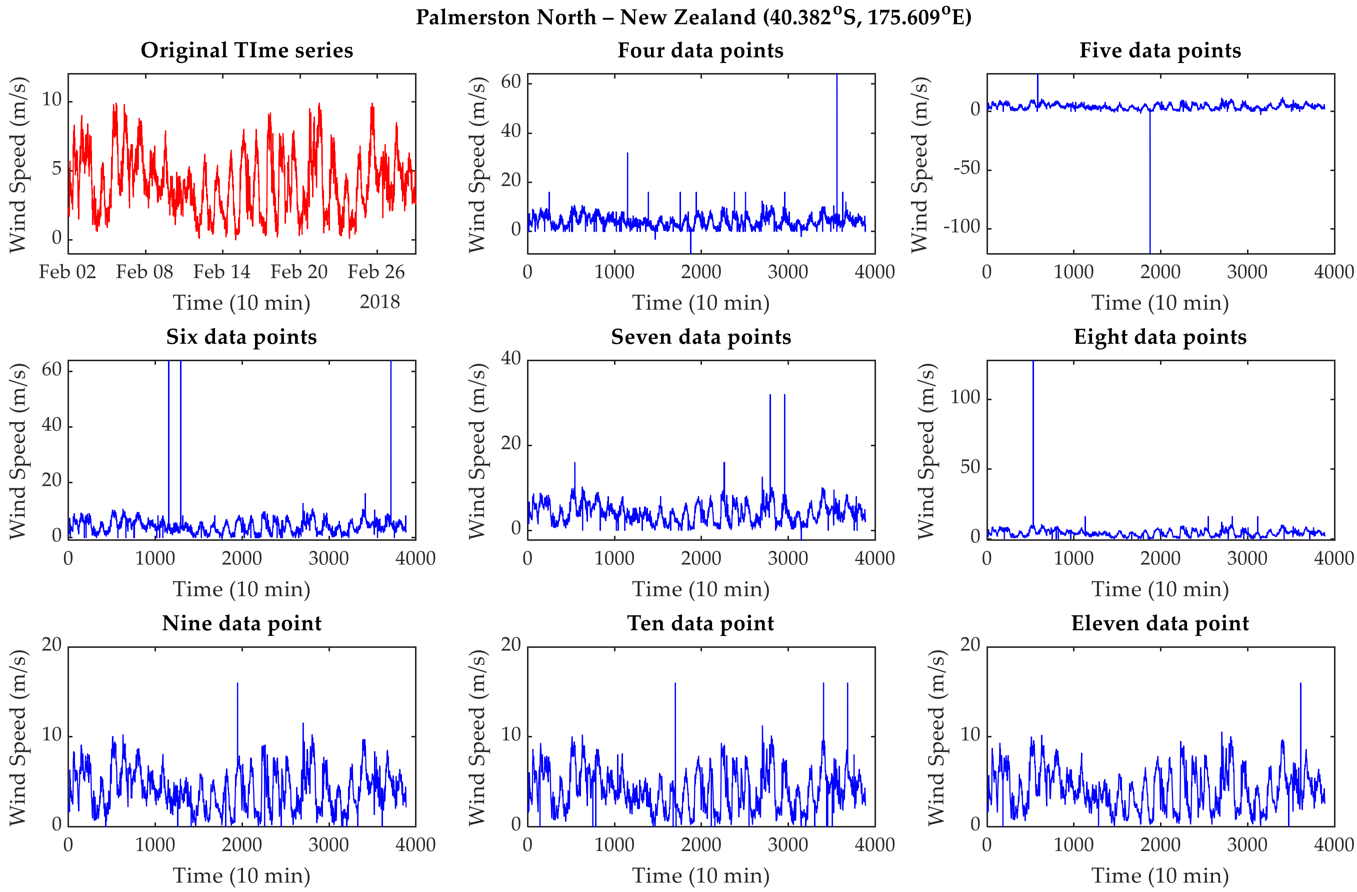


Fig. 2. One step ahead wind speed forecasting using traditional GM(1,1) model for window sizes varying from 4 to 11.

other considered eight models with a goodness of fit up to 98%. Zhang et al. [30] used fractional-order GM(1,1) combined with SVM optimized by PSO for a wind farm in China. The combined model is superior to other statistical models with an improvement in mean absolute error (MAE) up to 38%. Xia et al. [43] defined the new information priority accumulated GM(1,1) model with time power for the application of wind turbine capacity. The modified model improved the prediction accuracy in comparison with six other models. An analysis of a combined GM(1,1) and extreme learning machine model was carried out by Qolipour et al. [39] for wind speed data of Zanjan, Iran. The goodness of fit for the combined model is around 99%, which outperformed an individual ELM algorithm. In addition, other improved grey models such as FAGM(1,1), FOTP-GM(1,1), SAIGM(1,1), and non-homogeneous index sequence GM(1,1) [44] were also proposed for wind speed and wind power forecasting that obtained satisfactory prediction accuracy. The details of above-mentioned studies are given in Table 1. It is observed that variants of GM(1,1) model are more efficient than the traditional model. Also, if grey prediction model is combined with other advanced method, such as Grey-ELM [39], then the accuracy is further improved. Such hybrid models outperform AI/ML and statistical models.

Following the review and analysis of the above literature, three parameters are mostly analyzed to improve the predictive accuracy: initial value, window size (N), and weighting factor (α).

Initial value selection has very low significance for wind energy applications. Zhang et al. [30] considered the newer data point instead of the oldest one as the initial value. The results showed that the initial value does not influence wind speed prediction.

The selection of optimal window size is an important factor in improving the traditional model [30,38,40,45,46]. Considering all previous data might be useful for wind energy consumption where the time series shows an exponential trend [47–49]. However, the wind speed time series is nonstationary, and a smaller window size provides a more stable solution of the prediction model. From Table 1, the optimal window size varies from 4 to 10.

The most influential parameter to improve the prediction accuracy is background value and is mainly focused on literature. Background value is primarily affected by the weighting factor (α) [44]. A constant value of 0.5 is considered for the traditional model. To enhance the forecasting performance, researchers applied metaheuristic algorithms such as PSO [38,50] and GA [51] to optimize the background value. Cheng and Shi [52] optimized the background value by exponential, polynomial, and power function curve. Zhang et al. [30] reconstructed the background value in five different ways and suggested that the new point of background value provides the least prediction error.

Limitations of previous studies

Integrating a single optimal rolling window or improving the background value provides promising results for the considered case-studies; however, the problem of unrealistic prediction still exists.

We collected wind speed data for Palmerston North from the National Institute of Water and Atmospheric Research Ltd (NIWA) for the summer season, with a temporal resolution of 10 min [54]. A traditional GM(1,1) model is applied with a window size varying from four to eleven using MATLAB, and the results are shown in Fig. 2. It is visualized from the Fig. 2 that a traditional GM(1,1) predicts some erroneous values even with the smallest window size.

The weighting factor has two common approaches: i) constant weighting factor ii) adaptive weighting factor. We considered both the approaches and found out that none of the results are satisfactory

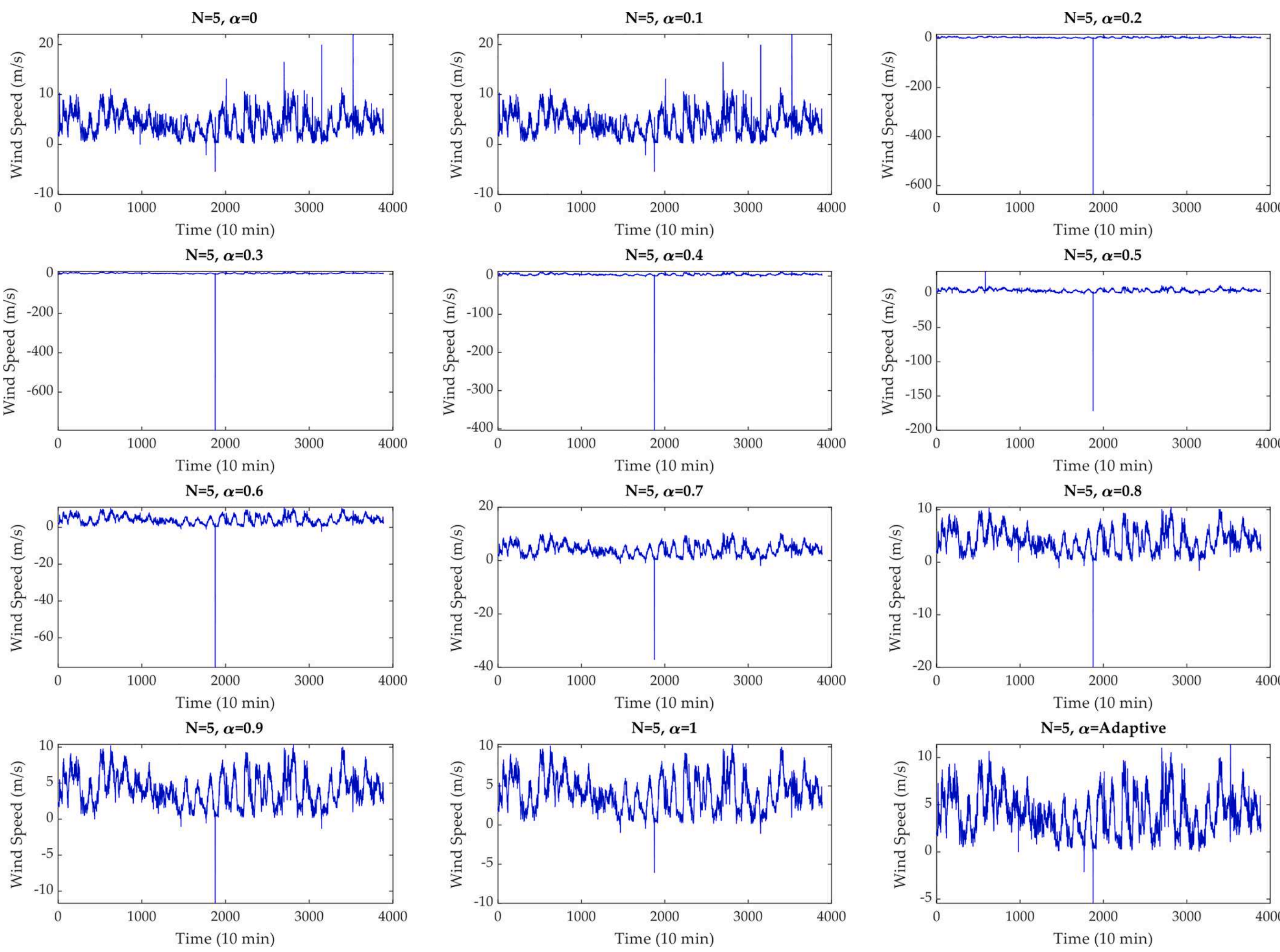


Fig. 3. One step ahead wind speed forecasting using traditional GM(1,1) model for varying weighting factor.

(Fig. 3). The problem of negative wind speed still exists, whereas some of the predicted values show substantial errors.

From the literature review, we observed that the necessary condition of developing coefficient a, is not considered in any of the studies, as mentioned earlier. In most of the high accuracy variants of the GM(1,1) model such as the FAGM, NGM, HFGM, DGGM, the common factor is $\left(v^{(0)}(1) - \frac{b}{a}\right)$ (See details of Eq. (8) in "*GM(1,1) model*" Section). When the developing coefficient equals zero, the factor approaches negative infinity. Similarly, if the developing coefficient is extremely small (in terms of 10^{-10}) then the factor gives a very high value. The model then predicts an unreasonable or sometimes impossible value. Also, if the developing coefficient is higher than the threshold value, the GM(1,1) model is not suitable for prediction. Furthermore, the adaptive weighting factor is considered in the literature, but no analysis has been carried out for variable rolling window size. Such problems have also been explored by other researchers [55,56]. However, those solutions partially satisfy the conditions, and a comprehensive model is still required for wind speed forecasting. To the best of authors' knowledge, such a model is not considered for wind speed time series in the literature, fulfilling all the necessary conditions. This is the primary motivation behind the work proposed in this paper. The three main contributions of this study are as follows.

- A modified GM(1,1) is proposed that integrates the traditional method with an optimal moving window and the adaptive weighting factor for very short-term wind speed forecasting. The effect of window size and weighting parameter is discussed to meet the accuracy levels.
- The necessary conditions of developing coefficient are considered in the proposed method that successfully solved previous models' major problems. The traditional model is modified by L' Hopital's rule and remnant model to fulfill the necessary and sufficient conditions.
- Datasets from four different cities are considered to confirm the validity of the proposed method. Results show that the modified GM (1,1) model is superior to the traditional GM(1,1) method.

The remainder of this paper is organized as follows: "*Materials and Methods*" Section presents the methodology of the modified grey prediction model. The problems of large prediction errors, singular phenomenon, and forecasting validity are also discussed. The experiments and analysis of the proposed forecasting system are presented in "*Results and Discussions*" Section. The robustness of the model is demonstrated by considering four case studies. Finally, the conclusions are presented in "*Conclusion*" Section.

Materials and methods

We first analyzed the raw data to find any missing or outlier values. Box-plot rule is used to identify anomalies, whereas Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) is applied to impute data inaccuracies. PCHIP is a three-degree piecewise polynomial function. Compared with cubic splines, it seeks to match only the first derivative of the data points with those of the intervals before and after. Also, it generally less undershoots or overshoots than cubic splines. MAPE is selected as the deciding parameter other than accuracy tests to find the optimal window size (N) for the rolling mechanism. MAPE is defined as Eq. (1).

$$MAPE = \sum_{k=1}^{n} \frac{\left| v^{(0)}(k) - \widehat{v}^{(0)}(k) \right|}{v^{(0)}(k)} \times 100\%, \tag{1}$$

where n is the number of observations.

GM(1,1) model

For a given non-negative time series with n observations i.e. $V^{(0)} = \{v^{(0)}(1), v^{(0)}(2), v^{(0)}(3), \cdots, v^{(0)}(n)\}$, the forecasting process through GM (1,1) is summarized in five steps as follows:

Step 1: First-order Accumulated Generating Operation (1-AGO)

The first step is to establish by $V^{(1)}$ using the first-order Accumulated Generating Operation (1-AGO) to the original time series as:

$$V^{(1)} = \{v^{(1)}(1), v^{(1)}(2), v^{(1)}(3), \cdots, v^{(1)}(n)\}, \tag{2}$$

$$v^{(1)}(k) = \sum_{i=1}^{k} v^{(0)}(i), k = 1, 2, 3, \cdots, n$$

Step 2: Background value array calculation

Same as $V^{(0)}$and $V^{(1)}$, the background value array $Z^{(1)}$ is defined as:

$$Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), z^{(1)}(3), \cdots, z^{(1)}(n)\}, \tag{3}$$

where $z^{(1)}(k) = \alpha v^{(1)}(k) + (1-\alpha)v^{(1)}(k-1)$, $k = 2, 3, \cdots, n$. α is a weighting factor and is usually specified as 0.5.

Step 3: Equation parameter calculation

The original form of the GM (1,1) model is a difference equation which expressed as:

$$v^{(0)}(k) + az^{(1)}(k) = b, \tag{4}$$

where a is the developing coefficient, and b is the grey actuating quantity. The parameter vector of the formula $\widehat{a} = [a, b]^T$ is estimated through the least square method, which satisfies:

$$\widehat{a} = (B^T B)^{-1} B^T Y, \tag{5}$$

where, $Y = \begin{bmatrix} v^{(0)}(2) \\ v^{(0)}(3) \\ \vdots \\ v^{(0)}(n) \end{bmatrix}$, $B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$ and $v^{(0)}(1) = v^{(1)}(1)$

Step 4: Whitenization (or image) equation and its solution

The following differential equation

$$\frac{dv^{(1)}}{dt} + av^{(1)} = b \tag{6}$$

is called whitenization (or image) equation of $v^{(0)}(k) + az^{(1)}(k) = b$ and its time response sequence is expressed as:

$$\widehat{v}^{(1)}(t) = \left(v^{(0)}(1) - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a} \tag{7}$$

Table 2
Developing coefficient and prediction length.

Developing Coefficient$\lvert a \rvert$	Prediction Length
$\lvert a\rvert \le 0.3$	Medium to long-term predictions
$0.3 < \lvert a\rvert \le 0.5$	Short to medium term predictions with less caution
$0.5 < \lvert a\rvert \le 0.8$	Short to medium term predictions with high caution
$0.8 < \lvert a\rvert \le 1$	Modified the traditional GM(1,1) model
$\lvert a\rvert > 1$	GM(1,1) is not suitable for predictions

The Eq. (7) is the solution of $V^{(1)}$.

Step 5: Inverse First-order Accumulated Generating Operation (Inverse 1-AGO)

The forecasting sequence $\widehat{V}^{(0)}$ can be obtained by the inverse 1-AGO as:

$$\widehat{V}^{(0)} = \begin{cases} \widehat{v}^{(0)}(1) = v^{(0)}(1) \\ \widehat{v}^{(0)}(k) = \widehat{v}^{(1)}(k) - \widehat{v}^{(1)}(k-1) = (1-e^a)\left(v^{(0)}(1) - \dfrac{b}{a}\right)e^{-a(k-1)} \end{cases}. \tag{8}$$

Remnant GM(1,1) model

The developing coefficient reflects the trend of $\widehat{v}^{(1)}(1)$ and $\widehat{v}^{(0)}(1)$. Increasing a rapidly increase the simulation errors, sometimes outside the predetermined accuracy requirement. Therefore, the condition of a must be analyzed first (as given in Table 2) before making any prediction [27].

One solution to modify the GM(1,1) model is to apply the additional grey model to the error sequence. This is termed as Remnant GM (1,1) model [27].

Let $\varepsilon^{(0)}$ is the error sequence with n observations such that $\varepsilon^{(0)} = \{\varepsilon(2), \varepsilon(3), \cdots, \varepsilon(n)\} = \left\{v^{(0)}(2) - \widehat{v}^{(0)}(2), v^{(0)}(3) - \widehat{v}^{(0)}(3), \cdots, v^{(0)}(n) - \widehat{v}^{(0)}(n)\right\}$. For absolute values of $\varepsilon^{(0)}$,

$$\widehat{\varepsilon}^{(0)}(k) = (1-e^{a_\varepsilon})\left(\varepsilon^{(0)}(2) - \frac{b_\varepsilon}{a_\varepsilon}\right)e^{-a_\varepsilon(k_\varepsilon-1)}, \tag{9}$$

where $k_\varepsilon = 3, 4\cdots, n$. Using Eqs. (8) and (9), the time response sequence is modified by adding or subtracting $\widehat{\varepsilon}^{(0)}(k)$from $\widehat{v}^{(0)}(k)$.

$$\widehat{V}^{(0)} = \begin{cases} \widehat{v}^{(0)}(1) = v^{(0)}(1) \\ \widehat{v}^{(0)}(k) = (1-e^a)\left(v^{(0)}(1) - \dfrac{b}{a}\right)e^{-a(k-1)} \pm (1-e^{a_\varepsilon})\left(\varepsilon^{(0)}(2) - \dfrac{b_\varepsilon}{a_\varepsilon}\right)e^{-a_\varepsilon(k_\varepsilon-1)} \end{cases}, \tag{10}$$

The Eq. (10) is referred to as the remnant GM(1,1) model, where the sign of the error modification value is the same as those in $\varepsilon^{(0)}$.

Singular phenomenon

Another potential problem of GM(1,1), discussed by Chen and Huang [55], is the singular phenomenon. We considered four different conditions of developing coefficient (a) in the previous section. One more possible case is $a = 0$. When developing coefficient is zero, the parameter b/a in Eq. (8) approaches infinity so that the factor $\left(v^{(0)}(1) - \frac{b}{a}\right)$ approaches negative infinity. In this situation, the indeterminate form results in the erroneous value of $\widehat{v}^{(0)}(k)$.

Generally, the researchers use some computing software for calcu-

Table 3
Levels of accuracy testing for GM(1,1).

Accuracy Level	Small error probability (p)	Post-error ratio (C)
I	$0.95 \le p$	$C \le 0.35$
II	$0.80 \le p < 0.95$	$0.35 < C \le 0.50$
III	$0.70 \le p < 0.80$	$0.50 < C \le 0.65$
IV	$0.60 \le p < 0.70$	$0.65 < C \le 0.80$

lations, where the floating-point truncation error causes the value of the developing coefficient extremely small (in terms of 10^{-10}) instead of zero. This results in predicting an unreasonable or sometimes impossible value by the model, as seen in Figs. 2 and 3. Therefore, the GM(1,1) model must be modified when $a \to 0$.

One possible solution for the singular phenomenon is to apply L' Hopital's rule, as discussed by [55]. From Eq. (8):

$$\lim_{a\to 0}\hat{v}^{(0)}(k) = \lim_{a\to 0}(1-e^{a})\left(v^{(0)}(1)-\frac{b}{a}\right)\left(e^{-a(k-1)}\right)$$

Rearranging the equation:

$$\lim_{a\to 0}\hat{v}^{(0)}(k) = \lim_{a\to 0}\frac{(1-e^{a})\left(e^{-a(k-1)}\right)}{\frac{1}{\left(v^{(0)}(1)-\frac{b}{a}\right)}}$$

$$\lim_{a\to 0}\hat{v}^{(0)}(k) = \lim_{a\to 0}\frac{1-e^{a}}{\left(e^{a(k-1)}\right)\Big/\left(v^{(0)}(1)-\frac{b}{a}\right)}$$

$$\lim_{a\to 0}\hat{v}^{(0)}(k) = \lim_{a\to 0}\frac{(1-e^{a})}{\left(ae^{a(k-1)}\right)/\left(av^{(0)}(1)-\text{b}\right)}$$

Differentiating the numerator and denominator, we get

$$\lim_{a\to 0}\hat{v}^{(0)}(k) = \lim_{a\to 0}\frac{-e^{a}}{\frac{\left(av^{(0)}(1)-b\right)\left\{a(k-1)e^{a(k-1)}+e^{a(k-1)}\right\}-\left(av^{(0)}(1)e^{a(k-1)}\right)}{\left(av^{(0)}(1)-b\right)^{2}}}$$

Applying limits, we get

$$= \frac{-1}{-b/b^{2}} = b.$$

Therefore, when $a = 0$, the modified GM(1,1) model is

$$\hat{V}^{(0)} = \begin{cases}\hat{v}^{(0)}(1) = v^{(0)}(1)\\ \hat{v}^{(0)}(k) = b\end{cases} \tag{11}$$

Accuracy testing of GM(1,1) model

Accuracy testing help to decide the appropriateness of a particular model. Each model chosen has to be tested through various methods. Once a model passed those tests, that can be meaningfully employed to make predictions. In this study, posterior error detection method

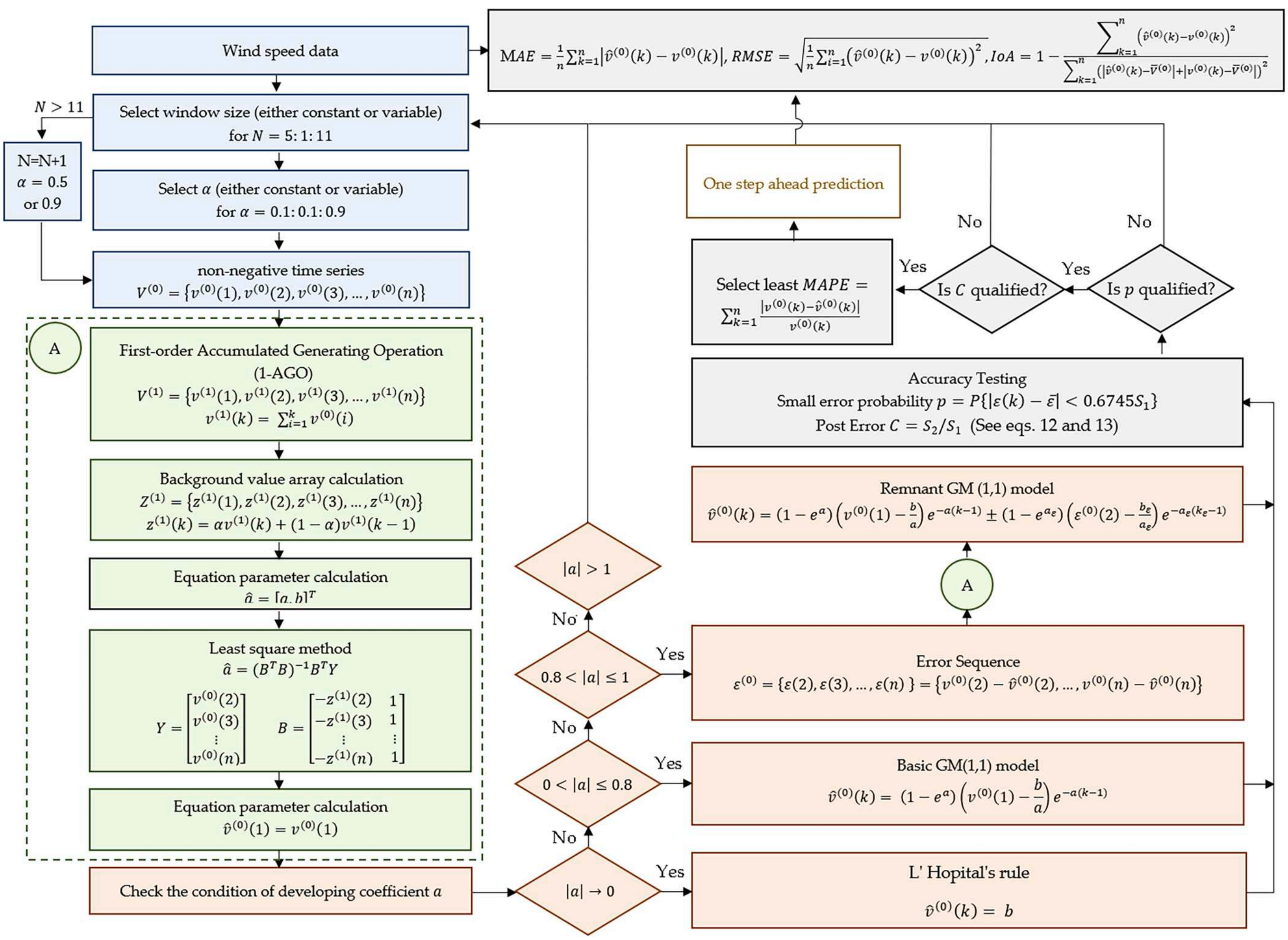


Fig. 4. The proposed modified GM(1,1) model.

Table 4
Summary of meteorological parameters at Palmerston North.

Mean monthly/annual wind speeds (m/s)											
Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
4.39	4.33	4.25	3.58	3.81	3.81	3.86	3.94	4.33	4.72	4.94	4.47
Average wind speed (m/s) for selected hours											
00:00	*03:00*	*06:00*	*09:00*	*12:00*	*15:00*	*18:00*	*21:00*				
3.42	3.28	3.28	4.28	5.44	5.72	4.61	3.67				
Average daily temperature range (°C)											
Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
10.3	10.3	10.2	9.9	8.9	8.2	8.3	8.5	8.5	8.6	9.1	9.4
Mean hourly temperatures for Jan. (°C)											
00:00	*01:00*	*02:00*	*03:00*	*04:00*	*05:00*	*06:00*	*07:00*	*08:00*	*09:00*	*10:00*	*11:00*
14.6	14.3	14	13.9	13.7	13.5	14.1	15.6	17.2	18.5	19.5	20.4
12:00	*13:00*	*14:00*	*15:00*	*16:00*	*17:00*	*18:00*	*19:00*	*20:00*	*21:00*	*22:00*	*23:00*
21.1	21.5	21.6	21.6	21.2	20.5	19.5	18.3	16.8	15.9	15.4	15
Mean hourly temperatures for Jul. (°C)											
00:00	*01:00*	*02:00*	*03:00*	*04:00*	*05:00*	*06:00*	*07:00*	*08:00*	*09:00*	*10:00*	*11:00*
7.1	7	6.9	6.8	6.7	6.6	6.5	6.6	6.7	7.8	9.4	10.6
12:00	*13:00*	*14:00*	*15:00*	*16:00*	*17:00*	*18:00*	*19:00*	*20:00*	*21:00*	*22:00*	*23:00*
11.3	11.7	11.9	11.7	11.3	10.3	9.1	8.5	8	7.7	7.5	7.3
Mean daily global solar radiation (MJ/m^2/day)											
Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
22.4	19.9	15.4	10.6	7	5.3	6.1	8.7	12.3	15.7	19.8	21.1
Mean monthly/annual 9 am vapor pressure (hPa)											
Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
14.9	15.3	14	12.7	11.4	9.7	9.3	9.7	10.7	11.6	12.2	14
Mean monthly/annual 9 am relative humidity (%)											
Jan	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*
75.3	77.7	79.4	81.2	85.8	86.8	86.8	84.6	79.7	80.5	76.7	76

criteria: small error probability (p) and post-error ratio (C) were used to test the accuracy of the forecasting models. The posterior error detection method also served as the forecasting validity of the grey model [27,54]. The prediction is only valid if the model passes both the accuracy tests.

Let S_1 and S_2 be the respective standard deviations of the original sequence $V^{(0)}$ and error sequence $\varepsilon^{(0)}$ and is calculated as in Eqs. (12) and (13), respectively.

$$S_1 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(v^{(0)}(k) - \overline{v}\right)^2} \quad \text{where} \quad \overline{v} = \frac{1}{n}\sum_{k=1}^{n} v^{(0)}(k), \tag{12}$$

$$S_2 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(\varepsilon(k) - \overline{\varepsilon}\right)^2} \quad \text{where} \quad \overline{\varepsilon} = \frac{1}{n}\sum_{k=1}^{n} \varepsilon(k) \tag{13}$$

The small error probability (p) and post-error ratio (C) is then evaluated as in Eqs. (14) and (15), respectively.

$$p = P\{|\varepsilon(k) - \overline{\varepsilon}| < 0.6745 S_1\}, \tag{14}$$

$$C = \frac{S_2}{S_1}. \tag{15}$$

A higher p value shows the ratio that errors fall within the acceptable range. The higher the p value, the superior the forecasting accuracy. The post-error ratio reflects the change rate of the difference between predicted and actual values. The lower the C value, the better the forecasting accuracy. Smaller C value shows that S_2 is relatively smaller than the S_1, i.e., errors are relatively more concentrated with little fluctuation than the original data. The acceptable levels of p and C are given in Table 3 with level I as the highest and level IV as the lowest. If no window size between 5 and 11 satisfies the accuracy conditions, then a constant weighting factor is used with the best possible window size. The overall methodology of the proposed modified GM(1,1) is outlined in Fig. 4. The final modified GM(1,1) model is:

$$\widehat{V}^{(0)} = \begin{cases} \widehat{v}^{(0)}(1) = v^{(0)}(1) & \\ \widehat{v}^{(0)}(k) = b & a = 0, k > 1 \\ \widehat{v}^{(0)}(k) = \widehat{v}^{(1)}(k) - \widehat{v}^{(1)}(k-1) = (1-e^{a})\left(v^{(0)}(1) - \dfrac{b}{a}\right)e^{-a(k-1)} & 0 < a \le 0.8, k > 1 \\ \widehat{v}^{(0)}(k) = (1-e^{a})\left(v^{(0)}(1) - \dfrac{b}{a}\right)e^{-a(k-1)} \pm (1-e^{a_\varepsilon})\left(\varepsilon^{(0)}(2) - \dfrac{b_\varepsilon}{a_\varepsilon}\right)e^{-a_\varepsilon(k_\varepsilon - 1)} & 0.8 < a \le 1, k > 1, k_\varepsilon > 2 \end{cases}$$

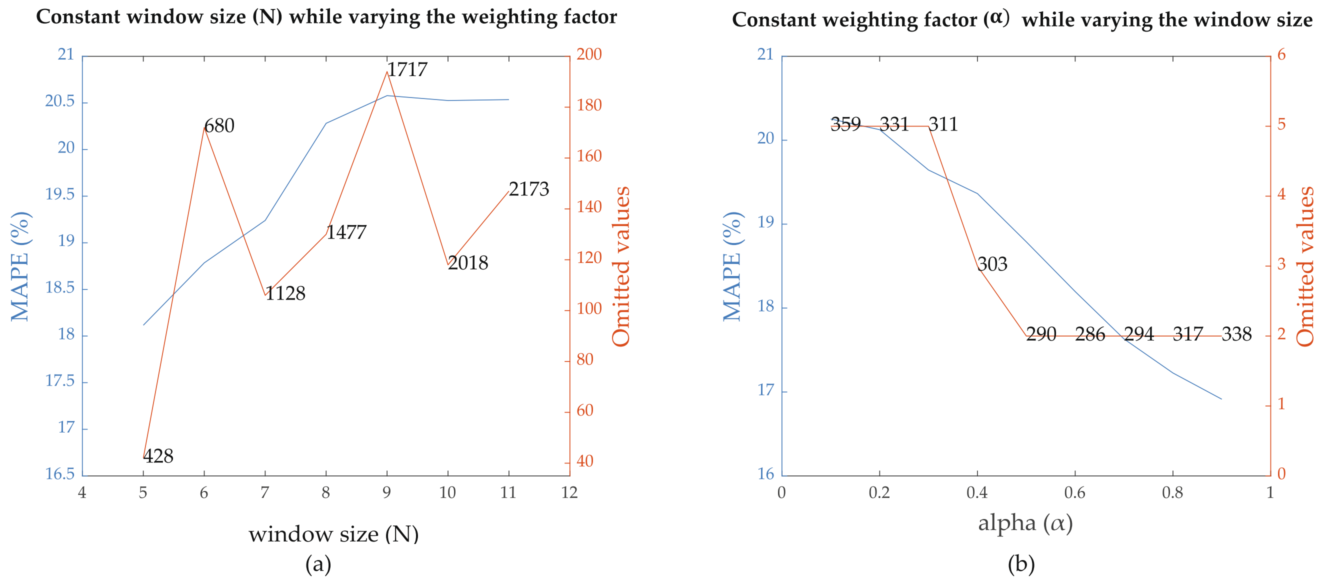


Fig. 5. Accuracy testing for: (a) N = constant, α = variable; (b) N = variable, α = constant. The numbers within the graph show the total data points that achieve the accuracy level I.

Results and discussions

Site specifications

Palmerston North (latitude: 40.382°S, longitude: 175.609°E, height: 21 m) is located in the Manawatu-Wanganui region and has a higher proportion of strong winds. According to the National Institute of Water and Atmospheric Research Ltd (NIWA) [57], the dominant wind direction is west-northwest. Spring is generally the windiest season. Among all recorded strong winds ($> 8.6m/s$), 37%, 26%, 19% and 18% occurred in spring, summer, winter, and autumn, respectively. For diurnal variation, the greatest wind speeds occur in the middle of the afternoon. Also, strong gust ($> 26m/s$) are infrequent at the studied location. The other details, including mean wind speed data (average speed taken over 10 min preceding each hour), daily temperature ranges, mean hourly temperature, mean daily global solar radiation, average 9 am vapor pressure, and relative humidity, is given in Table 4. For this study, wind speed data is collected from February 01, 2018 to February 28, 2018, with a time interval of 10 min at 10 m [54]. The finalized testing dataset comprised of 3,888 values, as shown in Fig. 2. The mean wind speed at the location is 4.1036 m/s, with a standard deviation of 2.1487 m/s.

One step ahead prediction based on traditional GM(1,1)

For a traditional GM(1,1) model, α is set as 0.5. It is seen from Fig. 2 that placing a weighting factor (α) value as 0.5 for a fixed window size (N) caused erroneous predictions when no optimization is considered. The reason for the incorrect predictions is the developing coefficient that approaches zero. Sometimes, erroneous predictions exceed physical limits. For example, $V^{(0)} = \{0.4, 0.3, 0.2, 0.1, 0.1, 0.7, 1.2\}$, predicts $\widehat{V}^{(0)} = -2.2893m/s$. In most of the cases, the singularity phenomenon occurs when two or more consecutive values are identical. For example, $V^{(0)} = \{8.8, 8.3, 9.2, 8.2, 7.5, 7.5, 8.8, 8.8\}$, predicts $\widehat{V}^{(0)} = 128m/s$. However, the consecutive number is not always the case. For example, $V^{(0)} = \{7.7, 8.1, 7.4, 7.3, 7.6, 8.3, 7.5\}$, predicts $\widehat{V}^{(0)} = 32m/s$. The permutation of time series that cause the singularity phenomenon is not the scope of the present study and will be considered a future direction.

A similar situation occurs with RGM (1,1) with adaptive alpha.

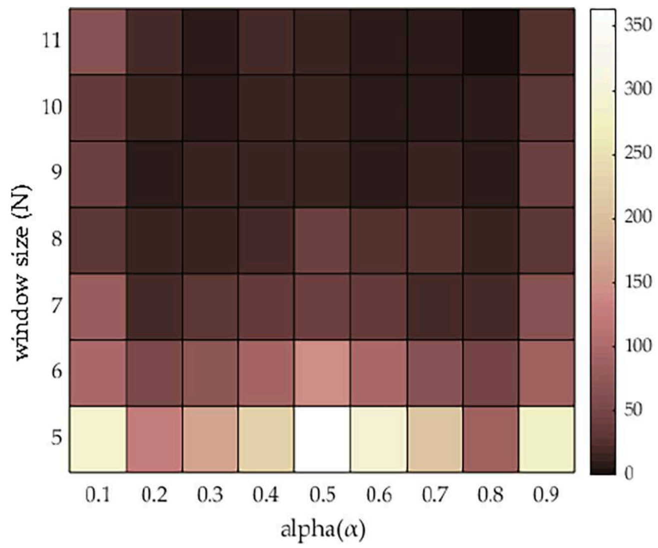


Fig. 6. Simulation results showing the combination of (N, α)to achieve the highest prediction accuracy level.

Although adjusting the alpha improves the forecasting performance of the traditional model, however, the problem of singular phenomena still exists, as shown in Fig. 3. Therefore, neglecting the necessary conditions of the developing coefficient limits the applicability of the GM(1,1) model.

One step ahead prediction based on the proposed methodology

The first choice is to select a constant window size (N) while varying the weighting factor ($0 < \alpha < 1$) with a step size of 0.1. The model behaves similarly to RGM(1,1) with adaptive alpha. We decided on the minimum value of N as 5. This is because the number of data points then available for the remnant GM(1,1) model is $(N-1) = 4$, which is the necessary condition to construct a grey model. Results show that the lower the window size, the higher is the prediction accuracy. Fig. 5(a) shows that the optimal window size is 5, with the lowest MAPE of

Table 5
Forecasting Validity, individual accuracy level I, and forecasting performance of five cases of the proposed model with improvement ratios.

Window Size(N)	Weighting Factor(α)	Predicted values failed to achieve a predetermined accuracy level	Improving Ratio (%)	Predicted values failed to achieve accuracy level I	Improving Ratio (%)	MAPE (%)	Improving Ratio (%)
5	0.5	88		803		18.56	
5	$0.1 \le \alpha \le 0.9$	42	52.27	428	46.70	18.12	2.37
$5 \le N \le 11$	0.5	2	97.73	290	63.89	18.79	−1.24
$5 \le N \le 11$	0.9	2	97.73	338	57.91	16.91	**8.89**
$5 \le N \le 11$	$0.1 \le \alpha \le 0.9$	1	98.86	115	**85.68**	18.56	0

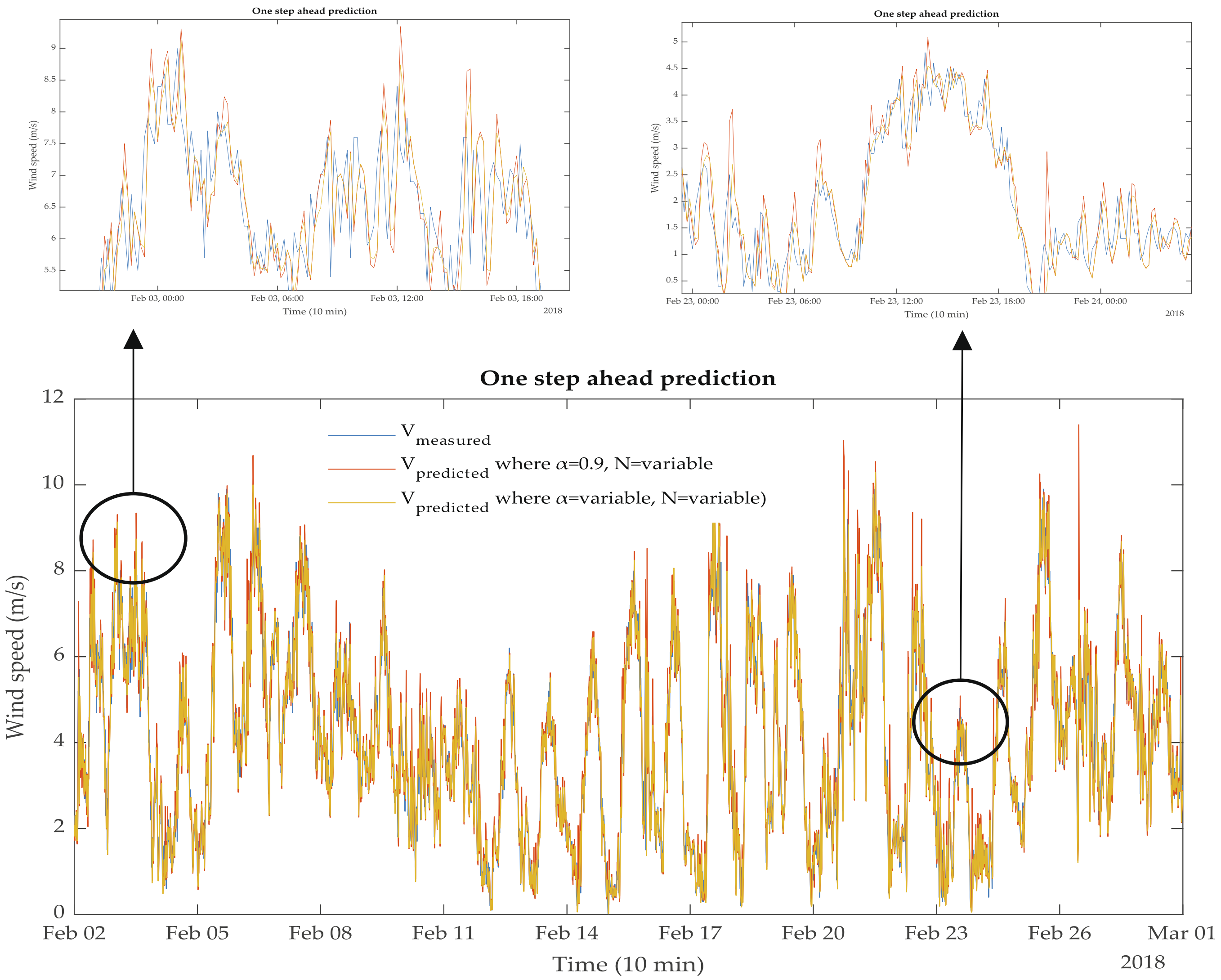


Fig. 7. One step ahead prediction for the proposed methodology.

18.12%. Also, we observed that some of the data points failed to attain the predetermined accuracy level. These omitted values are as low as 42 (1.08%) for $N = 5$ and as high as 194 (4.99% of total data) for $N = 9$. Increasing window size adversely affects the accuracy level of the predicted values as well. From the remaining predicted values that passed the predetermined accuracy level, 428 (11.12%) values failed to achieve small error probability accuracy level I when $N = 5$. Such values reached to 2173 (58.08%) when $N = 11$. Therefore, a fixed window size with varying weighting factor is not feasible for wind speed forecasting.

A constant weighting factor is selected for the second option while varying the window size from 5 to 11. The results show higher forecasting performance than the previous choice. From Fig. 5(b), we observe that the higher the α, the higher is the forecasting performance. Higher α shows that background value $z^{(1)}$ is based on the most recent value instead of the older one. Considering the MAPE of $\alpha = 0.5$ as a benchmark value, the accuracy for $\alpha = 0.9$ is improved by 9.98%. However, deviating α from the traditional value 0.5 increase the number of values that failed to achieve accuracy level I. A total of 290 (7.46%) values was unable to reach accuracy level I when $\alpha = 0.5$. The number of these values escalated to 338 (8.7%) when $\alpha = 0.9$, i.e., an increase of 16.62% from the traditional α.

Another possibility is to vary both α and N, and the optimal pair is selected based on accuracy levels. As per expectations from previous analysis, it is seen that most of the pairs comprise of $N = 5$ and $\alpha = 0.5$

Table 6
One step ahead wind speed prediction error for Palmerston North.

Case	Conditions	ME	MAE	MSE	RMSE	IoA
I	α = 0.9, N = variable	0.009684	0.489	0.4282	0.6544	97.67
II	α = variable, N = variable	−0.03483	0.544	0.5442	0.7377	97.07
% improvement	Case I over II	127.8	10.19	21.31	11.29	−0.61

(Fig. 6). The number of values failed to reach accuracy level I is 115, i.e., 65.97% lesser than the previous case ($\alpha = 0.9$). However, the forecasting performance is decreased as compared to the former case. MAPE is calculated as 18.56%, which was 16.91% in the second alternative.

The improving ratios of all the considered cases using the modified model are given in Table 5, where the bold values refer to the best model based on the individual error criterion. Improving ratio is defined as:

$$\text{Improving Ratio}(\%) = \frac{I_{GM} - I_{Model}}{I_{GM}} \times 100 \tag{16}$$

where I is a general index for comparison.

For calculations of MAPE, only those values are considered that reached the predetermined accuracy levels. From the analysis of the results stated in Table 5:

- A constant window size (N) with varying weighting factor (α) is better than the traditional model. Nearly 50% more data points passed the predetermined level and thus, improves the forecasting validity up to 50%. Also, the same number has achieved the accuracy level I. Overall, the model is 2.37% more accurate than the traditional model.
- The forecasting validity of varying window size is higher than the constant moving window. With varying window size, the forecasting validity is improved by 98% for constant alpha and 87% for adaptive alpha.
- The improving ratio based on individual accuracy levels of a constant α with adaptive N is higher by 64% than those of the traditional model. In comparison, varying both parameters increase this ratio to 86%. This shows that the adaptive weighting factor has an advantage over constant value. Also, MAPE of constant weighting factor is decreased by 1.24% while there is no effect on the latter case. Therefore, varying both window size (N) and weighting factor (α) provides the highest individual accuracy levels.
- The forecasting performance is mainly affected by the weighting factor. A higher value of the weighting factor provided the best performance. With varying window size, $\alpha = 0.9$ improved the traditional model by 8.89% and provided the best performance.

Based on the above analysis, adaptive window size with a 0.9 value of weighting factor is best suited to achieve the highest individual accuracy levels. Correspondingly, the optimal combination of both parameters provides the best forecasting performance.

Performance indicators

A comparison of one-step-ahead wind speed prediction for the latter two cases is given in Fig. 7. Other than MAPE, the overall forecasting performance of the last two cases of the prediction model is evaluated based on Mean Error (ME), Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (MAE), and Index of Agreement (IoA).

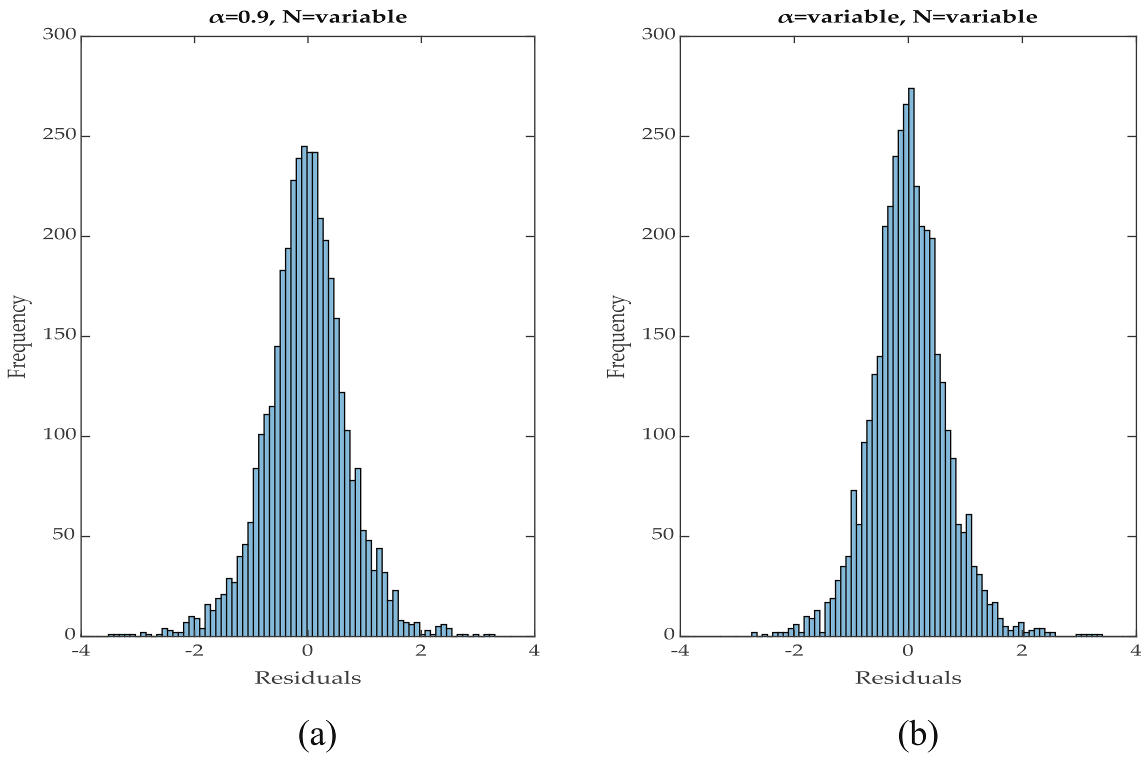


Fig. 8. Histograms of error for: (a) $\alpha = 0.9$, N = variable; (b) α = variable, N = variable.

Table 7
Average computational time per prediction for four models.

Model	Window Size	Weighting Factor	Computational time (sec)	Predict negative wind speed
Rolling GM (1,1) [8]	Constant	Constant	0.0149	Yes (Fig. 2)
Rolling GM (1,1) with adaptive alpha [30]	Constant	Variable	0.0926	Yes (Fig. 3)
Modified GM (1,1) (α = 0.9, N = variable)	Variable	Constant	0.0533	No (Fig. 7)
Modified GM (1,1) (α = variable, N = variable)	Variable	Variable	0.4104	No (Fig. 7)

Table 8
Site specifications of Mpika, Galicia and, Gwadar.

	Case Study I	Case Study II	Case Study III
Location	Mpika	Galicia	Gwadar
Latitude (°)	11.643°S	43.359°N	25.279 °N
Longitude (°)	31.341°E	7.879°W	62.346 °E
Climate	CWa	CSb	BWh
Start Date	February 01 2018	August 04 2018	November 03 2017
End Date	February 28 2018	August 31 2018	November 30 2017
Observing authority	ESMAP [58]	Sotavento [59]	ESMAP [58]
Total Observations	3888	3975	3932
Mean wind Speed (m/s)	4.3532	5.0884	3.8567
Standard deviation (m/s)	2.0313	1.9807	1.9877

B = dry, C = temperate, W = desert, S = steppe, a = hot summer, b = warm summer, h = hot.

$$ME = \frac{1}{n}\sum_{k=1}^{n}\left[\widehat{v}^{(0)}(k) - v^{(0)}(k)\right], \tag{17}$$

$$MAE = \frac{1}{n}\sum_{k=1}^{n}\left|\widehat{v}^{(0)}(k) - v^{(0)}(k)\right|, \tag{18}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{v}^{(0)}(k) - v^{(0)}(k)\right)^2, \tag{19}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{v}^{(0)}(k) - v^{(0)}(k)\right)^2}, \tag{20}$$

$$IoA = 1 - \frac{\sum_{k=1}^{n}\left(\widehat{v}^{(0)}(k) - v^{(0)}(k)\right)^2}{\sum_{k=1}^{n}\left(\left|\widehat{v}^{(0)}(k) - \overline{V}^{(0)}\right| + \left|v^{(0)}(k) - \overline{V}^{(0)}\right|\right)^2}, \tag{21}$$

where $\overline{V}^{(0)}$ is the mean of the given non-negative wind time series. Except for IoA, the lower values of parameters (closed to zero) show better prediction accuracy.

The results are presented in Table 6. On average, the forecasting performance of the constant weighting factor with the variable window size is at least 10% more than the latter case. Therefore, we conclude that the overall forecasting performance is lesser if we keep both α and N as a variable. IoA measures the degree of model prediction and is helpful for cross-comparison of models. Both models have a 97% index of agreement, which concludes that we can use any version of the modified GM(1,1) model for the wind speed prediction.

A precise understanding of the model accuracy is also determined from the residuals histogram, as shown in Fig. 8. It is observed that the frequency around zero is maximum for both cases. Thus, the shape of the curve endorses good forecasting performance models. Among both cases, a higher frequency around zero is witnessed in Fig. 8(b) that shows better individual accuracy levels for the second case. From visual inspection, the range of residuals is almost the same for both cases; however, the quantitative results presented in Table 6 show that the first case has a better overall forecasting performance.

Computational time

The average computational time per prediction for four models are given in Table 7. The models are programmed on MATLAB using Intel i5, 1.70 GHz processor with quad-core, and 16 GB RAM. For each model, the experiments were run ten times, and then the average value is considered. From Table 7, case II of the new model took the longest time, where the purpose is to achieve the highest individual accuracy levels. The RGM(1,1) took the lowest time.

In comparison with adaptive alpha RGM(1,1), case II of the present study is accurate and computationally efficient. Overall, the computational time is much lower than considered time resolution and is practically allowable. Also, for the RGM (1,1) with and without adaptive alpha, the problem of predicted negative wind speed exists, whereas the proposed model solved such problems.

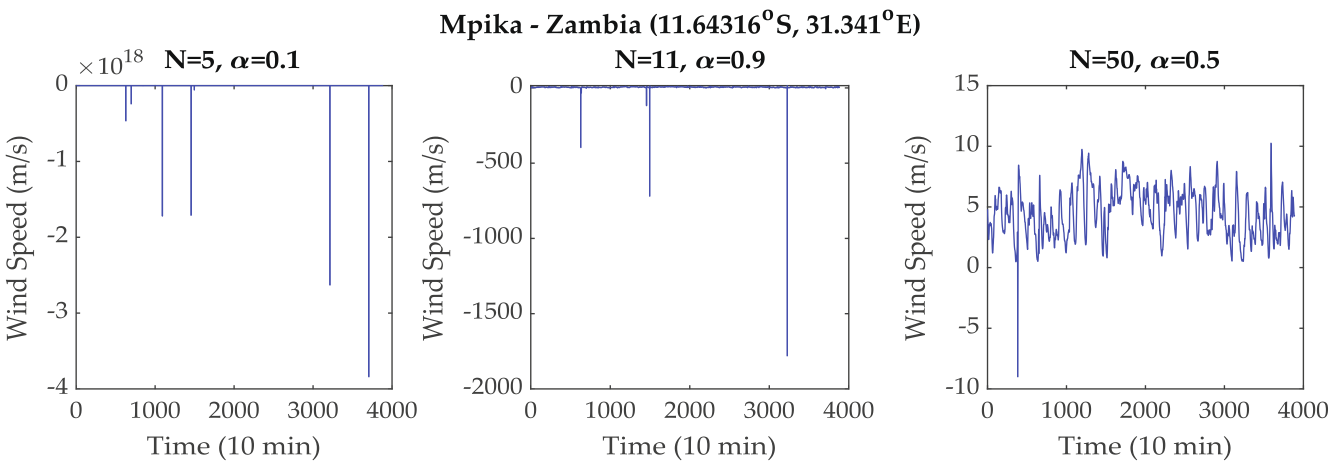


Fig. 9. One step ahead wind speed forecasting using traditional GM(1,1) model for (a) $N = 5, \alpha = 0.1$ (b) $N = 11, \alpha = 0.9$ (c)$N = 50, \alpha = 0.5$.

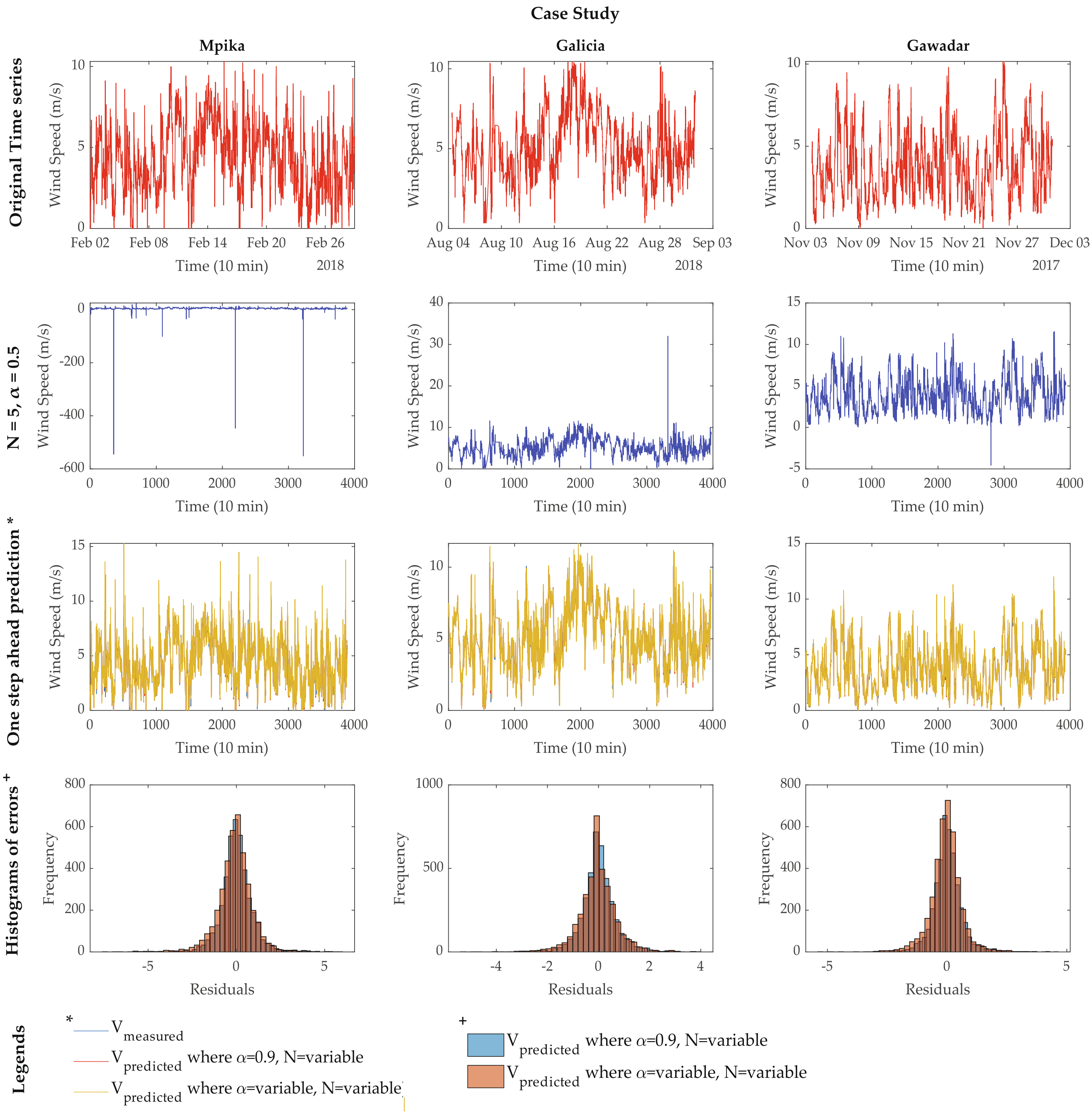


Fig. 10. One step ahead wind speed forecasting using the proposed model and analysis of residuals for Mpika, Galicia, and Gwadar.

Evaluation of robustness

Three independent case studies are further analyzed to verify the prediction performance of the new model to evaluate the robustness of the developed model. The details of the considered locations are given in Table 8.

The wind regime and energy production analysis at Mpika, Zambia, is a part of the Energy Sector Management Assistance Program (ESMAP) of the World bank. The site is located in an area of low-lying bush and rural agricultural land. The average wind speed is 6.2 m/s at 80 m [60]. Sotavento wind farm is situated in Galicia, Spain, with a nominal power of 17.56 MW. There are 24 turbines of nine models varying from 640 kW to 1.32 MW. The prevailing winds are on the east–west axis, with an average wind speed above 6 m/s [59]. Gwadar is in Pakistan and is one of the twelve sites for wind mapping in Pakistan managed by the World Bank and Alternative Energy Development Board. The mast is host by the Gwadar institute of technology. The land is flat with no obstruction. The mean wind speed is 4.7 m/s at 80 m [61].

Three different combinations of window size (5,11, and 50) and weighting factor (0.1, 0.5, and 0.9) for a case study of Mpika are shown in Fig. 9 to identify the shortcoming of the traditional model. Visual inspection demonstrates that the traditional model with constant window size is not suitable, even with an adaptive alpha. Also, increasing the window size is not solving the problem. The drawback of the traditional model to predict negative wind speed still exists with larger window size. The frequency of erroneous prediction for Mpika is much

Table 9
One step ahead wind speed prediction error for Mpika, Galicia, and Gwadar.

Case study	Conditions	ME	MAE	MSE	RMSE	IoA
Mpika	α = 0.9, N = variable	0.0273	0.6358	0.8324	0.9124	94.99
	α = variable, N = variable	−0.0938	0.7134	1.0639	1.0315	93.92
	Improvement of case I over II	129.12	10.87	21.76	11.54	−1.13
Galicia	α = 0.9, N = variable	0.0121	0.467	0.4557	0.675	97.11
	α = variable, N = variable	−0.033	0.5205	0.5657	0.7521	96.47
	Improvement of case I over II	136.71	10.28	19.44	10.24	−0.67
Gwadar	α = 0.9, N = variable	0.0176	0.4262	0.3744	0.6119	97.66
	α = variable, N = variable	−0.0469	0.4653	0.4554	0.6748	97.22
	Improvement of case I over II	137.53	8.41	17.88	9.33	−0.4492

higher than in Palmerston North. The same problem occurred for Galicia and Gwadar as well, as shown in Fig. 10.

In contrast, the modified model is superior to traditional GM(1,1) and solved all the issues mentioned earlier. Fig. 10 shows the one-step-ahead wind speed prediction by the modified model.

The comparison of several performance metrics for Mpika, Galicia, and Gwadar are given in Table 9. As our expectation, the forecasting performance of constant alpha with varying window size is at least 8%–10% higher than varying both the parameters. The index of the agreement is always above 90% that concludes we can use any of the models for the wind speed prediction.

Conclusions

The grey prediction has an advantage that it requires less information to address the uncertainty of raw data comprehensively. However, the traditional model may predict very high or even the negative values of wind speeds. Considering these issues, a modified GM(1,1) model is proposed and tested for four case studies. The generalized conclusions are as follows:

- The developing coefficient a is the most critical parameter. By using computational software, the floating-point truncation error causes the value of a extremely small. Therefore, it results in predicting negative wind speed. In this case, the traditional model is modified with L' Hopital's rule.
- Similarly, if $0.8 < |a| \leq 1$, it requires modification as the error tends to increase. For such a case, the traditional model is modified with a remnant GM(1,1) model.
- In this study, three cases are discussed for the modified model: constant window size while varying the weighting factor, constant weighting factor while varying window size, and finally, changing both window size and the weighting factor.
- A constant window size (N) with varying weighting factor (α) is not a suitable option for wind speed forecasting.
- If the objective is to increase overall forecasting performance, then vary window size (N) with a constant weighting factor ($\alpha = 0.9$).
- If the purpose is to achieve the highest individual accuracy levels, then vary both window size (N) and weighting factor (α).

The modified GM(1,1) model improved the forecasting performance of the traditional model by 9%. Therefore, integrating the modified model into already developed algorithms further improves the prediction accuracy of real-time applications. Also, integrating metaheuristic algorithms for background value optimization or analyzing model errors through other methods such as Markov Chains instead of the remnant model will further enhance forecasting performance.

Funding

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" 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.

CRediT authorship contribution statement

Muhammad Uzair Yousuf: Conceptualization, Methodology, Software, Validation, Formal, analysis, Investigation, Writing - original draft. **Ibrahim Al-Bahadly:** Resources, Supervision, Writing - review & editing. **Ebubekir Avci:** Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the National Institute of Water and Atmospheric Research Ltd (NIWA), Sotavento Galicia, S.A., World Bank, and Alternative Energy Development Board for their public reports, wind speed data, and kind cooperation.

References

[1] "Quarterly Electricity Generation and Consumption Data Updates." Ministry of Business, Innovation & Employment. https://www.mbie.govt.nz/building-and-energy/energy-and-natural-resources/energy-statistics-and-modelling/energy-statistics/electricity-statistics/ (accessed May 12, 2020).
[2] "Report of the Outcome Evaluation of the National Policy Statement for Renewable Electricity Generation," Ministry for the Environment, Wellington, 2016. Accessed: May 12, 2020. [Online]. Available: https://www.mfe.govt.nz/sites/default/files/media/RMA/npsreg-evaluation-report-final_0_0.pdf.
[3] "New Zealand's Wind Farms." New Zealand Wind Energy Association. http://www.windenergy.org.nz/wind-energy/nz-windfarms (accessed May 12, 2020).
[4] Yousuf MU, Al-Bahadly I, Avci E. Current perspective on the accuracy of deterministic wind speed and power forecasting. IEEE Access 2019;7:159547–64.
[5] Y. Zhang, Y. Li, and G. Zhang, "Short-term wind power forecasting approach based on Seq2Seq model using NWP data," Energy, p. 118371, 2020.
[6] Sim S-K, Maass P, Lind PG. Wind speed modeling by nested ARIMA processes. Energies 2019;12(1):69.
[7] Li W, Jia X, Li X, Wang Y, Lee J. A Markov model for short term wind speed prediction by integrating the wind acceleration information. Renew Energy 2020.
[8] El-Fouly T, El-Saadany E, Salama M. Grey predictor for wind energy conversion systems output power prediction. IEEE Trans Power Syst 2006;21(3):1450–2.
[9] Jahangir H, Golkar MA, Alhameli F, Mazouz A, Ahmadian A, Elkamel A. Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough ANN. Sustain Energy Technol Assess 2020;38:100601.
[10] Liu M, Cao Z, Zhang J, Wang L, Huang C, Luo X. Short-term wind speed forecasting based on the Jaya-SVM model. Int J Electr Power Energy Syst 2020;121:106056.
[11] Alamaniotis M, Karagiannis G. Application of fuzzy multiplexing of learning Gaussian processes for the interval forecasting of wind speed. IET Renew Power Gener 2019;14(1):100–9.

[12] Feng Z-K, Niu W-J, Zhang R, Wang S, Cheng C-T. Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization. J Hydrol 2019;576:229–38.
[13] Memarzadeh G, Keynia F. A new short-term wind speed forecasting method based on fine-tuned LSTM neural network and optimal input sets. Energy Convers Manage 2020;213:112824.
[14] Hu H, Wang L, Tao R. Wind speed forecasting based on variational mode decomposition and improved echo state network. Renewable Energy 2020.
[15] Hong Y-Y, Satriani TRA. Day-ahead spatiotemporal wind speed forecasting using robust design-based deep learning neural network. Energy 2020;209:118441.
[16] Feng Z-K, et al. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. J Hydrol 2020;583:124627.
[17] Liu H, Chen C. Data processing strategies in wind energy forecasting models and applications: a comprehensive review. Appl Energy 2019;249:392–408.
[18] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. Renew Sustain Energy Rev 2014;34: 243–54.
[19] Xiao L, Wang J, Dong Y, Wu J. Combined forecasting models for wind energy forecasting: a case study in China. Renew Sustain Energy Rev 2015;44:271–88.
[20] X. Yang, J. Zou, D. Kong, and G. Jiang, "The analysis of GM (1, 1) grey model to predict the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China," Medicine, vol. 97, no. 34, 2018.
[21] Wang Y-W, Shen Z-Z, Jiang Y. Comparison of ARIMA and GM (1, 1) models for prediction of hepatitis B in China. PLoS ONE 2018;13(9):e0201987.
[22] Yang X, Fang Z, Yang Y, Mba D, Li X. A novel multi-information fusion grey model and its application in wear trend prediction of wind turbines. Appl Math Model 2019;71:543–57.
[23] Zhang P, Ma X, She K. Forecasting Japan's Solar energy consumption using a novel incomplete gamma grey model. Sustainability 2019;11(21):5921.
[24] Rajesh R. Social and environmental risk management in resilient supply chains: a periodical study by the Grey-Verhulst model. Int J Prod Res 2019;57(11): 3748–65.
[25] Wang C-N, Le AP. Application of multi-criteria decision-making model and gm (1, 1) theory for evaluating efficiency of fdi on economic growth: a case study in developing countries. Sustainability 2019;11(8):2389.
[26] Ju-Long D. Control problems of grey systems. Syst Control Lett 1982;1(5):288–94.
[27] Liu S, Yang Y, Forrest J. Grey data analysis. Springer; 2017.
[28] Akay D, Atak M. Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. Energy 2007;32(9):1670–5.
[29] Wu L, Liu S, Yao L, Yan S. The effect of sample size on the grey system model. Appl Math Model 2013;37(9):6577–83.
[30] Zhang Y, Sun H, Guo Y. Wind power prediction based on PSO-SVR and grey combination model. IEEE Access 2019;7:136254–67.
[31] Dejamkhooy A, Dastfan A, Ahmadyfard A. Modeling and forecasting nonstationary voltage fluctuation based on grey system theory. IEEE Trans Power Deliv 2014;32 (3):1212–9.
[32] Zhou D, Al-Durra A, Zhang K, Ravey A, Gao F. A robust prognostic indicator for renewable energy technologies: A novel error correction grey prediction model. IEEE Trans Ind Electron 2019;66(12):9312–25.
[33] Şahin U. Forecasting of Turkey's greenhouse gas emissions using linear and nonlinear rolling metabolic grey model based on optimization. J Cleaner Prod 2019;239:118079.
[34] Yang J, Liu Q, Li X, Cui X. Overview of wind power in China: status and future. Sustainability 2017;9(8):1454.
[35] F. Marzbani, A. Osman, M. Hassan, and A. Noureldin, "Hybrid GM(1,1)-NARnet one hour ahead wind power prediction," in 2013 3rd International Conference on Electric Power and Energy Conversion Systems, 2-4 Oct. 2013 2013, pp. 1-6, doi: 10.1109/EPECS.2013.6713087.
[36] S. Chen L. Ye G. Zhang C. Zeng S. Dong C. Dai in 2011 International Conference on Advanced Power System Automation and Protection 2011 IEEE, pp. 1705–1711.
[37] Lin W-M, Hong C-M, Huang C-H, Ou T-C. Hybrid control of a wind induction generator based on Grey-Elman neural network. IEEE Trans Control Syst Technol 2013;21(6):2367–73.
[38] Wu D, Gao C. Short-term wind power generation forecasting based on the SVM-GM approach. Electr Power Compon Syst 2018;46(11–12):1250–64.
[39] Qolipour M, Mostafaeipour A, Saidi-Mehrabad M, Arabnia HR. Prediction of wind speed using a new grey-extreme learning machine hybrid algorithm: a case study. Energy Environ 2019;30(1):44–62.
[40] An X, Jiang D, Liu C, Zhao M. Wind farm power prediction based on wavelet decomposition and chaotic time series. Expert Syst Appl 2011;38(9):11280–5.
[41] Samet H, Marzbani F. Quantizing the deterministic nonlinearity in wind speed time series. Renew Sustain Energy Rev 2014;39:1143–54.
[42] H. Jiang, K. Zhang, J. Wang, and T. Yang, "A modified unbiased GM(1,1) prediction model based EEMD," in 2016 IEEE Chinese Guidance, Navigation and Control Conference (CGNCC), 12-14 Aug. 2016 2016, pp. 2027-2031, doi: 10.1109/CGNCC.2016.7829102.
[43] Xia J, Ma X, Wu W, Huang B, Li W. Application of a new information priority accumulated grey model with time power to predict short-term wind turbine capacity. J Cleaner Prod 2020;244:118573.
[44] Li S, Ma X, Yang C. A novel structure-adaptive intelligent grey forecasting model with full-order time power terms and its application. Comput Ind Eng 2018;120: 53–67.
[45] El-Fouly T, El-Saadany E, Salama M. Improved grey predictor rolling models for wind power prediction. IET Gener Transm Distrib 2007;1(6):928–37.
[46] M. Guo, Z. Wei, H. Zang, G. Sun, H. Li, and K. W. Cheung, "Short term wind speed forecasting using wavelet transform and grey model improved by particle swarm optimization," in 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT), 2015: IEEE, pp. 1879-1884.
[47] Qian W, Wang J. An improved seasonal GM (1, 1) model based on the HP filter for forecasting wind power generation in China. Energy 2020;209:118499.
[48] Zhang P, Ma X, She K. A novel power-driven fractional accumulated grey model and its application in forecasting wind energy consumption of China. PLoS ONE 2019;14(12):e0225362.
[49] Wu W, Ma X, Zeng B, Wang Y, Cai W. Forecasting short-term renewable energy consumption of China using a novel fractional nonlinear grey Bernoulli model. Renew Energy 2019;140:70–87.
[50] D. Wu, R. Xiao, C. Gao, Y. Shen, J. Wang, and Z. Zhao, "Short-term wind power data forecasting by particle swarm optimization dynamic GM (1, 1) model," in 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018: IEEE, pp. 395–399.
[51] B. Zeng, J. X. Zou, K. Li, and X. S. Xin, "Wind speed prediction based on improved grey neural network model," in Applied Mechanics and Materials, 2014, vol. 548: Trans Tech Publ, pp. 1235–1240.
[52] M. Cheng and G. Shi, "Improved methods for parameter estimation of gray model GM (1, 1) based on new background value optimization and model application," Communications in Statistics-Simulation and Computation, pp. 1-23, 2019.
[53] Chengxian F, Shuqin L. Wind speed forecasting method: gray related weighted combination with revised parameter. Energy Procedia 2011;5:550–4.
[54] Zhicun X, Meng D, Lifeng W. Evaluating the effect of sample length on forecasting validity of FGM (1, 1). Alexandria Eng J 2020.
[55] Chen C-I, Huang S-J. The necessary and sufficient condition for GM (1, 1) grey prediction model. Appl Math Comput 2013;219(11):6152–62.
[56] Zhou W, He J-M. Generalized GM (1, 1) model and its application in forecasting of fuel production. Appl Math Model 2013;37(9):6234–43.
[57] Chappell PR. The climate and weather of Manawatu-Wanganui. Taihoro Nukurangi: NIWA; 2015.
[58] Wind Data obtained from World Bank via ENERGYDATA.info, under a project funded by the Energy Sector Management Assistance Program (ESMAP). (accessed May 25, 2019).
[59] "Historical-Real Time data " Sotavento. http://www.sotaventogalicia.com/en/technical-area/real-time-data/historical/ (accessed May 25, 2019).
[60] "Wind Resource Mapping in Zambia 12 Month Site Resource Report," May 08 2018, issue C. [Online]. Available: http://documents1.worldbank.org/curated/en/528711526549758961/pdf/Renewable-energy-wind-mapping-for-Zambia-12-month-site-resource-report.pdf.
[61] "Wind Resource Mapping in Pakistan Site Installation Report: All Masts," Feb. 2017. [Online]. Available: http://documents1.worldbank.org/curated/en/839661487934625743/pdf/112987-ESM-P146140-PUBLIC-PakistanWindMappingSiteInstallationReportsWBESMAPFeb.pdf.

CHAPTER 4

SHORT-TERM WIND SPEED FORECASTING BASED ON HYBRID MODWT-ARIMA-MARKOV MODEL

This chapter contains content from the following article.

M. U. Yousuf, I. Al-Bahadly, and E. Avci, " Short-term wind speed forecasting based on hybrid MODWT-ARIMA-Markov model," *IEEE Access*, vol. 9, pp. 79695-79711, 2021.
DOI: https://doi.org/10.1109/ACCESS.2021.3084536

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.

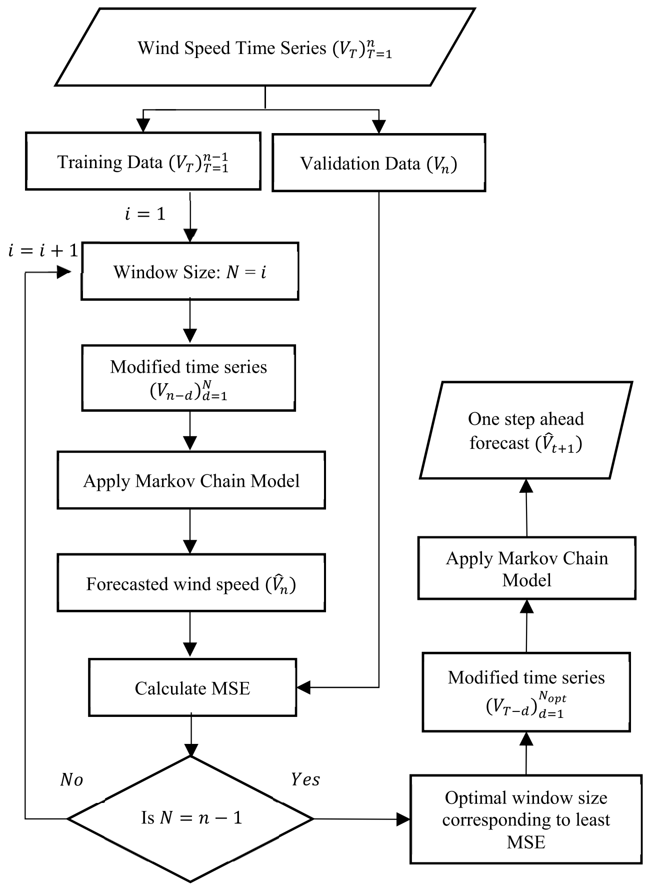


FIGURE 1. Selection of optimal window based on a similar approach to the leave-one-out method.

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,

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

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 $p_{ii}^{(n)} < 0$.

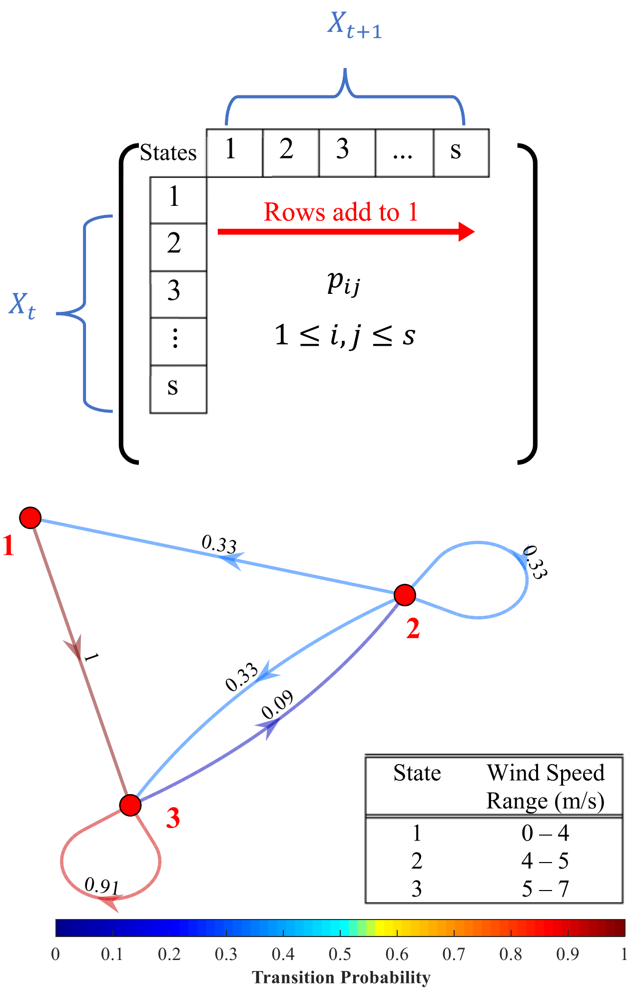


FIGURE 2. Properties of a transition probability matrix with a transition graph of sequence discussed in the state categorization step.

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]:

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

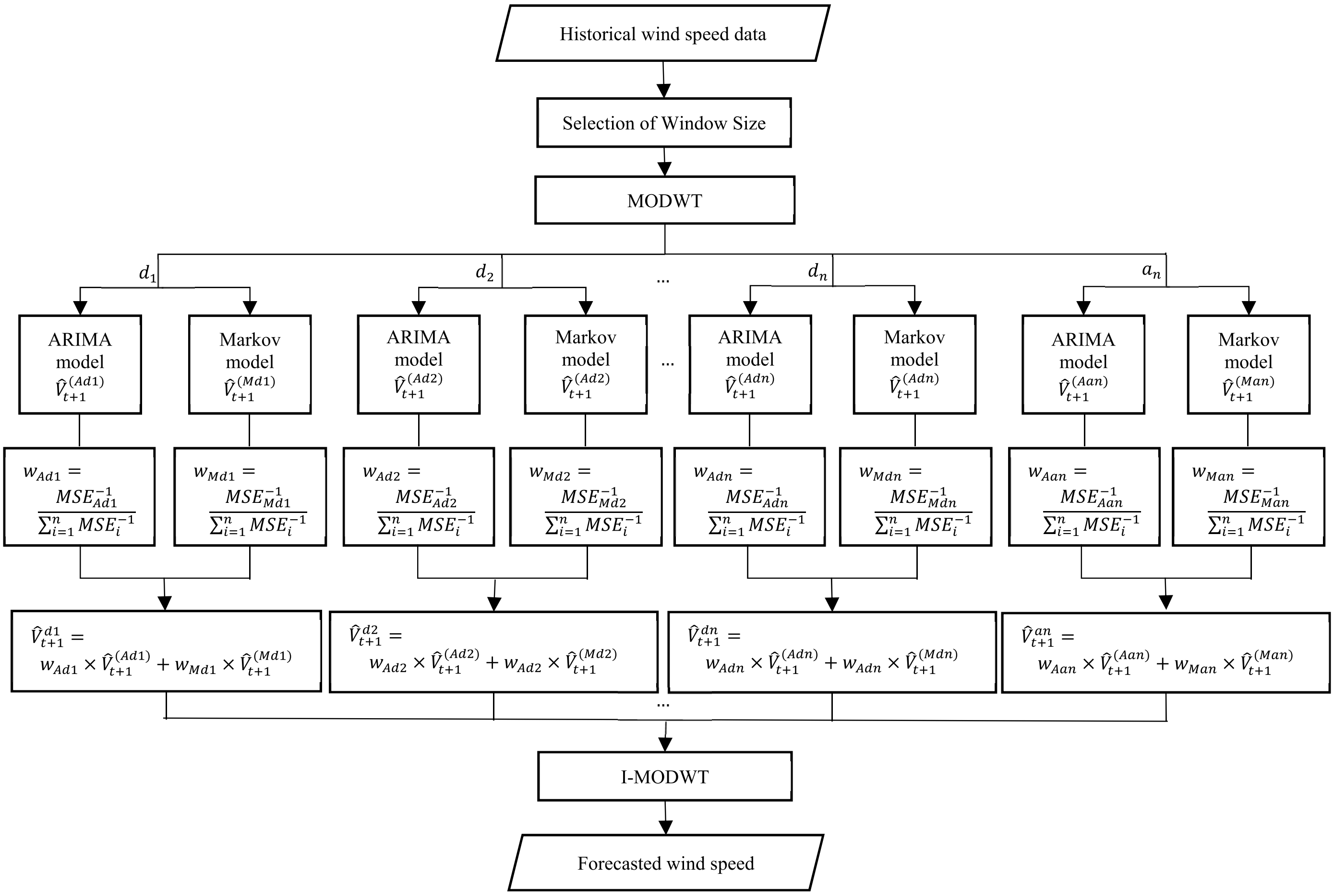


FIGURE 3. **The flowchart of the proposed hybrid MODWT-ARIMA-Markov model.**

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).

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

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

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

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

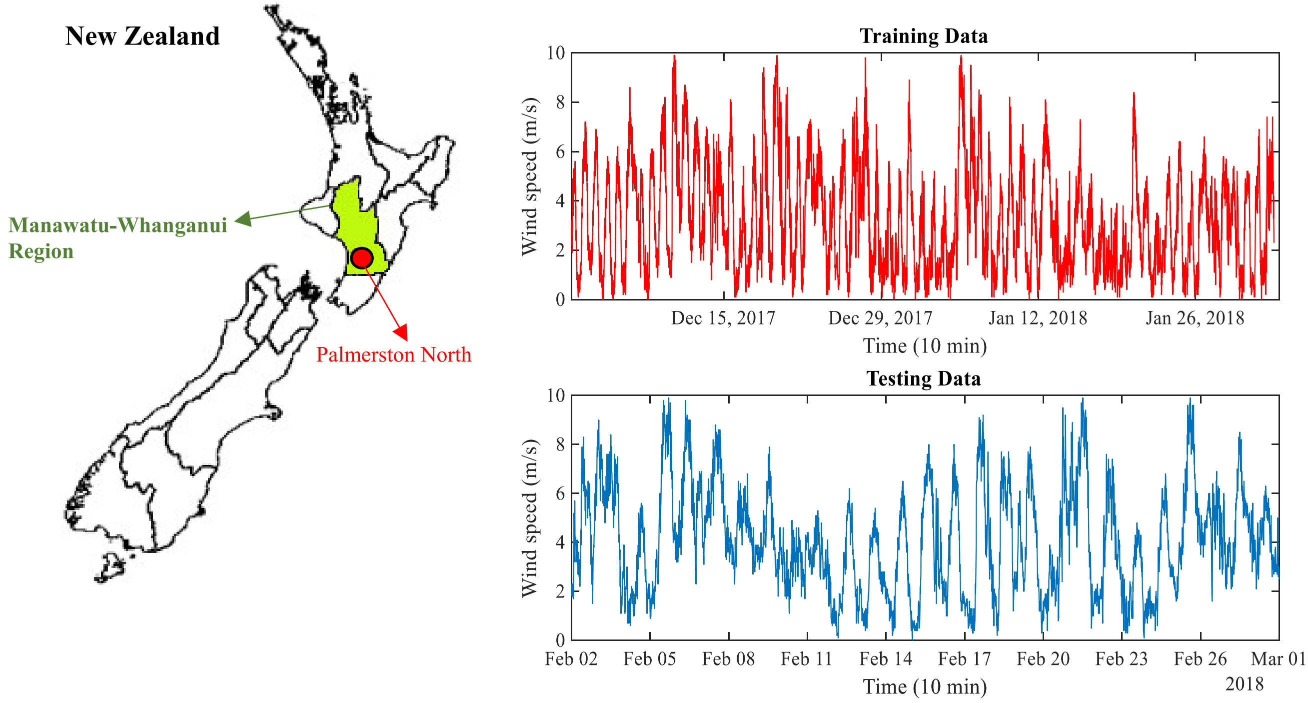
Fig. 5(a) shows one step ahead of wind speed forecasting with the testing dataset. The prediction results show excellent agreement with the measured data with $R = 0.94$. Fig. 5(b) shows that 81.3% of the data requires a maximum of a prior week's observation. Also, more than half of the values require 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.

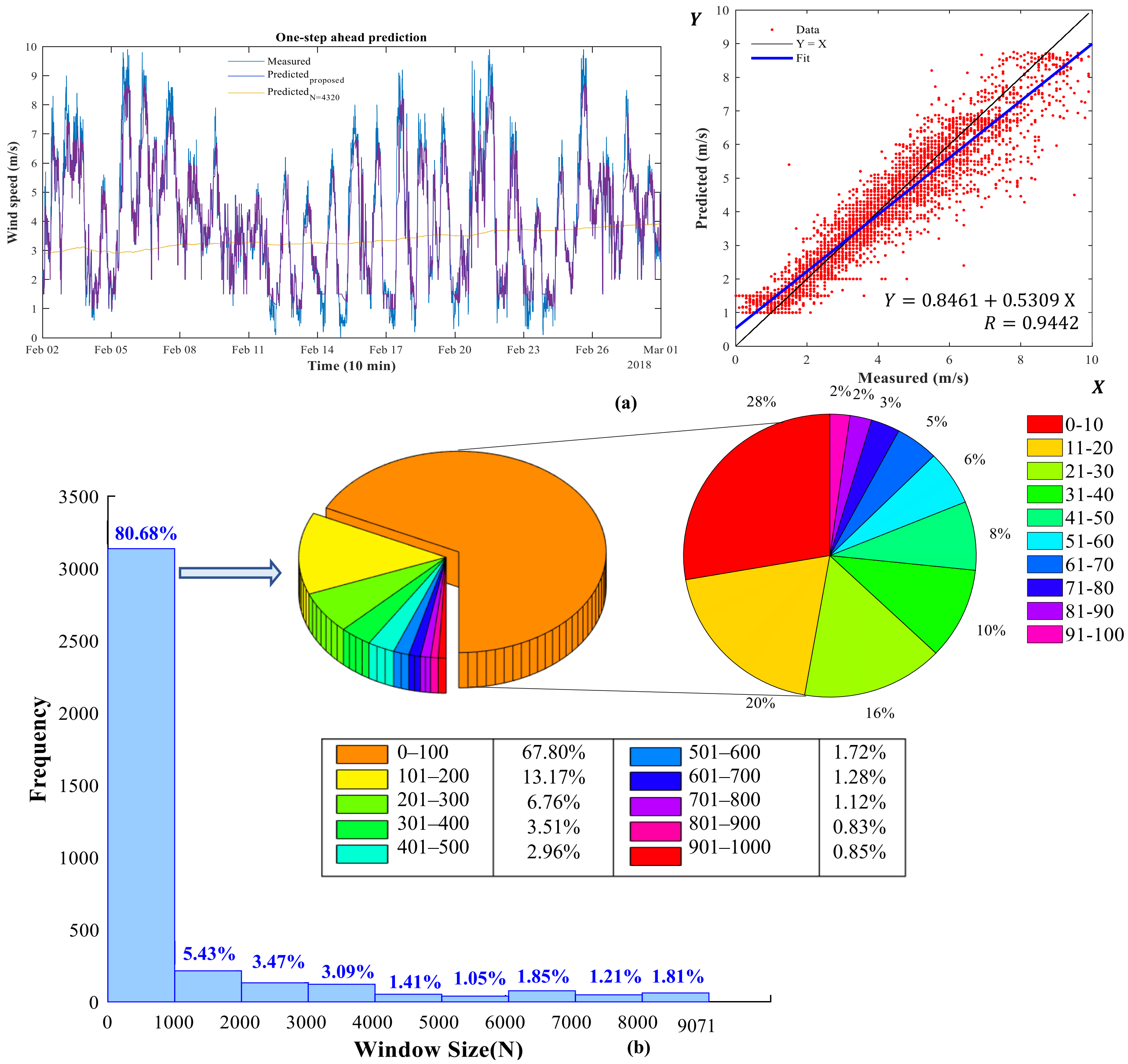


FIGURE 5. **(a) Forecasting results of the proposed model (b) Frequency of optimal moving window.**

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

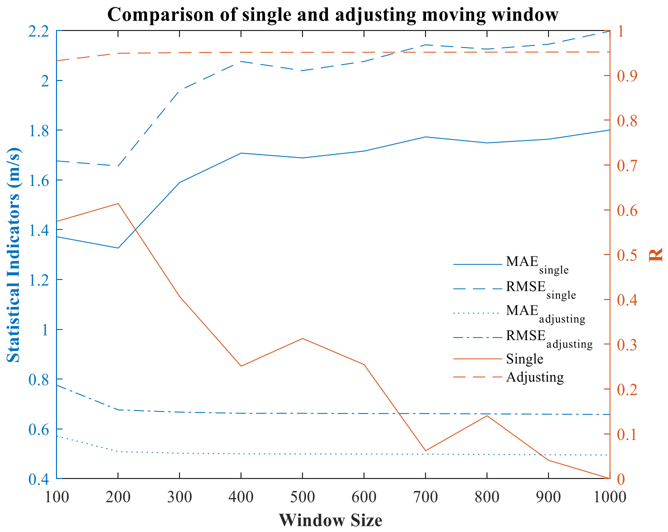


FIGURE 6. Effect of moving window size on statistical indicators.

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.

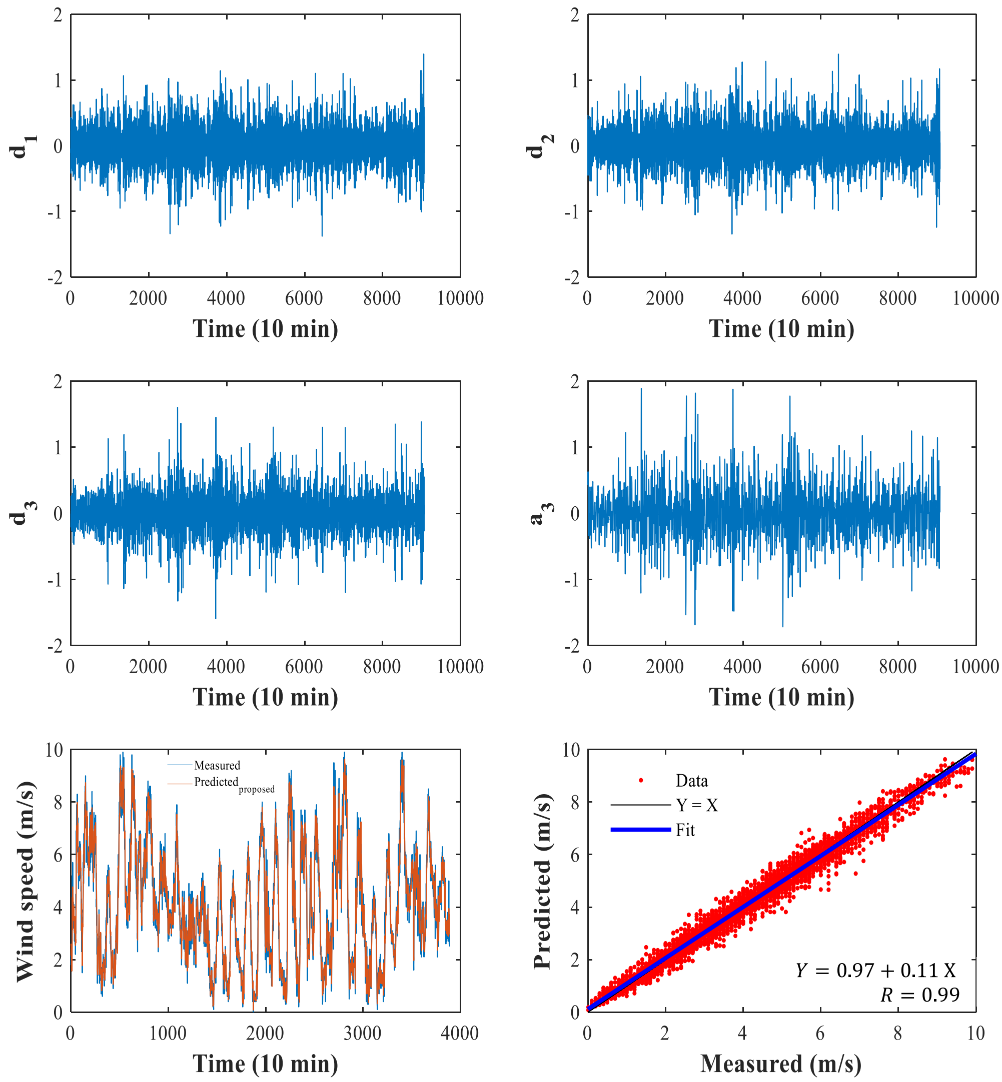


FIGURE 7. Decomposed signals of training dataset and forecasting results of the testing dataset.

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

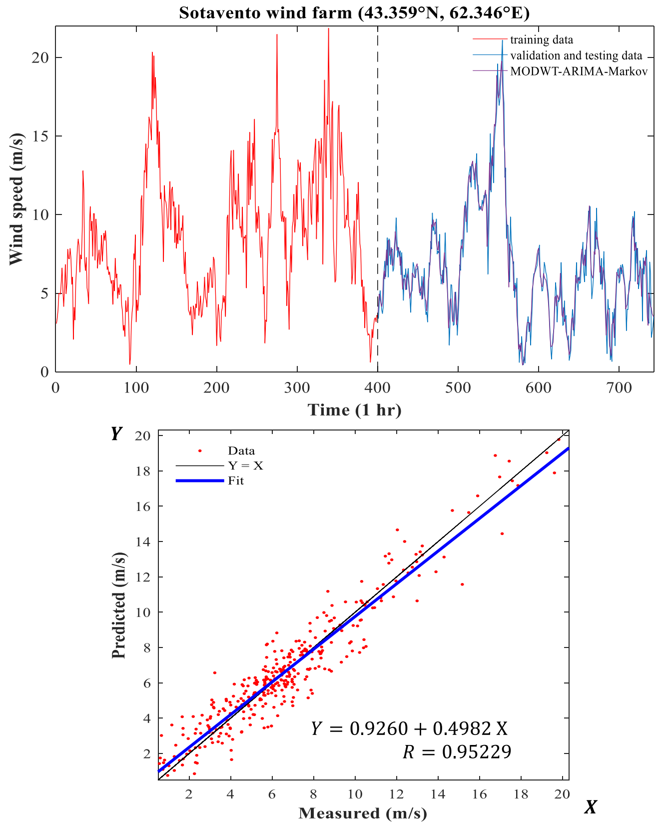


FIGURE 8. Forecasting results of the proposed model for a case study of Sotavento wind farm.

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.

CHAPTER 5

WIND SPEED PREDICTION FOR SMALL SAMPLE DATASET USING HYBRID FIRST-ORDER ACCUMULATED GENERATING OPERATION-BASED DOUBLE EXPONENTIAL SMOOTHING MODEL

This chapter contains content from the following article,

M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Wind speed prediction for small sample dataset using hybrid first-order accumulated generating operation-based double exponential smoothing model," Energy Science and Engineering, vol. 10 (3), pp. 726-739, 2022. DOI: https://doi.org/10.1002/ese3.1047

Received: 17 October 2021 | Revised: 28 November 2021 | Accepted: 16 December 2021

DOI: 10.1002/ese3.1047

Energy Science & Engineering Open Access

ORIGINAL ARTICLE

Wind speed prediction for small sample dataset using hybrid first-order accumulated generating operation-based double exponential smoothing model

Muhammad Uzair Yousuf[1,2] | **Ibrahim Al-Bahadly**[1] | **Ebubekir Avci**[1]

[1]Department of Mechanical and Electrical Engineering, Massey University, Palmerston North, New Zealand

[2]Department of Mechanical Engineering, NED University of Engineering and Technology, Karachi, Pakistan

Correspondence

Muhammad Uzair Yousuf, Department of Mechanical and Electrical Engineering, Massey University, Palmerston North 4442, New Zealand.
Email: M.U.Yousuf@massey.ac.nz

Funding information

This research was supported by Tilt Renewables Tararua Wind Farm Research Bursary, Department of Mechanical and Electrical Engineering, Massey University, and Higher Education Commission (HEC) Pakistan Grant 5-1/HRD/UESTPI(Batch VI)/6082/2019/HEC

Abstract

Wind power generation has recently emerged in many countries. Therefore, the availability of long-term historical wind speed data at various potential wind farm sites is limited. In these situations, such forecasting models are needed that comprehensively address the uncertainty of raw data based on small sample size. In this study, a hybrid first-order accumulated generating operation-based double exponential smoothing (AGO-HDES) model is proposed for very short-term wind speed forecasts. Firstly, the problems of traditional Holt's double exponential smoothing model are highlighted considering the wind speed data of Palmerston North, New Zealand. Next, three improvements are suggested for the traditional model with a rolling window of six data points. A mixed initialization method is introduced to improve the model performance. Finally, the superiority of the novel model is discussed by comparing the accuracy of the AGO-HDES model with other forecasting models. Results show that the AGO-HDES model increased the performance of the traditional model by 10%. Also, the modified model performed 7% better than other considered models with three times faster computational time.

KEYWORDS

exponential, forecasting, Holt, hybrid, sample size, statistical, wind speed

1 | INTRODUCTION

Wind power has proliferated in recent decades. Despite massive disruptions from the global pandemic and the decline in the gross domestic product (GDP) in 2020, wind capacity grew almost double its previous highest annual increase (equivalent to 111 GW).[1] As of the end of the last year, installed wind capacity increased by 17.84% to 733.3 GW, and wind power generation increased by 12% to 1591 TWh.[2] Overall, wind power accounts for 5.93% of global power generation.

According to renewable capacity statistics 2021,[1] China, the USA, Germany, and India account for 38.46%, 16.06%, 8.48%, and 5.26% share, respectively. In addition, an impressive increase in new installations is observed for Colombia and Russia, with almost 28 and 9 times more capacity than the previous year. Other notable countries include Sri Lanka (1.96 times), Kazakhstan (1.71 times), Argentina (1.63 times), and Vietnam (1.6 times) for a remarkable increase in wind power from the preceding year. An overall summary of the total installed wind capacity is visualized in Figure 1.

Although wind energy is one of the rapidly evolving technologies, the main disadvantage of wind energy is its intermittent, stochastic, and non-stationary nature. The randomness in wind speed hinders the stability of wind energy, which leads to additional operating costs. Therefore, effective wind speed forecasts are needed to improve wind energy utilization in power systems and to reduce the additional reserve capacity.

Wind speed forecasting methods are mainly classified into five categories: persistence, physical, statistical, artificial intelligence/machine learning (AI/ML), and hybrid. A brief description of these models is given below, whereas detailed reviews are available in the literature.[3-5]

- The persistence method is also known as the naïve predictor. In this method, the current wind speed is strongly correlated with the immediate future wind speeds as $v(t + \Delta t) = v(t)$. This method is only applicable for very short-term wind speed forecasts (0 – 30min).
- Physical models are based on numerical weather predictions (NWP) as input.[6] These models require extensive information of various meteorological data and characteristics of wind farms. Therefore, physical models are computationally extensive and are appropriate for medium- (6–24 h) to long-term (>24 h) predictions, such as 12 h,[7] 24 h,[8] 48 h,[9] and 72 h.[10]
- Statistical models map internal relationships among historical data. Such models are fast to calculate and easy to build. In addition, these models have strong linear estimation ability and are best suited for very short to short-term (30 min to 6 h) wind forecasts. Commonly used statistical models include ARIMA,[11] exponential smoothing,[12] gray predictors,[13] and Markov chain[14] models.
- Artificial intelligence/machine learning (AI/ML) models are motivated by how the human brain would solve the problem. These algorithms learn from past data to predict future wind speeds through computer

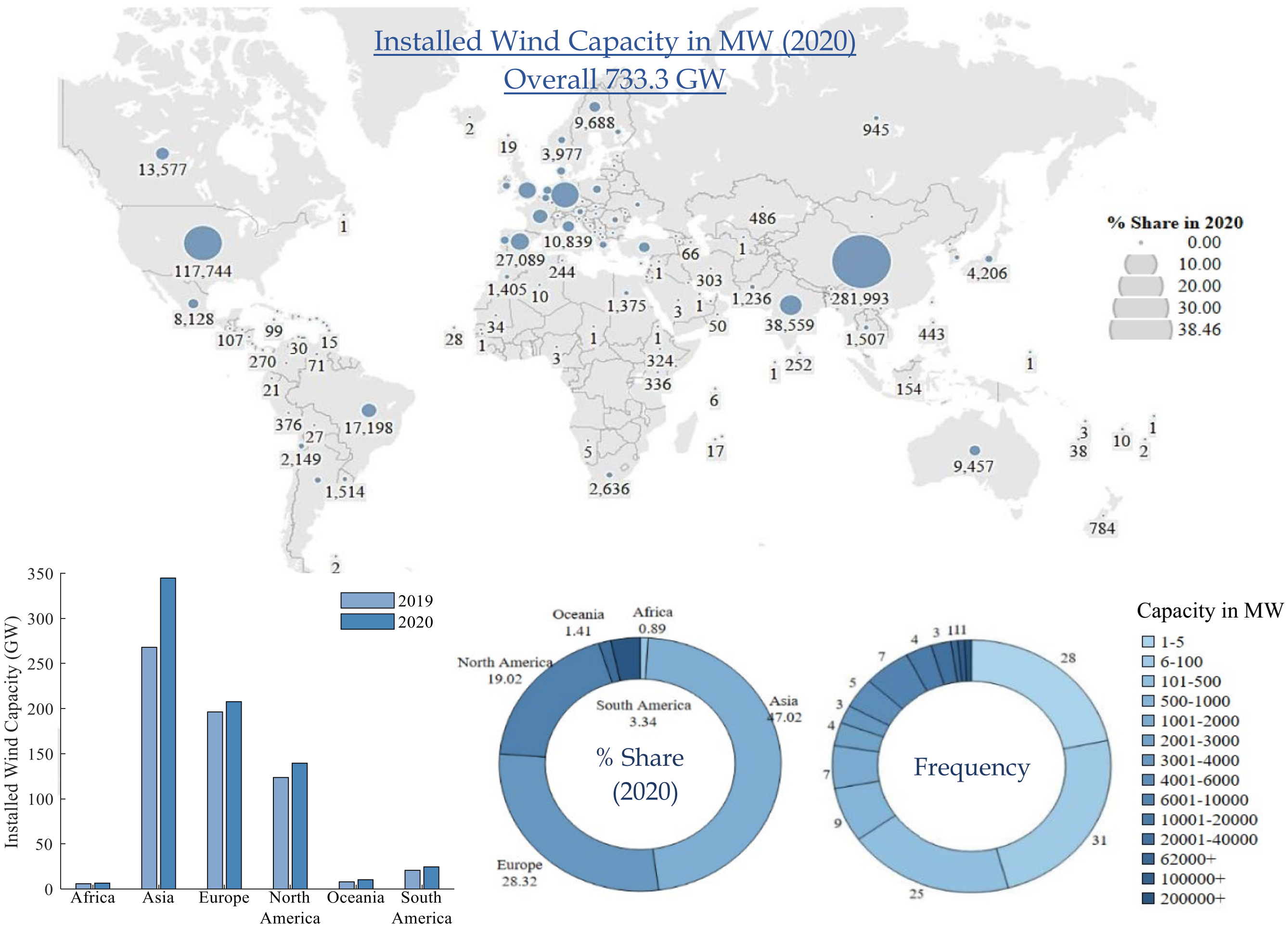


FIGURE 1 Global distribution of wind capacity (in MW) for the year 2020

simulations considering logical thinking, reasoning, and group behavior. Also, AI/ML models have powerful feature extraction and nonlinear mapping abilities. Such models effectively deal with incomplete and/or uncertain data and are suitable for very short- to short-term forecasts. AI/ML models are further subcategorized into traditional and deep learning models. Commonly applied traditional models include ANN,[15] SVM,[14] and Fuzzy Logic,[16] while deep learning models include LSTM,[14] DBN,[17] CNN,[18] ESN,[19] and ELM.[20]

- Hybrid models are the combination of two or more methods from other categories. Since a single predictive model does not fully capture the complex relationship of wind speed, therefore, a hybrid model can recompense for some of the deficiencies of individual models at the expense of additional complexity. Hybrid models can further be classified into four subcategories: weighted models, feature selection models, decomposition models, and error processing models. Some of the recent architectures include VMD-DE-ESN,[21] MODWT-ARIMA-Markov,[14] and ICEEMDAN–LSTM–GWO.[22]

While analyzing the global distribution of wind capacity for the year 2020 (as shown in Figure 1), it is observed that more than 20% of countries have a maximum installed wind capacity of 5 MW. Also, wind power generation has recently emerged in many countries. For example, in Saudi Arabia, the first wind farm with a capacity of 400MW started generating electricity in August 2021.[23] Hence, the availability of long-term historical data at various potential wind farm sites is limited. In these situations, such forecasting models that use only a small sample size are needed to comprehensively address the uncertainty of raw data.

Literature studies of wind speed forecasting showed that grey models are the most applied method for small sample datasets. Satisfactory prediction accuracy can be achieved with a minimum dataset of 5–11.[13] For this reason, many researchers focused on improving the grey models for wind forecasting and proposed improved versions, including FOTP-GM(1,1), NDGM(1,1), and FAGM(1,1).[24] Other statistical models such as ARIMA and Markov chain models are applied for datasets of 15[25] and 200[14] samples, respectively. On the contrary, exponential smoothing (ES) models are rarely considered for small sample datasets.

1.1 | Previous studies on exponential smoothing model

The exponential smoothing (ES) forecasting models[26-28] have been popular since their development. Forecasts generated from ES models are a weighted average of past observations, with previous values assigned exponentially decreasing weights. It means that the recent observations are given more weight as compared to the older ones. This simple structure provides fast and reliable forecasts for various applications, including wind speed predictions.

Standard ES models are based on the trend and seasonality of time series. These include simple ES, Holt's linear trend, Brown's linear trend, damped trend, and Holt-Winter ES methods. Most of these methods are successfully applied in wind speed forecasting.[12,29-38] However, as this study aims to analyze wind speed forecasts for a small dataset, therefore, double exponential smoothing models (trends with no seasonality) are considered for further discussions.

Kusiak and Zhang[34] applied Holt's linear trend model for six steps ahead wind speed forecast considering a resolution of 10 s. The evolutionary strategy algorithm is used to identify the optimal smoothing constants. The results showed that the prediction performance of the double exponential smoothing (DES) model is comparable to that of the artificial neural network. However, a dataset of 80,000 instances was considered for the analysis. Another method of DES, Brown's DES model, is applied by Yang et al.[32] The training is performed for a dataset of 1500. The smoothing constant is selected as 0.9 with no further discussion on the selection criteria. Furthermore, the DES method performed the least as compared to other models. Hybrid exponential smoothing models are also considered to improve forecasting performance. These include DES-PSO-BPNN-Elman[31] and CEEMDAN-CC-FA-Holt-SVR.[36] Both the models applied for a training dataset of 4500 and 1300, respectively. The optimal smoothing constants are evaluated based on the sum of squared errors of prediction. Results showed that the DES model is a better substitute for other statistical models. Besides technical programming, built-in software functions are also commonly used for the double exponential smoothing forecasting model. These include Minitab,[39] Eviews,[40] forecast package of *R*,[41] and SAS.[42] Table 1 summarizes the studies implementing the DES model for wind speed forecasting.

1.2 | Limitations of previous studies

From Table 1, three major observations are summarized as follows:

- None of the studies considered the DES model for a small wind speed dataset, and hence, the problems associated with the small dataset are not analyzed.

TABLE 1 Studies implementing the DES model for wind speed forecasting

Double Exponential Smoothing (DES) Model		
	Brown's method	**Holt's method**
Studies	Niu et al.,[31] Yang et al.,[32] Zhang et al.[33]	Kusiak and Zhang,[34] Prema and Rao,[35] Jiang et al.,[36] Priya and Arulanand[12]
Training data	1500,[32] 4500[31]	1300,[36] 80,000,[34] 4320[35]
Data resolution	10 min[31]	10 s,[34] 10 min,[35,36] 30 min[36]
Initial values	$S_0^{(2)} = S_0^{(1)} = v_1$[31] $S_0^{(1)} = \frac{\sum_{t=1}^{n+1} v_t}{2}$[39]	$\hat{v}_0 = L_0 + T_0$ where $L_0 = v_1,\ T_0 = \frac{\sum_{t=-5}^{-2} v_t - \sum_{t=-4}^{-1} v_t}{4}$[34] $L_0 = v_1,\ T_0 = 0$[36] $L_0 = \hat{\beta}_{o,T},\ T_0 = \hat{\beta}_{1,T}$
Smoothing constant	$\alpha = 0.9$[32]	evolutionary strategy algorithm[34] forecast package of *R*[35]
Hybrid model	DES-PSO-BPNN-Elman[31]	CEE-CC-Holt-GRNN[36]
Software	Eviews[39]	Minitab,[40,41] Eviews[39] forecast package of R[37]
Other model	Simple exponential smoothing model[29,30]	Holt-Winter exponential smoothing model[37,38]

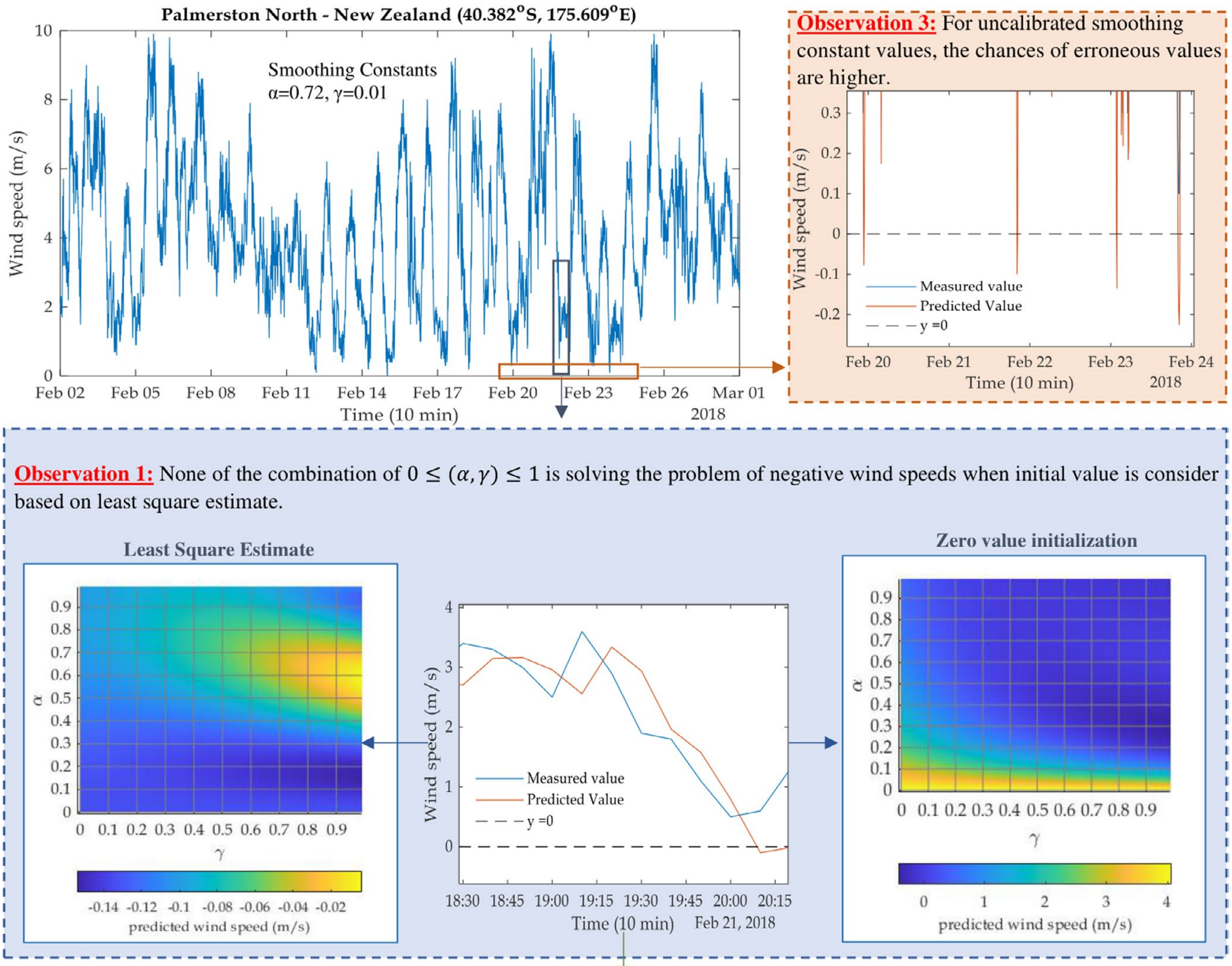
- There is no specific method to initialize the level and trend. However, several studies suggest a least-square estimate to be the best choice. This is why some software, such as Minitab,[39] SAS,[42] and Eviews,[40] are working on least-square estimates. A linear regression model is fit in time to available data as $\hat{v}_o = \hat{\beta}_{o,T} + \hat{\beta}_{1,T} t$. This would imply y-intercept ($\hat{\beta}_{o,T}$) and slope ($\hat{\beta}_{1,T}$) of the trend line as L_o and T_o.
- A common approach to finding optimal values of smoothing constants is to search a combination of parameters that minimizes the prediction errors. The values are constrained between 0 and 1. Furthermore, once the smoothing constants are evaluated, they remain constant for the rest of the dataset without recalibration.

To address the first observation, we analyze the wind speed data of Palmerston North, New Zealand. The summer season data are gathered from NIWA with a temporal resolution of 10 min. The traditional Holt's DES model is applied with a rolling window of 6, and the results are displayed in Figure 2. It is observed that the traditional model starts predicting negative values for non-negative time series. Even, for a time series fragment of $V = \{3.6, 2.9, 1.9, 1.8, 1.1, 0.5\}$, all the combinations of smoothing constants $0 < (\alpha, \gamma) < 1$ are providing negative predictions when the least-square estimate method is considered to initiate the level and trend. This led our focus to the second observation of how to initialize the level and trend.

In an early study, Makridakis and Hibon[43] analyzed seven methods to initiate the first forecast. The major conclusion is that type of initial values does not affect the accuracy of the prediction. Also, initializing by least square provides satisfactory results.[44] However, the conclusion is not suitable for a small-size dataset. As seen in Figure 2, the choice of the initial value is critical when the sample size is small. For the considered example of time series fragment, the Holt function of the forecast package of *R* applied a convenient initial value method, whereas Minitab used a least-square estimate. With $(\alpha, \gamma) = (0.72, 0.01)$, both software predicted negative wind speeds, that is, −0.2259 and −0.1002 m/s, respectively. However, if some other combination of zero-value initialization method is considered, then the problem is rectified. This brought our attention to the third major observation, that is, calibrating smoothing constants at every instant.

From the analysis of Figure 2, it is observed that if uncalibrated smoothing constants are considered, and then, there are more chances for negative wind speeds. It is because the addition of the latest observation does not change the parameters of the prediction model. Hence, unoptimized smoothing constants imposed an arbitrary handicap on the forecasting model.[45] As depicted in the heat map of Figure 2, with the zero-value initialization method, more than half combinations of $0 < (\alpha, \gamma) < 1$ are predicting erroneous values, while the rest are under the acceptable limits. Therefore, optimizing at each time origin is required than optimizing only once.[46]

To the best of the authors' knowledge, these problems are not discussed in any wind forecasting literature. Therefore, in this study, we improve the forecasting performance of the traditional Holt's method by considering the above-discussed observations to avoid erroneous wind speed forecasts. The main contributions of this study are as follows:



Observation 2: The selection of initial value has high effect on the forecasting results with small size dataset.

Initialization	**Method**					
	Least square estimate	Convenient initial values	Zero Values			Two values
Level	Intercept	v_1	0	Intercept	0	v_1
Trend	Slope	$v_2 - v_1$	0	0	Slope	v_2
Predicted value	-0.1002	-0.2259	0.7426	0.6903	-0.0479	4.5172

FIGURE 2 Forecasting issues with the traditional Holt's Double exponential smoothing model for a rolling widow of 6

- A hybrid model integrating accumulated generating operator and Holt's DES model (AGO-HDES) is proposed. Three improvements are suggested to improve the performance of the traditional model for very short-term wind speed forecasts.
- A mixed initialization method is introduced, which has not been discussed in any literature. Instead of considering a single initial value throughout the process, the first initialization method is applied to identify the optimal damping constant, while the second method is used for final forecasts. Overall, eight choices of initial values from four categories are considered for the comparison.
- A comparative analysis is carried out for AGO-HDES with traditional Holt's DES, Brown's DES, and damped trend DES models.
- The superiority of the proposed model is discussed by comparing the forecasting performance of the AGO-HDES model with other statistical forecasting models available in the literature.

The remainder of this paper is organized as follows. In section 2, the traditional models are defined. In section 3, three improvements are suggested to enhance the forecasting accuracy. The performance of the novel hybrid model is discussed in Section 4, considering wind speed data of Palmerston North, New Zealand. Finally, conclusions are highlighted in Section 5.

2 | MATERIALS AND METHODS

2.1 | Double exponential smoothing model

Double exponential smoothing models are preferred when data exhibit a trend. Two commonly used DES models are Brown's method and Holt's method.

Brown's method is the extension of simple exponential smoothing model and is given by

$$S_t^{(1)} = \alpha v_t + (1-\alpha)\, S_{t-1}^{(1)} \tag{1}$$

$$S_t^{(2)} = \alpha S_t^{(1)} + (1-\alpha)\, S_{t-1}^{(2)} \tag{2}$$

where α is the smoothing constant and $S_t^{(1)}$ and $S_t^{(2)}$ are first- and second-order exponential smoothing values. Based on $S_t^{(1)}$ and $S_t^{(2)}$, k-step ahead forecast can be evaluated as

$$\hat{v}_{t+k} = a_t + k b_t \tag{3}$$

where $a_t = 2S_t^{(1)} - S_t^{(2)}$ and $b_t = \frac{\alpha}{1-\alpha}\left(S_t^{(1)} - S_t^{(2)}\right)$

Different from Brown, the time series is divided into two components in Holt's method: Level (L_t) and Trend (T_t). The two components can be calculated as

$$L_t = \alpha v_t + (1-\alpha)\left(L_{t-1} + T_{t-1}\right) \tag{4}$$

$$T_t = \gamma\left(L_t - L_{t-1}\right) + (1-\gamma)\, T_{t-1} \tag{5}$$

where α and γ are the smoothing constants. In Holt's method, the trend also updates itself via equation (5), expressed as the difference between the last two smoothed values.[47] Based on level and trend, k-step ahead forecast can be evaluated as

$$\hat{v}_{t+k} = L_t + kT_t \tag{6}$$

In Holt's method, generated forecast shows an indefinite increasing or decreasing trend. Gardner and McKenzie[48] introduced a damped parameter to the traditional model to flatten the trend line sometime in the future. The level and trend for the damped model are calculated as:

$$L_t = \alpha v_t + (1-\alpha)\left(L_{t-1} + \phi T_{t-1}\right) \tag{7}$$

$$T_t = \gamma\left(L_t - L_{t-1}\right) + (1-\gamma)\,\phi T_{t-1} \tag{8}$$

where ϕ is the damping parameter. If $\phi = 1$, then the damped model is identical to Holt's DES model.

3 | IMPROVED DOUBLE EXPONENTIAL SMOOTHING MODEL

For any of the above-mentioned DES models, forecasting is summarized in three major steps: initial value calculation, optimal smoothing constants selection, and applying forecasting equations (as discussed in Section 2). In this section, we discuss the improvements related to the first two steps. Hereafter, the traditional model with the first two improvements is termed as HDES.

3.1 | Improvement # 01: optimal smoothing constant

A common approach for finding the optimal values of smoothing constants (α, γ) is to explore the parameter combination that minimizes the sum or mean of squared error of predictions,

$$\mathrm{SSE} = \sum_{t=1}^{n} \left(v_t - \hat{v}_t\right)^2 \tag{9}$$

$$\mathrm{MSE} = \frac{\mathrm{SSE}}{n} \tag{10}$$

However, there are two issues with the traditional approach. It is possible that a combination of (α, γ) with least MSE might predict negative wind speed. As an example, a time series fragment of $V = \{1.1, 0.8, 0.6, 0.2, 0.2, 0.3\}$, suggests the best combination as (α = 0.41, γ = 0.01) based on least MSE. However, the predicted value is −0.0506 m/s. The same problem occurred with other rolling windows (N) as well, as shown in Table 2. One solution is to apply Box-Cox transformation with $\lambda = 0$ for imposing a positivity constraint.

Secondly, in most cases, the range of smoothing constants are constrained between $0 \le (\alpha, \gamma) \le 1$. As discussed in Figure 2, it might be possible that none of the combinations is applicable for prediction, and thus, the traditional constraints limit the model applicability. Although Hyndman et al.[49] proved that the model is forecastable if $0 < \alpha < 2$ and $0 < \gamma < 4 - 2\alpha$ (see condition 1 of Theorem 10.2 in ref. [49]); however, the concept of admissible constraints is rarely used for wind speed predictions (see Table 1).

TABLE 2 One step ahead wind speed forecast for the best possible combination

N	V (m/s)	Method[a]	$\hat{v}_0 = L_t + T_t$ (m/s)	α	γ	$\hat{v}_{t+1}$ (m/s)	Sign ($\hat{v}_{t+1}$)
5	{1.6, 1.4, 1.1, 0.4, 0.3}	Least MSE	1.68	0.45	0.01	−0.1208	−
		Minitab	—	—	—	—	
		Forecast package of R	1.4	0	0	0.4	+
6	{2.1, 1.6, 1.1, 0.8, 0.6, 0.3}	Least MSE	1.9619	0.7	0.01	−0.0799	−
		Minitab	2.0921	1.7259	0.0999	−0.0629	−
		Forecast package of R	1.6	0.351	0.9123	−0.0261	−
7	{1, 1.2, 0.5, 0.4, 0.2, 0.2, 0.3}	Least MSE	1.0143	0.28	0.01	−0.0670	−
		Minitab	1.0143	1.7453	0.01	0.0131	+
		Forecast package of R	1.2	0.9235	0.3931	0.2521	+
8	{1.7, 1.8, 1.5, 1.3, 0.6, 0.6, 0.6, 0.2}	Least MSE	1.8667	0.89	0.01	−0.0219	−
		Minitab	1.8309	0.3629	0.4207	−0.0119	−
		Forecast package of R	1.8	0.9092	0.3915	−0.0037	−

[a]Initial Value calculation

Least MSE: Least-square estimate

Minitab: Backcasting

Forecasting package of R: Convenient value method

In this study, instead of identifying a single combination of (α, γ), the model stores a vector of all combinations. Initially, the smoothing constants are considered between $0 < (\alpha, \gamma) < 1$ with a step size of 0.1. Therefore, a vector of 81 combinations is stored in memory. If a particular combination predicts a negative wind speed, it is discarded from the primary vector. The final vector is then sorted by the least MSE value to identify the ascending series of combinations. However, if all the combinations predict erroneous values, then admissible constraints are considered to find the optimal smoothing constants in the next step.

3.2 | Improvement # 02: Initial Value calculation

Considering the simplest form of exponential smoothing model with no trend,

$$\begin{gathered}
\hat{v}_1 = \alpha v_1 + (1-\alpha)\hat{v}_0 \\
\hat{v}_2 = \alpha v_2 + (1-\alpha)\hat{v}_1 = \alpha v_2 + (1-\alpha)\left\{\alpha v_1 + (1-\alpha)\hat{v}_0\right\} \\
\hat{v}_2 = \alpha\left\{v_2 + (1-\alpha)v_1\right\} + (1-\alpha)^2\hat{v}_0 \\
\hat{v}_3 = \alpha\left\{v_3 + (1-\alpha)v_2 + (1-\alpha)^2 v_1\right\} + (1-\alpha)^3\hat{v}_0 \\
\vdots \\
\hat{v}_t = \alpha\left\{v_t + (1-\alpha)v_{t-1} + \ldots + (1-\alpha)^{t-1}v_1\right\} + (1-\alpha)^t\hat{v}_0
\end{gathered} \tag{11}$$

The equation shows that the final forecast $(\hat{v}_t)$ is influenced by the initial value $(\hat{v}_0)$ with a factor $(1-\alpha)^t$. Therefore, for a small sample dataset, the estimation of initial values has high relevance to the final forecast.

Different methods are identified in the literature to initialize the level and trend. A general approach is to obtain by least-square estimates. A linear regression model is fit in time to available data as,

$$\hat{v}_o = \hat{\beta}_{o,T} + \hat{\beta}_{1,T}t \tag{12}$$

where $\hat{\beta}_{o,T}$ and $\hat{\beta}_{1,T}$ are considered as the initial values L_o and T_o.

Other methods include convenient initial values, zero values, and first data point. In this study, we considered eight different initialization methods, including least-square estimates. Such methods include ($L_o = 0$, $T_o = 0$), ($L_o = v_1$, $T_o = 0$), ($L_o = 0$, $T_o = v_1$), ($L_o = v_1$, $T_o = v_1$), ($L_o = v_1$, $T_o = v_2 - v_1$), ($L_o = \hat{\beta}_{o,T}$, $T_o = 0$), and ($L_o = 0$, $T_o = \hat{\beta}_{1,T}$). These methods are labeled as IV1–IV8.

This study introduces a mixed initial value method. The first initialization method is applied to identify the optimal damping constant, while the second method is used for final forecasts.

3.3 | Improvement # 03: first-order accumulated generating operation

The third modification is to apply First-order Accumulated Generating Operation (1-AGO) to the original time series as:

$$V^{(1)} = \left\{v_1^{(1)}, v_2^{(1)}, v_3^{(1)}, \cdots, v_n^{(1)}\right\} \tag{13}$$

where $v_k^{(1)} = \sum_{i=1}^{k} v_i,\ k = 1, 2, 3, \cdots, n$

Once the prediction is completed, then the forecasting series can be obtained by inverse 1-AGO as:

$$\hat{v}_k = \hat{v}_k^{(1)} - \hat{v}_{k-1}^{(1)} \tag{14}$$

Based on the above improvements, the modified model is summarized as follows:

1. Apply 1-AGO on the original wind speed sequence to obtain $V^{(1)}$.
2. Select the first method to initialize level and trend. This method is used to obtain the optimal damping constants.
3. Calculate the mean squared error for all the combinations of smoothing constants considering traditional constraints, that is, $0 < (\alpha, \gamma) < 1$ with a step size of 0.1 and stored in a vector and then sort the matrix in ascending order corresponding to the least MSE against the values of smoothing constants. For more optimized results, a smaller step size such as 0.01 can also be used at the expense of higher computational time and storage.
4. Select the second method to initialize level and trend and then compute the predictive value using the DES model based on the order of optimal smoothing constant matrix.
5. Apply inverse 1-AGO on the forecasting sequence.
6. Discard the combination if the model forecasts negative wind speed, then use the next combination for prediction.
7. If none of the combinations is applicable, then apply admissible constraints, that is, $1 < \alpha < 2$, to find the best order and repeat steps 3 to 6.

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

4 | RESULTS AND DISCUSSIONS

4.1 | Site description

New Zealand has 17 wind farms operating with a total installed capacity of 690 MW. It supplies about 6% of New Zealand's annual electricity production that roughly corresponds to the electricity consumption of 300,000 households per year. Of these wind farms, more than 40% of the wind capacity is installed near Palmerston North. Tararua wind farm is the largest one (161 MW) and is 10 km from Palmerston North. Other operating wind farms near the city are Te Apiti Wind Farm (90 MW) and Te Rere Hau Wind Farm (48.5 MW). In the near future, Turitea wind farm will be the largest wind farm of New Zealand (222 MW) and is planned approximately 10 km southeast of Palmerston North.[50]

Palmerston North (40.382°S, 175.609°E) has a large number of strong winds (>8.6 m/s) with west-northwest as the predominant direction. From the analysis of the recorded data, the seasonal percentage of strong winds is 37%, 26%, 19%, and 18% in spring, summer, winter, and autumn, respectively.[51] In this study, wind speed data from December 1, 2017 to February 28, 2018 is considered.[52] The collected data are first preprocessed to identify any anomalies using Box-plot analysis. A total of 30 inaccuracies were found that were imputed using Piecewise Cubic Hermite Interpolating Polynomials (PCHIP). Next, the finalized observations are divided into training and testing datasets with a ratio of 3:1. Thus, the training dataset contains 9072 observations, whereas the testing dataset has 3888 observations.

4.2 | Mixed initialization method

The results of the mixed initialization method for HDES, damped HDES, and AGO-HDES are tabulated in Tables 3–5. First initialization is applied to evaluate the optimal smoothing constants, whereas the latter is used for wind speed forecast. Four categories of initial values are considered in the analysis. These include zero level and non-zero trend (IV4 and IV8), non-zero level and zero trend (IV3 and IV7), both zero (IV2) and both non-zero (IV1, IV5, and IV6).

Results show that the mixed method strategy is improving the model performance. The best initial values to evaluate optimal smoothing parameters are IV2 and IV8 for all three models. Similarly, the best initialization method for the final forecast is IV3 and IV7 for HDES, IV2, IV3 and IV7 for damped HDES, and IV1 and IV8 for AGO-HDES. More specifically, zero trend is the best choice for the original time series, whereas zero level provides the best forecast for the AGO time series. It is because the AGO time series always have an uptrend.

Compared with HDES and damped HDES models, it is observed that the damped model provides three times more suitable initialization combinations than the HDES model. Among 64 combinations, HDES provides four options with the least MSE, while the damped HDES model has 13 appropriate options for the same case study.

While comparing all three models, it is observed that the novel AGO-HDES is lesser affected by incorrect selection of initialization method. The MSE values for AGO-HDES are adjusted between 0.3705 and 3.9469 whereas, for HDES and damped HDES, MSE varies from 0.3839 to

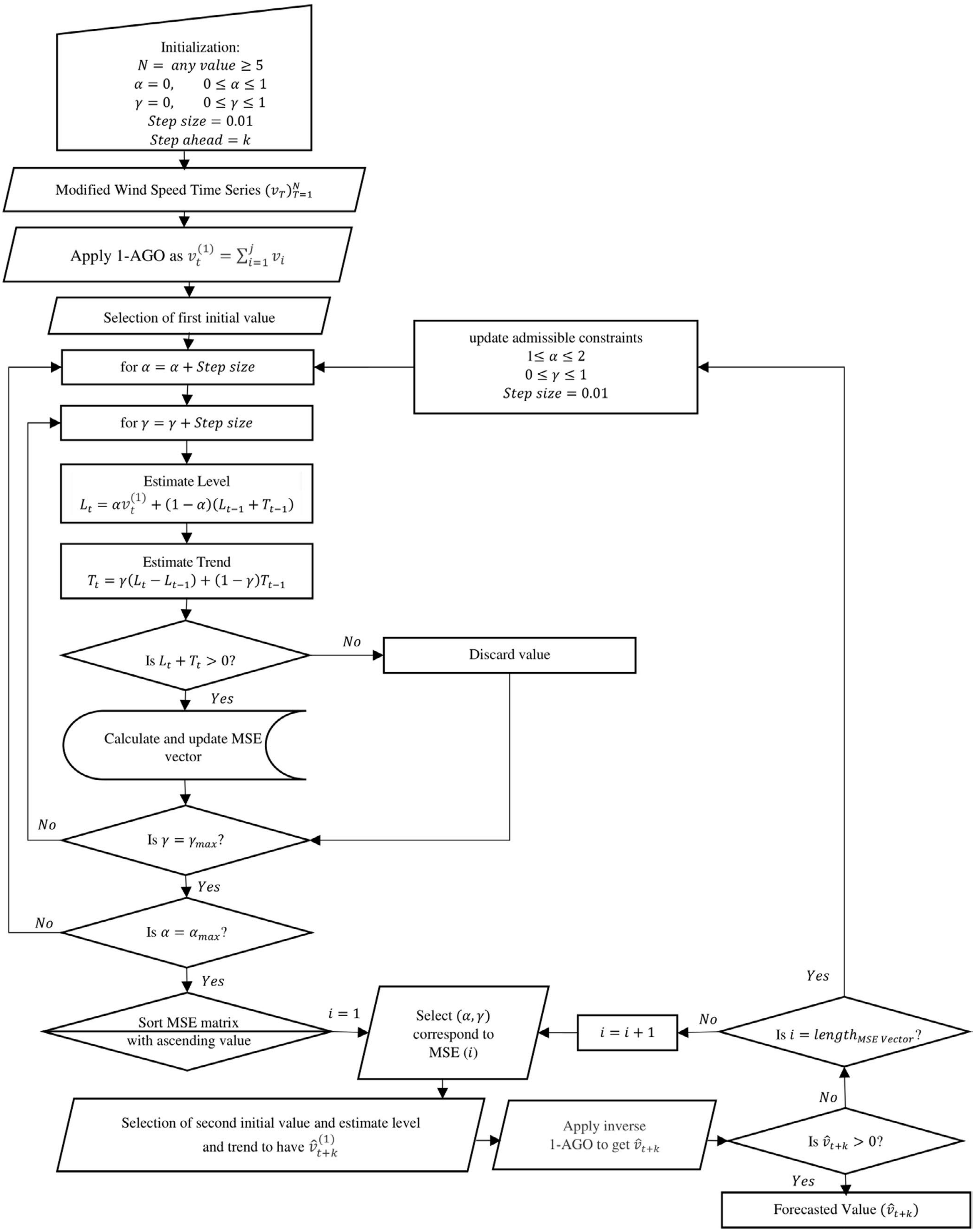


FIGURE 3 Proposed framework for AGO-HDES forecasting model

TABLE 3 MSE of the mixed initialization method for HDES

	Evaluation of optimal smoothing constants							
Forecast	**IV1**	**IV2**	**IV3**	**IV4**	**IV5**	**IV6**	**IV7**	**IV8**
IV1	0.4652	0.4456	0.4723	0.6357	0.6414	0.5045	0.4512	0.4409
IV2	3.0052	0.4996	2.1920	0.6369	0.6463	1.2354	2.3876	0.4858
IV3	0.4696	**0.3891**	0.4656	0.6351	0.6437	0.5196	0.4457	**0.3846**
IV4	323.03	11.3610	126.67	0.6451	0.6440	54.212	109.43	11.170
IV5	371.16	9.2163	139.37	0.6471	0.6425	60.461	119.56	9.0643
IV6	5.0482	0.5548	2.1818	0.6513	0.6573	0.5248	2.1044	0.5529
IV7	0.4998	**0.3913**	0.4869	0.6359	0.6444	0.5319	0.4428	**0.3839**
IV8	3.0900	0.5297	2.1905	0.6356	0.6430	1.2174	2.3870	0.5294

Bold indicates best options of initialization method based on MSE criterion.

TABLE 4 MSE of the mixed initialization method for damped HDES

	Evaluation of optimal smoothing constants							
Forecast	**IV1**	**IV2**	**IV3**	**IV4**	**IV5**	**IV6**	**IV7**	**IV8**
IV1	0.4006	0.3816	0.4485	0.5174	0.4778	0.4760	0.4164	**0.3796**
IV2	2.6392	0.3853	1.9697	1.6949	1.3779	3.4763	1.9718	0.3800
IV3	0.4220	0.3825	0.4359	0.5264	0.4788	0.5411	0.4295	**0.3782**
IV4	75.6383	0.5929	11.7941	0.3813	1.1462	16.4322	14.2378	0.6081
IV5	94.2902	0.5501	12.1401	1.8834	0.4499	22.5899	15.2632	0.5651
IV6	1.5031	0.3853	0.5258	0.5072	0.4687	0.4255	0.5526	0.3802
IV7	0.4418	0.3801	0.4566	0.5914	0.4877	0.5704	0.4258	**0.3762**
IV8	2.6229	0.3872	1.9600	1.6113	1.3598	3.4594	1.9642	0.3834

Bold indicates best options of initialization method based on MSE criterion.

TABLE 5 MSE of the mixed initialization method for AGO-HDES

	Evaluation of optimal smoothing constants							
Forecast	**IV1**	**IV2**	**IV3**	**IV4**	**IV5**	**IV6**	**IV7**	**IV8**
IV1	0.4396	**0.3705**	0.4479	0.6337	0.6384	0.4770	0.4257	**0.3711**
IV2	3.0052	0.4996	2.1920	0.6369	0.6463	1.2354	2.3876	0.4858
IV3	3.9469	0.5365	2.6106	0.6425	0.6524	1.5322	2.8676	0.5154
IV4	0.4696	0.3891	0.4656	0.6351	0.6437	0.5196	0.4457	0.3846
IV5	0.6462	0.3960	0.6318	0.6403	0.6496	0.6564	0.6793	0.3842
IV6	0.6637	0.3919	0.6526	0.6398	0.6485	0.6432	0.6795	0.3814
IV7	2.9736	0.4942	2.1734	0.6355	0.6427	1.2117	2.3659	0.4825
IV8	0.4741	**0.3744**	0.4639	0.6341	0.6411	0.4957	0.4444	**0.3734**

Bold indicates best options of initialization method based on MSE criterion.

371.16 and 0.3762 to 22.59, respectively. Therefore, it is concluded that the novel AGO-HDES model is a better option than HDES and damped HDES models.

It should be noted that this conclusion is only applicable for small sample dataset. For larger rolling windows, the initialization method does not affect the accuracy of the prediction, as shown in Figure 4. Therefore, it follows the same conclusion as inferred by Makridakis and Hibon.[43]

4.3 | One step ahead wind speed forecast

Figure 5 shows one step ahead wind speed forecast for multiple models. For the AGO-HDES method, the most frequent pairs for α and γ belonged to the intervals (0.7, 1.0) and (0, 0.3), respectively. It shows that higher weights are given to the most recent observations, whereas lower weights are given to the earlier observations. Overall, it concludes that forecasts are more responsive to recent levels and less emphasized by trend estimates. Figure 5 is a screenshot of the sample real-time forecasting video. The complete video is attached in the Video S1.

This study evaluates the forecasting models' performance based on commonly used mean absolute error (MAE) and root mean square error (RMSE). Furthermore, the enhancement of the traditional model is estimated through improving ratio. The formulations of the considered three metrics are given in equations 15 to 17.

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|\hat{v}_t - v_t\right| \tag{15}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(\hat{v}_t - v_t\right)^2} \tag{16}$$

$$\text{Improving ratio }(\%) = \frac{I_{\text{reference}} - I_{\text{modified}}}{I_{\text{reference}}} \tag{17}$$

where $\hat{v}_t$ and v_t are the predicted and measured wind speeds at instant t, n is the number of observations, and I is a general index for comparison.

Table 6 shows the performance of the forecasting models. Holt DES method is better than the Brown DES method. This is because Holt's method updates both level and trend. However, the traditional DES methods are lesser accurate than other statistical models. The proposed modifications enhanced the prediction accuracy. Analysis shows that the mixed initialization method increased the performance of the traditional model by 10%. The MAE and RMSE values of traditional DES model dropped to 0.452 m/s and 0.608 m/s from 0.509 and 0.678 m/s, respectively. Similarly, the modified models performed 7% better than GM (1,1) and Markov chain models. The accuracy measures show that AGO-HDES performed the best among all models with the least MAE and RMSE values of 0.452 m/s and 0.609 m/s. With the lowest RMSE, the AGO-HDES model captures the intermittency of small sample size wind speed in finer detail with fewer outlier errors. Furthermore, Figure 6 shows the scatter plot between measured and predicted wind speed by the AGO-HDES model for one step ahead forecast. The correlation coefficient of 0.96 also endorses the strong correlation between measured and predicted wind speed.

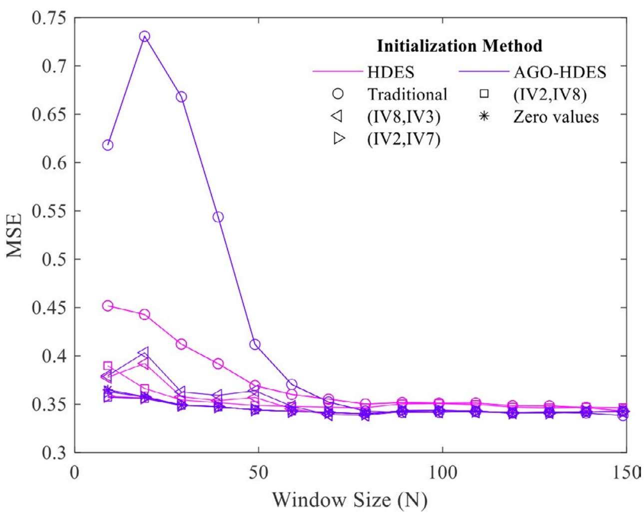


FIGURE 4 Effect of the initialization method on the performance of the proposed forecasting models for varying rolling window

In addition, the best performance is also accompanied by very short computational time. All models were programmed in MATLAB with an Intel i5, 1.70 GHz processor with quad-core, and 16 GB RAM. For each model, the test was run ten times, and the average computational time was considered. Table 7 shows that the AGO-HDES model is almost three times faster than the GM (1,1) model. Therefore, the proposed model is practically applicable for wind speed forecasts.

5 | CONCLUSION

The performance of most of the existing data-driven wind speed forecasting models decreases considerably for small datasets. Also, long-term wind speed historical data are not always available for newly identified sites. Therefore, in this study, a hybrid first-order accumulated generating operation-based double exponential smoothing model is proposed for small-size wind datasets. The generalized conclusions are as follows:

- The traditional model might predict negative wind speed values because of two reasons. Firstly, the most common initialization method is the least-square estimate, which is biased to a small-size dataset. Secondly, a combination of smoothing constants with the least MSE might cause erroneous wind speed predictions.
- In comparison, we proposed three modifications and named the model as AGO-HDES. Firstly, the model stores a vector of all combinations of smoothing constants instead of identifying a single combination. If a particular combination predicts a negative wind speed, it is discarded from the primary vector. The final vector is then sorted by w.r.t least MSE value to identify the ascending series of combinations. Secondly, a mixed initial value method is introduced. Instead of considering a single initial value throughout the process, the first initialization method is applied to identify the optimal

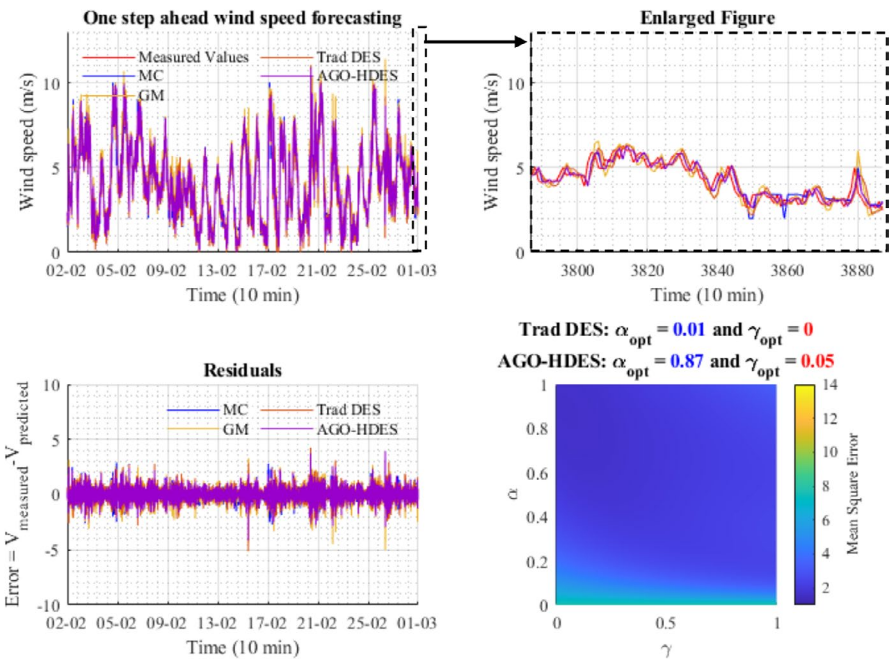


FIGURE 5 One step ahead wind speed forecast for Markov Chain, GM (1,1), Traditional DES and AGI-HDES models

TABLE 6 Performance evaluation of the forecasting models

Type	Models	MAE (m/s)	Improving ratio w.r.t. the best model (%)	RMSE (m/s)	Improving ratio w.r.t. the best model (%)
Traditional	DES	0.509	11.20	0.678	10.17
	Damped DES	0.476	5.04	0.635	4.09
	Brown	0.515	12.23	0.695	12.37
Modified	HDES	0.459	1.53	0.620	1.77
	Damped HDES	0.457	1.09	0.613	0.65
	AGO-HDES	**0.452**	—	**0.609**	—
Comparison	$GM(1,1)_{\alpha'=0.9}$[13a]	0.489	7.57	0.654	6.88
	$GM(1,1)_{\alpha'=\text{var}}$[13]	0.544	16.91	0.738	17.48
	Markov Chain[14]	0.487	7.19	0.650	6.31

[a]Here α' is the weighting factor.

Bold indicates best options of initialization method based on MSE criterion.

damping constant, while the second method is used for final forecasts. Thirdly, First-order Accumulated Generating Operation (AGO) is applied to the original time series. Once the prediction is completed, then the forecasting series can be obtained by inverse-AGO (I-AGO).

- In a comparison of both HDES and AGO-HDES models, it is observed that the novel AGO-HDES is lesser affected by incorrect selection of initialization method. As a result, the MSE values for AGO-HDES are adjusted between 0.3705 and 3.9469 whereas, for HDES, it varies from 0.3839 to 371.16.
- The novel AGO-HDES model increased the performance of the traditional model by 10%. Similarly, the modified model performed 7% better than GM (1,1) and Markov chain models.
- The AGO-HDES model is almost three times faster than the GM (1,1) model with the same rolling window.

The novel AGO-HDES model is well-suited for short-term wind speed forecasts. However, the problem of trend lag is not eliminated entirely, and a study is needed in this regard. Dixit et al.[53] addressed the issue of "prediction lag" or "timing error" in wave height forecasting by proposing a multilevel neuro-wavelet technique. The central idea of the study is to apply wavelet transformation multiple times such that the correlation is removed. Therefore, decomposition-based hybrid models are

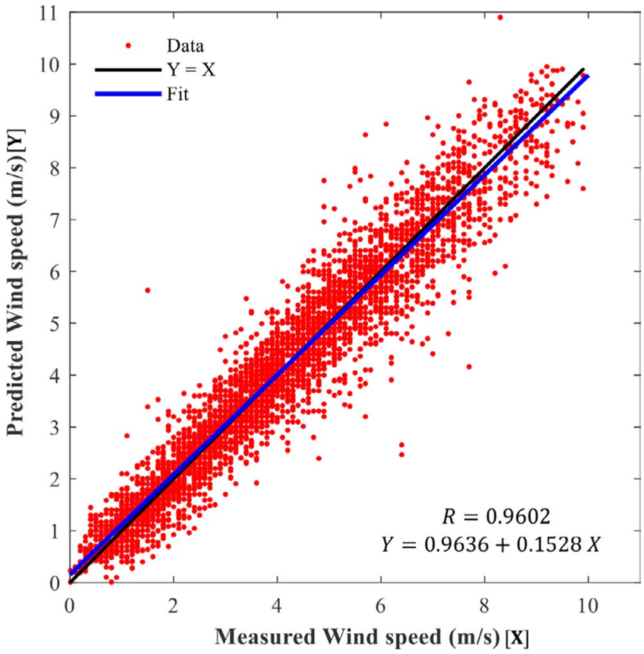


FIGURE 6 Scatter plot for measured and predicted wind speed

TABLE 7 Average computational time per prediction

Models	Computational time (s)
Markov Chain[14]	7.82
GM (1,1)[13]	0.4104
AGO-HDES	0.1496

required to address the same in wind speed forecasting. Similarly, a dataset of only six values was considered, with a resolution of 10 min, to predict one step ahead wind speed. Hence, the cyclic component is not considered. Therefore, an adaptive higher order exponential smoothing model will also be considered in the future for low-resolution data to address the diurnal effects of wind speed.

ACKNOWLEDGMENTS

The authors would like to acknowledge the NIWA for their public data.

ORCID

Muhammad Uzair Yousuf https://orcid.org/0000-0003-2177-6804
Ibrahim Al-Bahadly https://orcid.org/0000-0002-0451-1786
Ebubekir Avci https://orcid.org/0000-0002-4687-0859

ENDNOTE

[1] In some articles, trend is termed as growth.

REFERENCES

1. Renewable capacity statistics. 2021. Accessed July 30, 2021. https://www.irena.org/Statistics/Download-Data
2. Statistical review of world energy. 2021. Accessed July 30, 2021. https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2021-full-report.pdf
3. Yousuf MU, Al-Bahadly I, Avci E. Current perspective on the accuracy of deterministic wind speed and power forecasting. *IEEE Access*. 2019;7:159547-159564.
4. Santhosh M, Venkaiah C, Vinod Kumar D. Current advances and approaches in wind speed and wind power forecasting for improved renewable energy integration: a review. *Engineering Reports*. 2020;2(6):e12178.
5. Wang Y, Zou R, Liu F, Zhang L, Liu Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl Energy*. 2021;304:117766.
6. Salazar A, Che Y, Zheng J, Xiao F. Multivariable neural network to postprocess short-term, hub-height wind forecasts. *Energy Sci Eng*. 2021.
7. Zjavka L, Mišák S. Direct wind power forecasting using a polynomial decomposition of the general differential equation. *IEEE Trans Sustain Energy*. 2018;9(4):1529-1539.
8. Yang M, Shi C, Liu H. Day-ahead wind power forecasting based on the clustering of equivalent power curves. *Energy*. 2021;218:119515.
9. Gao J, Chongfuangprinya P, Ye Y, Yang B. A three-layer hybrid model for wind power prediction. In: *IEEE Power & Energy Society General Meeting (PESGM)*. IEEE; 2020:1-5.
10. Hoolohan V, Tomlin AS, Cockerill T. Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data. *Renewable Energy*. 2018;126:1043-1054.
11. Liu X, Lin Z, Feng Z. Short-term offshore wind speed forecast by seasonal ARIMA-A comparison against GRU and LSTM. *Energy*. 2021;227:120492.
12. Priya B, Arulanand N. Univariate and multivariate models for Short-term wind speed forecasting. *Mater Today Proc*. 2021.
13. Yousuf MU, Al-Bahadly I, Avci E. A modified GM (1, 1) model to accurately predict wind speed. *Sustain Energy Technol Assess*. 2021;43:100905.
14. Yousuf MU, Al-Bahadly I, Avci E. Short-term wind speed forecasting based on hybrid MODWT-ARIMA-Markov model. *IEEE Access*. 2021;9:79695-79711.
15. Jahangir H, Golkar MA, Alhameli F, Mazouz A, Ahmadian A, Elkamel A. Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough ANN. *Sustain Energy Technol Assess*. 2020;38:100601.
16. Vanitha V, Sophia J, Resmi R, Raphel D. Artificial intelligence-based wind forecasting using variational mode decomposition. *Computational Intelligence*. 2021;37(3):1034-1046.
17. Wang H, Wang G, Li G, Peng J, Liu Y. Deep belief network based deterministic and probabilistic wind speed forecasting approach. *Appl Energy*. 2016;182:80-93.
18. Dhiman HS, Deb D. Machine intelligent and deep learning techniques for large training data in short-term wind speed and ramp event forecasting. *Int Trans Elect Energy Syst*. 2021;e12818.

19. Hu H, Wang L, Lv S-X. Forecasting energy consumption and wind power generation using deep echo state network. *Renewable Energy*. 2020;154:598-613.
20. Xiao L, Shao W, Jin F, Wu Z. A self-adaptive kernel extreme learning machine for short-term wind speed forecasting. *Appl Soft Comput*. 2021;99:106917.
21. Hu H, Wang L, Tao R. Wind speed forecasting based on variational mode decomposition and improved echo state network. *Renewable Energy*. 2021;164:729-751.
22. Altan A, Karasu S, Zio E. A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer. *Appl Soft Comput*. 2021;100:106996.
23. Saudi Arabia's first wind farm begins electricity production. *Al Jazeera*. Al Jazeera Media Network, 8th Aug 2021. Accessed August 12, 2021. https://www.aljazeera.com/news/2021/8/8/saudi-arabias-first-wind-farm-begins-electricity-production
24. Li S, Ma X, Yang C. A novel structure-adaptive intelligent grey forecasting model with full-order time power terms and its application. *Comput Ind Eng*. 2018;120:53-67.
25. Aasim SS, Mohapatra A. Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting. *Renewable Energy*. 2019;136:758-768.
26. Holt. *Forecasting Seasonals and Trends by Exponentially Weighted Averages (ONR Memorandum No. 52)*, Vol. 10. Carnegie Institute of Technology; 1957.
27. Winters PR. Forecasting sales by exponentially weighted moving averages. *Manage Sci*. 1960;6(3):324-342.
28. Brown RG. *Statistical Forecasting for Inventory Control*. McGraw/Hill; 1959.
29. Wang J, Zhang W, Wang J, Han T, Kong L. A novel hybrid approach for wind speed prediction. *Inf Sci*. 2014;273:304-318.
30. Huang X, Wang J, Huang B. Two novel hybrid linear and nonlinear models for wind speed forecasting. *Energy Convers Manage*. 2021;238:114162.
31. Niu M, Sun S, Wu J, Zhang Y. Short-term wind speed hybrid forecasting model based on bias correcting study and its application. *Math Probl Eng*. 2015;2015:1-13.
32. Yang H, Jiang Z, Lu H. A hybrid wind speed forecasting system based on a 'decomposition and ensemble' strategy and fuzzy time series. *Energies*. 2017;10(9):1422.
33. Zhang K, Qu Z, Dong Y, et al. Research on a combined model based on linear and nonlinear features-a case study of wind speed forecasting. *Renewable Energy*. 2019;130:814-830.
34. Kusiak A, Zhang Z. Short-horizon prediction of wind power: a data-driven approach. *IEEE Trans Energy Convers*. 2010;25(4):1112-1122.
35. Prema V, Rao KU. Time series decomposition model for accurate wind speed forecast. *Renew Wind Water Solar*. 2015;2(1):1-11.
36. Jiang P, Wang B, Li H, Lu H. Modeling for chaotic time series based on linear and nonlinear framework: application to wind speed forecasting. *Energy*. 2019;173:468-482.
37. Ferreira M, Santos A, Lucio P. Short-term forecast of wind speed through mathematical models. *Energy Rep*. 2019;5:1172-1184.
38. Camelo HN, Lucio PS, Leal Junior JBV, Carvalho PCM, Santos DG. Innovative hybrid models for forecasting time series applied in wind generation based on the combination of time series models with artificial neural networks. *Energy*. 2018;151:347-357.
39. Methods and formulas for double exponential smoothing. Accessed October 11, 2021. https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/double-exponential-smoothing/methods-and-formulas/methods-and-formulas/
40. Exponential smoothing. Accessed October 11, 2021. http://www.eviews.com/help/helpintro.html#page/content/series-Exponential_Smoothing.html#ww148060
41. Hyndman RJ, et al. Package 'forecast'. 2020. Accessed October 11, 2021. https://cran.r-project.org/web/packages/forecast/forecast.pdf
42. SAS Institute Inc. *Chapter 16: The FORECAST Procedure. SAS/ETS® 13.2 User's Guide*. SAS Institute Inc.; 2014.
43. Makridakis S, Hibon M. Exponential smoothing: the effect of initial values and loss functions on post-sample forecasting accuracy. *Int J Forecast*. 1991;7(3):317-330.
44. Gardner ES Jr. Exponential smoothing: the state of the art—Part II. *Int J Forecast*. 2006;22(4):637-666.
45. Tashman LJ. Out-of-sample tests of forecasting accuracy: an analysis and review. *Int J Forecast*. 2000;16(4):437-450.
46. Fildes R, Hibon M, Makridakis S, Meade N. Generalising about univariate forecasting methods: further empirical evidence. *Int J Forecast*. 1998;14(3):339-358.
47. Holt's forecasting model. In: Swamidass PM, ed. *Encyclopedia of Production and Manufacturing Management*. Springer US; 2000.
48. Gardner ES Jr, McKenzie E. Forecasting trends in time series. *Manage Sci*. 1985;31(10):1237-1246.
49. Hyndman R, Koehler AB, Ord JK, Snyder RD. *Forecasting with Exponential Smoothing: The State Space approach*. Springer Science & Business Media; 2008.
50. New Zealand Wind Energy Association. Accessed October 11, 2021. https://www.windenergy.org.nz/
51. Chappell PR. *The Climate and Weather of Manawatu-Wanganui*. NIWA Science and Technology; 2015. https://docs.niwa.co.nz/library/public/NIWAsts66.pdf
52. National Institute of Water and Atmospheric Research (NIWA). The National Climate Database. Accessed October 11, 2021. https://cliflo.niwa.co.nz/
53. Dixit P, Londhe S, Dandawate Y. Removing prediction lag in wave height forecasting using Neuro-Wavelet modeling technique. *Ocean Eng*. 2015;93:74-83.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Yousuf MU, Al-Bahadly I, Avci E. Wind speed prediction for small sample dataset using hybrid first-order accumulated generating operation-based double exponential smoothing model. *Energy Sci Eng*. 2022;00:1–14. doi:10.1002/ese3.1047

CHAPTER 6

STATISTICAL WIND SPEED FORECASTING MODELS FOR SMALL SAMPLE DATASETS: PROBLEMS, IMPROVEMENTS, AND PROSPECTS

This chapter contains content from the following article, republished with permission. In accordance with Elsevier's republishing policy, any full text that has been included is the unmodified accepted article. This means that while the content is identical to the published article, there may be stylistic differences.

M. U. Yousuf, I. Al-Bahadly, and E. Avci, " Statistical wind speed forecasting models for small sample datasets: Problems, Improvements, and prospects," *Energy Conversion and Management*, vol. 261, p. 115658, 2022.
DOI: https://doi.org/10.1016/j.enconman.2022.115658

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Statistical wind speed forecasting models for small sample datasets: Problems, Improvements, and prospects

Muhammad Uzair Yousuf [a,b,*], Ibrahim Al-Bahadly [a], Ebubekir Avci [a]

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

ARTICLE INFO

Keywords:
Statistical
Performance
Wind speed
Forecasting
Problems
Improvements

ABSTRACT

Wind speed forecasting models have seen significant development and growth in recent years. In particular, hybrid models have been emerging since the last decade. Hybrid models combine two or more techniques from several categories, with each model utilizing its distinct strengths. Mainly, data-driven models that include statistical and machine learning models are deployed in hybrid models for shorter forecasting time horizons (<6hrs). Literature studies show that machine learning models have gained enormous potential owing to their accuracy and robustness. On the other hand, only a handful of studies are available on the performance enhancement of statistical models, even though hybrid models are incomplete without statistical models. To address the knowledge gap, this paper addressed the problems of traditional statistical models while enhancing prediction accuracy. Initially, the multi-step ahead wind speed forecasting performance of eight commonly used statistical approaches is evaluated using four different case studies and three rolling windows. The reasons for erroneous wind speed forecasts are discussed in detail. Next, four enhanced models were considered while addressing the shortcomings of conventional methods. In addition, four machine learning models are also analyzed for comparison. Moreover, the outcomes of the comparisons are discussed, explaining the higher prediction accuracy of improved models. A Global Performance Indicator (GPI) is used to rank the investigated wind speed forecasting models. Results showed that exponential smoothing models have a greater GPI in most cases, whereas Markov Chain models are among the poorest. The grey models are more suitable for smaller samples of data. In machine learning models, Support Vector Machine (SVM) has proven to be the best choice. Overall, the improved models show between 4% and 28% higher accuracy than their counter traditional models. Lastly, the future directions are highlighted that need subsequent research to further improve forecasting performance.

1. Introduction

Wind energy plays an essential role in addressing environmental and resource challenges. It is a widely available and environmentally friendly renewable energy source. According to the BP Statistical Review of World Energy 2021 [1], the highest contribution to renewable energy growth came from wind in 2020. In addition, the wind capacity is expected to expand further, as onshore wind power costs have dropped by around 40% in the last five years [1]. Despite being one of the fast-expanding technologies, the fundamental drawback is the stochastic nature of wind. Wind speed intermittency has a significant influence on power system stability. As a result, several studies have been carried out to accurately forecast the wind speed that would support power generation planning while reducing the need for extra reserve capacity. According to a study conducted for the California Independent System Operator (CASIO), the annual total cost savings from improved short-term wind forecasting are estimated to range from \$5.05 million to \$146 million [2]. Similarly, for the Irish electricity system with 33% wind penetration, it is concluded that the total system costs reduced between 0.5% and 1.6%, with a decrease in wind forecast mean absolute error from 8% to 4% [3]. Xu et al. [4] evaluated that the improper selection of the forecasting model resulted in additional 57 MW reserves for 50% wind power generation of the total installed capacity. Also, Zhang et al. [5] concluded that if dispatchers do not take wind forecasting errors into account when identifying reserve plans, the probability of insufficient reserve capacity in a 15-minute period would be unacceptably high.

* Corresponding author at: Department of Mechanical and Electrical Engineering, Massey University, Palmerston North 4442, New Zealand.
E-mail address: M.U.Yousuf@massey.ac.nz (M.U. Yousuf).

https://doi.org/10.1016/j.enconman.2022.115658
Received 13 February 2022; Received in revised form 6 April 2022; Accepted 19 April 2022
Available online 27 April 2022

Nomenclature

Abbreviations

ACF	Autocorrelation Function
AGO	Accumulated Generating Operation
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BPNN	Back Propagation Neural Network
CEEMDAN	Complementary Ensemble Empirical Mode Decomposition with Adaptive Noise
CFD	Computational Fluid Dynamics
DES	Double Exponential Smoothing
DT	Decision Trees
ELM	Extreme Learning Machine
EMD	Empirical Mode Decomposition
ESMAP	Energy Sector Management Assistance Program
ESN	Echo State Network
FFOTR	Fuzzy First-Order Transition Rules
GM	Grey Model
GPI	Global performance Indicator
GPR	Gaussian Process Regression
IGA	Improved Genetic Algorithm
LSSVM	Least Squares Support Vector Machine
LSTM	Long Short-Term Memory
MC	Markov Chain
modDES	Modified Double Exponential Smoothing
modGM	Modified Grey Model
modMC	Modified Markov Chain
MSE	Mean Square Error
NASA	National Aeronautics and Space Administration
NIWA	National Institute of Water and Atmospheric Research
nMAE	normalized Mean Absolute Error
NN	Neural Networks
nRMSE	normalized Root Mean Square Error
PCHIP	Piecewise Cubic Hermite Interpolating Polynomials
PRBF	Principal component analysis- Radial Basis kernel Function
PSO	Particle Swarm Optimization
SSE	Sum of Squares Error
SVM	Support Vector Machine
VMD	Variational Mode Decomposition
WDD	Wavelet Domain Denoising
WFGM	Weighted Fractional Grey Model
WPD	Wavelet Packet Decomposition

Roman

a	developing coefficient
b	grey actuating quantity
c	constant term
c_v	scale factor
d	degree of differencing
f	fractional order-AGO
k	steps ahead forecast
$k(x_i, x)$	kernel function
k_v	shape factor
p	order of autoregressive
q	order of moving average
t	time
v	measured wind speed
$\widehat{v}$	predicted wind speed
w	weighting parameter
B	backward difference operator
C	post-error ratio
E'	centered RMSE
I	scaled value of an indicator
L	level
N	rolling window sizes
N_f	frequency of transition
N_t	the theoretical minimum value for rolling window
P	probability
R	correlation coefficient
S	state-space
$S^{(1)}$	first-order exponential smoothing values
$S^{(2)}$	second-order exponential smoothing values
T	Trend
Z	background value array

Greek

α	weighting factor
β	smoothing constant for the level
$\widehat{\beta}_o$	y-intercept
$\widehat{\beta}_1$	Slope
γ	smoothing constant for the trend
δ	Lagrange multiplier
ε	error term
ξ	weight factor in GPI (-1 or 1)
θ	moving average parameter coefficient
μ	mean
σ	standard deviation
φ	mapping function
ϕ	autoregressive parameter coefficient
Π	state probability vector

Existing forecasting methods can be classified into five categories: persistence model, physical model, statistical model, Artificial Intelligence/Machine Learning (AI/ML) model, and Hybrid model. In the persistence model, the immediate-future wind speed ($v_{t+\Delta t}$) is assumed to be the same as the present time (v_t). This model is termed a naïve predictor. However, it cannot be upgraded to enhance further.

Physical models are established on meteorological and geographic data, such as temperature, pressure, terrain structure, and obstacles [6]. There are two types of physical methods: diagnostic models and computational fluid dynamics (CFD) models [7]. Diagnostic models employ boundary layer parameterizations and are suitable for flow over flat terrain. CFD models simulate wind flow fields dynamically and are useful for flow over complicated terrain. Physical models use initial conditions to solve complex numerical systems to provide global and regional forecasts [8] without requiring historical data [9]. However, detailed information on surface roughness and wind farm features is essential. As a result, these models need a significant amount of work to set up.

Statistical models are data-driven models that use historical wind speed data to generate forecasts. Grey models [10,11], Markov Chain [12,13], exponential smoothing [14,15], ARMA [16,17], and ARIMA [18,19] models appear to be the top among the several approaches that have been studied and evaluated.

Artificial Intelligence/ Machine learning (AI/ML) models are also data-driven models. However, different from statistical models, machine learning algorithms are more capable of addressing the nonlinearity of wind speed. Neural Networks (NN) [20], Support Vector Machine (SVM) [21,22], Decision Trees (DT) [23,24], Gaussian Process Regression (GPR) [25,26], and Extreme Learning Machine (ELM) [27] are some of the approaches explored within machine learning models.

Hybrid models combine two or more techniques from several categories, with each model utilizing its distinct strengths [28]. A single

Table 1
Examples of studies that integrated statistical and machine learning algorithms.

Statistical Model	Hybrid Models	Purpose
ARMA	WDD-WPD-ARMA-EMD-ELM [43] VMD-PRBF-ARMA-E [44] ARMA-ESN [45]	To capture the linear patterns
ARIMA	ELM-CEEMDAN-ARIMA [46] EMD-LSTM-ARIMA [47] ARIMA–NN–FFOTR [48]	Applied as an error correction model to improve the accuracy of the prediction model
Exponential Smoothing	CEEMDAN-CC-Holt-GRNN [49] DES-PSO-BPNN-Elman [50] CEEMD-ES-CS-BPNN [51]	Suitable option for uncertain systems with poor information
GM (1,1)	Grey-ELM [52] PSO-SVR and Grey Combination Model [53] SVM-GM [54]	
Markov Chain	ANN-Markov [55] CEEMD-IGA-FNN-Markov [56] LSSVM-Markov [57]	Less operation data is needed

prediction model cannot fully represent the inherent complex relationship of wind speed such as nonlinearity and randomness w.r.t. time. Hence, a hybrid model can compensate for some of the deficiencies of distinct models while adding to the computational complexity. Weighted models [29], feature selection models [30], decomposition models [31], and error processing models [32] are the four subcategories of hybrid models. The details of each subcategory are readily available in review articles [7,33].

1.1. *Objectives and motivation*

It is observed from the literature review that AI/ML approaches have gained more attraction in recent years in the field of wind speed forecasting and have occupied "irreplaceable dominance" [34,35]. The success of machine learning models is based on algorithms capable of learning by trial and error and increasing their performance over time, rather than the traditional programming of step-by-step coding instructions based on logic and if-then rules [36]. For this reason, most of the current research is mainly focused on improving the performance of AI/ML models. As an example, the variants of support vector machines (SVM) that are applied for wind speed forecasting include linear epsilon-insensitive SVM (ε-SVM), reduced SVM (RSVM), least-square SVM (LSSVM), twin SVM (TSVM), ε-twin SVM (ε-TSVM), primal least square SVM (PLSSVM), primal least square twin SVM (PLSTSVM), iterative Lagrangian twin parametric insensitive SVM (ILTPISVM), and proximal SVM (PSVM) [37–41]. Such vast improvements and advancements have not been observed in most statistical models. It is important to note that the statistical models are not only easy to build and fast to calculate, but they are also robust and less prone to overfitting than machine learning approaches. In that respect, statistical models are valuable not just on their own but also as part of hybrid models that incorporates more advanced methodologies [42]. Keeping in view the advantages of statistical models, numerous studies integrated statistical and machine learning algorithms together for wind speed forecasting. A few of them are discussed in Table 1. Hence, accuracy improvements of statistical models are as significant as machine learning models. To the best of the authors' knowledge, only a few works reported in this domain, and a detailed study is still required to address the problems of traditional statistical models while enhancing the prediction performance. Therefore, this research has three primary objectives to fill this gap.

- The first objective is to identify the prediction problems of statistical models in terms of forecasting results acceptability. Four distinct case studies are considered to evaluate the performance of eight frequently used statistical methods. A total of n-100 observations, with three rolling windows, were used to develop the prediction models. One to three steps ahead forecast performance was assessed by comparing them to actual values not employed in the forecasting model development.
- The second purpose is to explore four improved models while addressing the limitations of existing approaches. In addition, four common AI/ML models are also considered for comparison.
- The final goal is to describe the results of the comparisons and try to explain why updated models have better prediction performance. Further issues that restrict the utilization of statistical forecasting models are also mentioned as potential future areas.

The rest of the paper is organized as follows. In section 2, case studies are discussed initially. Next, the theoretical framework of traditional statistical models is presented while highlighting the problems of conventional models. Also, the improved models and machine learning models are introduced in this section. Next, the results and discussions of the forecasting models are presented in Section 3, followed by future directions. Finally, conclusions are presented in Section 4.

2. Materials and methods

2.1. *Case studies*

Keeping in view the objectives of the study, wind speed data for diversified regions were selected to compare the effectiveness of the forecasting models. Therefore, sites from New Zealand, Pakistan, Zambia, and Spain were chosen, as shown in Fig. 1. The descriptive data of the selected sites are provided in Table 2. The raw data is retrieved from ESMAP (Energy Sector Management Assistance Program) [58], NIWA (National Institute of Water and Atmospheric Research) [59], and Sotavento wind farm databases [60].

The preliminary step is to post-process the collected raw data to identify the anomalies. Data repositories such as ESMAP and NASA show that inaccuracies may exist in the data in terms of no value (NaN), missing value (no data recorded, 999, −999), erroneous value (all zeros), and out of range value (outliers). The first three anomalies can easily be recognized in the datasets. To identify the outliers, two commonly applied approaches are median absolute deviation [61,62] and the boxplot rule [63,64]. Jeong et al. [65] applied both techniques to a hydrogeological dataset and concluded that both methods worked well at a 5% outlier level. Because the outlier detection method is not the primary objective of the study, the boxplot rule is selected to identify the outliers.

To eliminate anomalies, multiple approaches are applied by the researchers. Hence, there is no standard procedure for processing the outlier. Schlechtingen and Santos [66] are in favor of neglecting the missing input and target values from the large dataset. As we already have a small sample dataset, therefore, deletion of a certain row is not a solution. Wang and Xiong [67] applied cubic spline interpolation, Zhou et al. [63] applied a modified boxplot method, Ouyang et al. [68] applied simple interpolation, and Qiu et al. [61] applied a moving average filter to estimate the outlier data points. Based on available literature [69–71], it is observed that Piecewise Cubic Hermite Interpolating Polynomials (PCHIP) is better than piecewise linear functions and cubic spline and is used in solar [72] and wind energy [73]. Furthermore, PCHIP offers a smooth slope transition between data points [74] and is more robust [69]. Based on these discussions, PCHIP is selected to synthesize the anomalies.

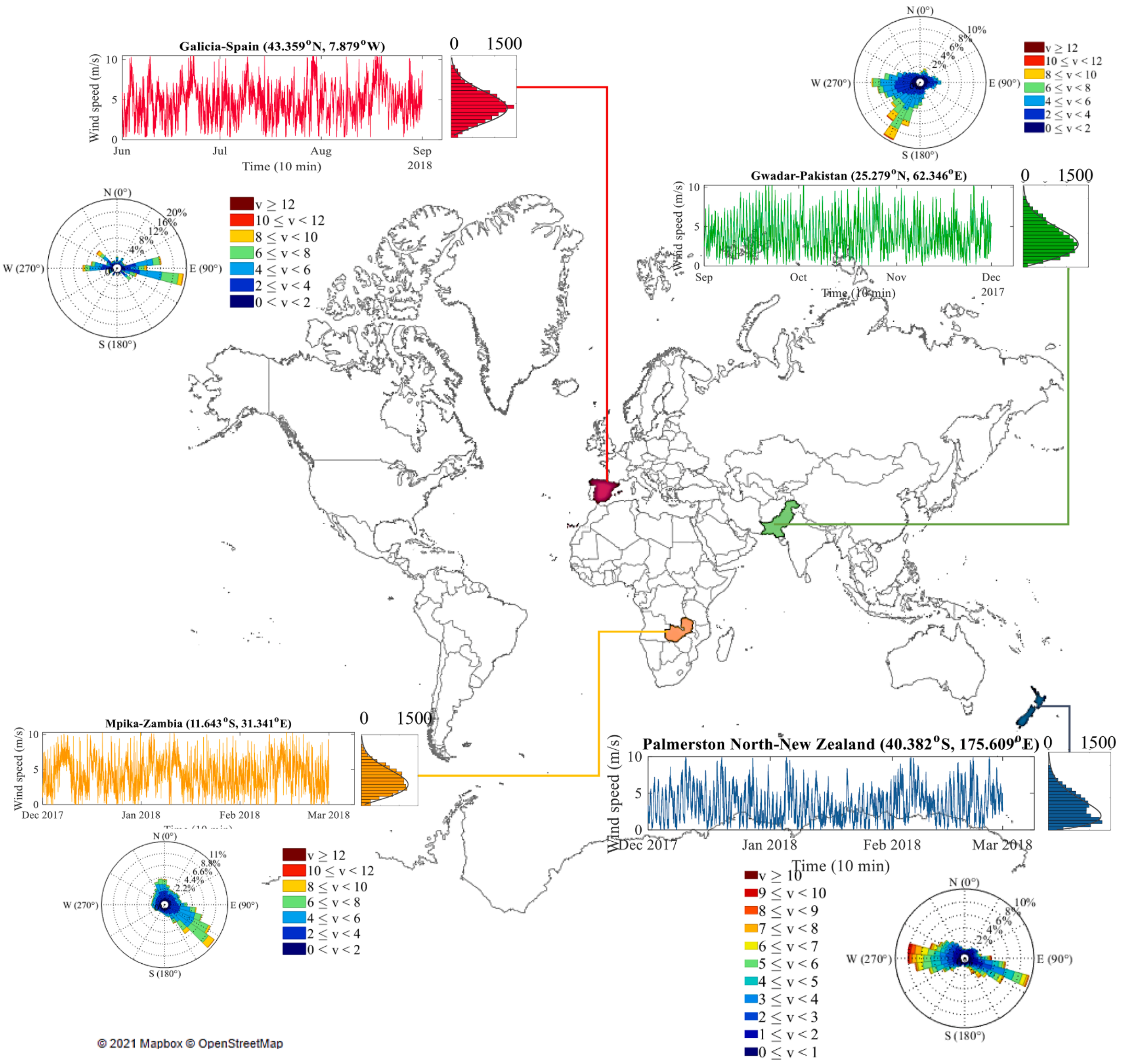


Fig. 1. Sites from New Zealand, Pakistan, Zambia, and Spain are selected as case studies.

2.2. Traditional statistical models

Literature studies show that the most commonly applied statistical models are Auto-Regressive Integrated Moving Average (ARIMA), Double Exponential Smoothing (DES), Grey models, and Markov Chains. Considering the conventional and modified versions of these models, eight statistical models are included in this study (see Fig. 2), whose brief description is provided next.

2.2.1. Auto-Regressive integrated moving average

The most often used statistical models are ARIMA. The generalized non-seasonal model structure is ARIMA (p, d, q), where p and q shows the order of autoregressive (AR) and moving average (MA) parts. For stationary time series, the model is ARMA (p, q). For non-stationary time series, the stationarity condition is achieved by taking d degree of differencing as shown in (1).

$$B^d v_t = v_{t-d} \tag{1}$$

Hence, with (p, d, q), the linear equation is expressed as,

$$v_t = c + \left(\sum_{i=1}^{p} \phi_i v_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \right) \tag{2}$$

where ϕ_i and θ_j are the i^{th} autoregressive and j^{th} moving average parameters coefficients respectively. The model is a three-step iterative process as follows.

Step 1. Model identification: First, the unit root test (here, the Dickey and Fuller tests) is conducted to determine the stationarity. Next, the differencing is applied, if required. Once the stationarity is achieved, the plots of sample and partial Auto Correlation Function (ACF) are utilized to identify the orders of autoregressive (AR) and moving average (MA).

Step 2. Parameter Estimation: The parameters are then estimated using the maximum likelihood method.

Step 3. Diagnostic Checking: To assure the performance of the model, it is necessary to confirm the model adequacy conditions. The residual values will behave like a white noise process if the specified model is acceptable.

2.2.2. Double exponential smoothing model

Exponential Smoothing (ES) techniques are also frequent in forecasting applications. In contrast with the moving average methods, ES

Table 2
Descriptive data of the selected case studies.

Description	Case Study			
	Palmerston North	Gwadar	Mpika	Galicia
Country	New Zealand	Pakistan	Zambia	Spain
Latitude	40.382 °S	25.279 °N	11.643 °S	43.359 °N
Longitude	175.609 °E	62.346 °E	31.341 °E	7.879 °W
Observing authority	NIWA [59]	ESMAP [58]	ESMAP [58]	Sotavento [60]
Data Duration	01/12/2017 –28/02/2018	01/09/ 2017 –30/ 11/2017	01/12/ 2017 –28/ 02/2018	01/06/2018 –31/08/2018
Data resolution (min)	10	10	10	10
Maximum – v_{max} (m/s)	9.90	10.26	10.31	10.49
Mean – $\bar{v}$ (m/s)	3.52	4.14	4.35	4.76
Standard deviation – σ_v (m/s)	2.15	2.14	2.03	2.09
*Scale factor – c_v (m/s)	3.92	4.67	4.85	5.36
*Shape factor –k_v	1.64	2.03	2.12	2.42

$$*k_v = \left[\frac{\sum_{i=1}^{n} v_i^{k_v} ln(v_i)}{\sum_{i=1}^{n} v_i^{k_v}} - \frac{\sum_{i=1}^{n} ln(v_i)}{n}\right]^{-1}, c_v = \left[\frac{\sum_{i=1}^{n} v_i^{k_v}}{n}\right]^{\frac{1}{k_v}}$$

assigns exponentially decreasing weights to the older values to predict the future state.

Double Exponential Smoothing (DES) is used when data shows a trend. Brown [75] and Holt [76] are two widely used DES models. Brown's approach is an expansion of the simple ES model, and it is defined as,

$$S_t^{(1)} = \beta v_t + (1-\beta) S_{t-1}^{(1)} \tag{3}$$

$$S_t^{(2)} = \beta S_t^{(1)} + (1-\beta) S_{t-1}^{(2)} \tag{4}$$

where $S_t^{(1)}$ and $S_t^{(2)}$ are first and second-order exponential smoothing values.Holt DES model has two components: level. 'L_t' (smoothed estimates of the value) and trend 'T_t' (smoothed estimates of the average growth). The two components can be calculated as:

$$L_t = \beta v_t + (1-\beta)(L_{t-1} + T_{t-1}) \tag{5}$$

$$T_t = \gamma(L_t + L_{t-1}) + (1-\gamma) T_{t-1} \tag{6}$$

where β and γ are smoothing constants that decide the weights given to the observations. The more the value closer to 1, the higher the weight given to the newer observation. Forecasting through DES is summarized in three steps.

Step 1. Initial value calculation: As seen in equations (3)-(6), initial values are needed in the recursive calculations. Literature studies suggested that the least square estimate is the most common method to calculate the initial values [77]. A linear regression model is fit to available wind speed time series data as $\hat{v}_o = \hat{\beta}_{o,T} + \hat{\beta}_{1,T} t$.

For Brown's Model [78]:

$$S_0^{(1)} = \hat{\beta}_{o,T} - \left(\frac{1-\beta}{\beta}\right)\hat{\beta}_{1,T} \tag{7}$$

$$S_0^{(2)} = \hat{\beta}_{o,T} - 2\left(\frac{1-\beta}{\beta}\right)\hat{\beta}_{1,T} \tag{8}$$

For Holt's Model [79]:

$$L_o = \hat{\beta}_{o,T} \tag{9}$$

$$T_o = \hat{\beta}_{1,T} \tag{10}$$

Step 2. Optimal smoothing constants: Next is to obtain the optimal values of smoothing constants (β, γ) that minimize the sum or mean of squared error of predictions.

$$SSE = \sum_{t=1}^{n} (v_t - \hat{v}_t)^2 \tag{11}$$

$$MSE = \frac{SSE}{n} \tag{12}$$

Step 3. Forecasting Equation: The last step is to generate *k*-step ahead forecast as:

For Brown's Model:

$$\hat{v}_{t+k} = a_t + k b_t \tag{13}$$

where $a_t = 2S_t^{(1)} - S_t^{(2)}$ and.$b_t = \frac{1}{1-\beta}\left(S_t^{(1)} - S_t^{(2)}\right)$

For Holt's Model:

$$\hat{v}_{t+k} = L_t + k T_t \tag{14}$$

2.2.3. Markov Chain model

The Markov Chain (MC) is a stochastic process that sequentially travels from one state to the next in state space. MCs have the benefit of being able to represent wind time dependency features. The following five steps outline the forecasting process using MC.

Step 1. State categorization: The categorization of states is the first step of the MC model. Instead of manually giving the state-space (S), a self-adaptive method is applied in this study. At each time step, the state-space is defined as $\{0, ceil(v_{min}), sort(v, ascend), ceil(v_{max})\}$ with no repetitive values.

Step 2. State Transition Matrix: As per definition, a first-order MC satisfies:

$$P\{X_{t+1} = j | X_t = i_t, X_{t-1} = i_{t-1}, \cdots X_o = i_o\} = P\{X_{t+1} = j | X_t = i_t\} = p_{ij} \tag{15}$$

Equation (15) shows that the probability of the future state (X_{t+1}) depends only on the current state (X_t). Therefore, $\{X_{t+1} = j | X_t = i\}$ for all t and all i and j.

A matrix that describes the probability (P) of the observed frequency of transition (N_f) is termed a transition matrix. Thus,

$$N_f = \begin{bmatrix} n_{11} \cdots n_{1s} \\ \vdots \ddots \vdots \\ n_{s1} \cdots n_{ss} \end{bmatrix}, P = \begin{bmatrix} p_{11} \cdots p_{1s} \\ \vdots \ddots \vdots \\ p_{s1} \cdots p_{ss} \end{bmatrix} \tag{16}$$

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

Step 3. Confirmation of Ergodic Properties: The next step is to verify the ergodic properties. If MC is irreducible and its states are positive recurrent, and aperiodic, it is ergodic. For the ergodic MC, the limit $\pi_i = \lim_{n\to\infty} P\{X_n = i\} = \frac{1}{\mu_i}$ exists and is independent of the initial distribution, where μ_i determines the average time it takes to return to a state i [80].

Step 4. State transition probability vector: Based on the initial state probability vector (Π_0) and state transition matrix (P), the state transition probability of the future state is determined as $\Pi_k = \Pi_0 \times P^k$. Except for the element corresponding to the present state at time t, all the elements of Π_0 are zero.

Step 5. Forecasting Equation: Finally, one step ahead wind speed is computed as.

$$\hat{v}_{t+1} = \sum_{i=1}^{s} \Pi_i \bar{S}_i \tag{17}$$

where $(\bar{S})$ is an average wind speed of the specific state.

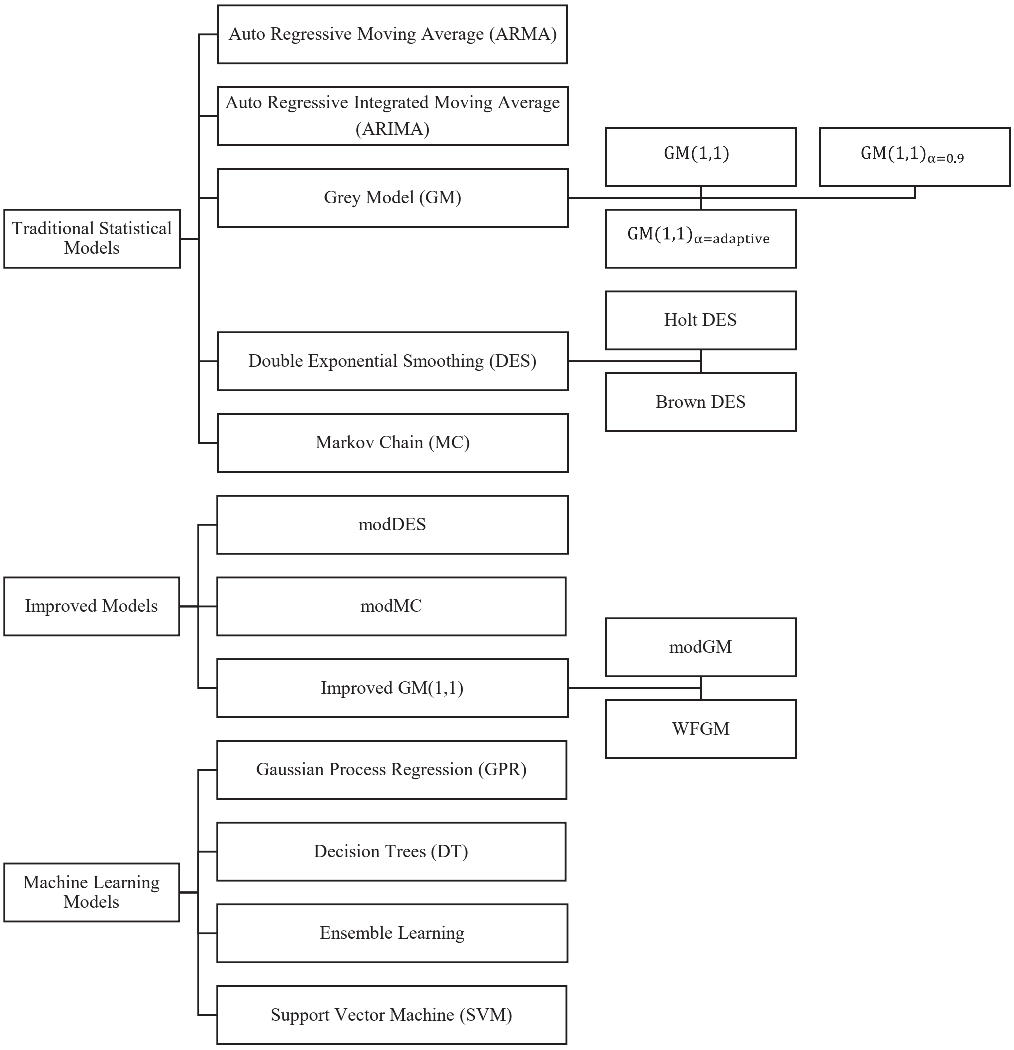


Fig. 2. Taxonomy of the selected 16 models from three categories.

2.2.4. Grey model [GM (1,1)]

The Grey system theory was devised by Deng [81], which is the focus of the grey prediction model. Following are six steps to forecast wind speed using GM (1,1).

Step 1. First-order Accumulated Generating Operation (1-AGO): For a given wind speed time series $V^{(0)} = \{v^{(0)}(1), v^{(0)}(2), \cdots, v^{(0)}(n)\}$, perform 1-AGO as:

$$V^{(1)} = \{v^{(1)}(1), v^{(1)}(2), \cdots, v^{(1)}(n)\} \tag{18}$$

where $v^{(1)}(k) = \sum_{i=1}^{k} v^{(0)}(i),\ k = 1, 2, 3, \cdots, n$

Step 2. Background value array calculation: Next is to establish the background value array $Z^{(1)}$ same as. $V^{(1)}$

$$Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \cdots, z^{(1)}(n)\} \tag{19}$$

where $z^{(1)}(k) = \alpha v^{(1)}(k) + (1-\alpha)v^{(1)}(k-1),\ k = 2, 3, \cdots, n.$ α is a weighting factor. The value of α defines three different models.

Traditional model: $\alpha = 0.5$

Best selection model: $\alpha = 0.9$

Adaptive model: the value of $0.1 < \alpha < 0.9$ corresponds to the least MSE.

Step 3. Equation parameter calculation: The GM (1,1) model is stated as a difference equation in its original form as (20),

$$v^{(0)}(k) + az^{(1)}(k) = b \tag{20}$$

Table 3
Effect of developing coefficient on prediction length.

Developing Coefficient $\lvert a \rvert$	Prediction Length
$\lvert a \rvert \leqslant 0.3$	Medium to long-term predictions
$0.3 < \lvert a \rvert \leqslant 0.8$	Short to medium term predictions
$0.8 < \lvert a \rvert \leqslant 1$	Modified the traditional GM(1,1) model
$\lvert a \rvert \rangle 1$	GM(1,1) is not suitable for predictions

where a and b are the developing coefficient and grey actuating quantity, respectively. The vector of parameters in the expression $\widehat{a} = [a, b]^T$ is calculated using the least square approach, which meets the following conditions:

$$\widehat{a} = (B^T B)^{-1} B^T Y \tag{21}$$

where $Y = \begin{bmatrix} v^{(0)}(2) \\ v^{(0)}(3) \\ \vdots \\ v^{(0)}(n) \end{bmatrix}$, $B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$ and. $v^{(0)}(1) = v^{(1)}(1)$

Table 3 shows the suitability of the prediction w.r.t. developing coefficient and prediction length. In Table 3, short-, medium- and long-term predictions is considered as two, five, and ten steps ahead forecasts, respectively [82].

Step 4. Whitenization equation: The whitenization equation of (20) is a

Table 4
Accuracy levels for posterior error detection measures.

Accuracy Level	Small error probability (P)	Post-error ratio (C)
I	$0.95 \leqslant P$	$C \leqslant 0.35$
II	$0.80 \leqslant P < 0.95$	$0.35 < C \leqslant 0.50$
III	$0.70 \leqslant P < 0.80$	$0.50 < C \leqslant 0.65$
IV	$0.60 \leqslant P < 0.70$	$0.65 < C \leqslant 0.80$

differential equation $\frac{dv^{(1)}}{dt} + av^{(1)} = b$ whose time response is.

$$\hat{v}^{(1)}(t) = \left(v^{(0)}(1) - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a} \tag{22}$$

Hence equation (22) is a solution of equation (20).

Step 5. Inverse 1-AGO: Final forecasting can be achieved by obtaining inverse 1-AGO as in (23).

$$\hat{V}^{(0)} = \begin{cases} \hat{v}^{(0)}(1) = v^{(0)}(1) \\ \hat{v}^{(0)}(k) = \hat{v}^{(1)}(k) - \hat{v}^{(1)}(k-1) = (1 - e^{a})\left(v^{(0)}(1) - \frac{b}{a}\right)e^{-a(k-1)} \end{cases} \tag{23}$$

Step 6. Diagnostic Checking: To ensure the appropriateness of the grey model, accuracy testing is recommended. As mentioned in literature studies, the posterior error detection method, including small error probability (P) and post-error ratio (C), is employed to assess the accuracy of grey models. Based on standard deviations of original time series and error sequence (σ_1, σ_2), the formulation for both the measures are given in equations (24)-(25) and the acceptable levels in Table 4. Only if the model passes both accuracy tests can the forecast be considered valid.

$$p = P\{|\varepsilon(k) - \bar{\varepsilon}|\langle 0.6745\sigma_1\} \tag{24}$$

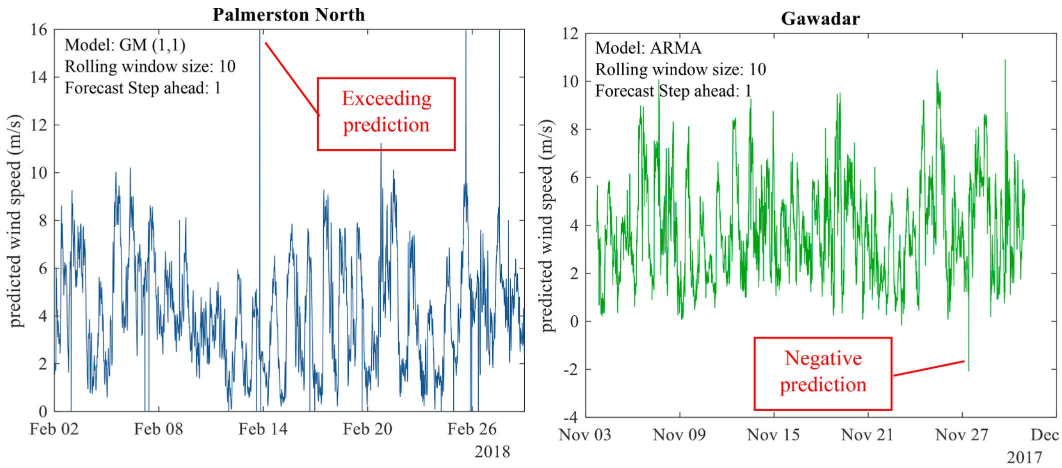


Fig. 3. Examples of erroneous predictions: exceeding and negative.

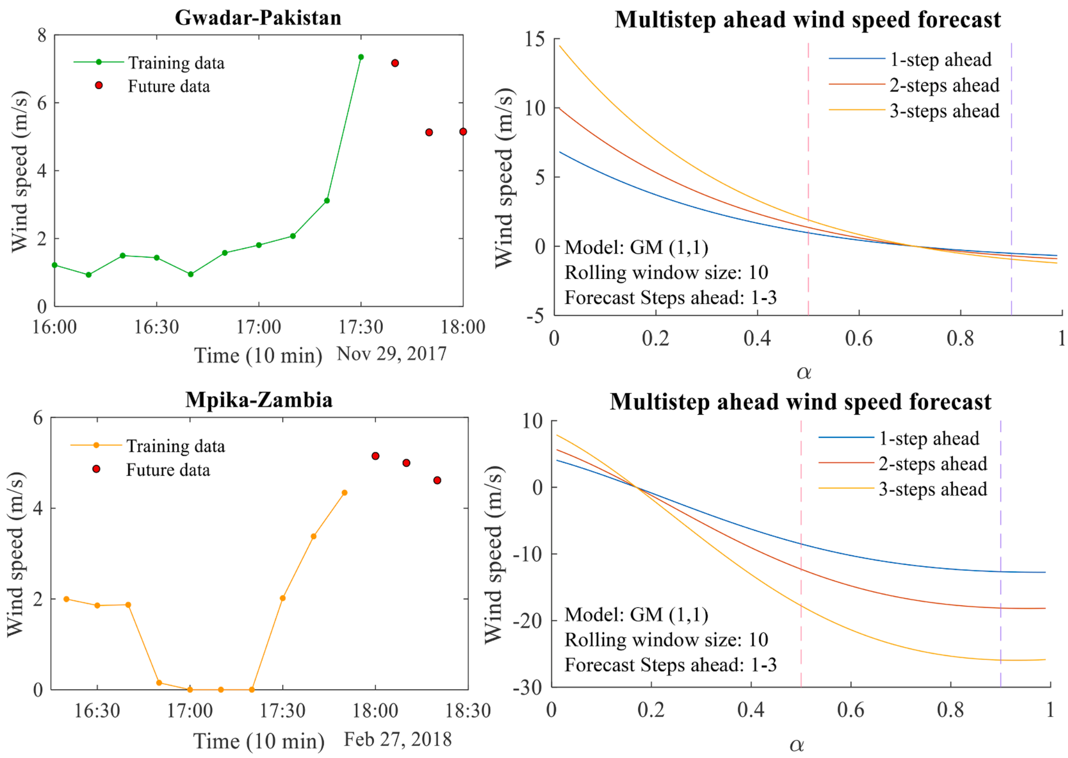


Fig. 4. The traditional value of $\alpha = 0.5$ is not always the best selection, and hence an optimal parameter is required.

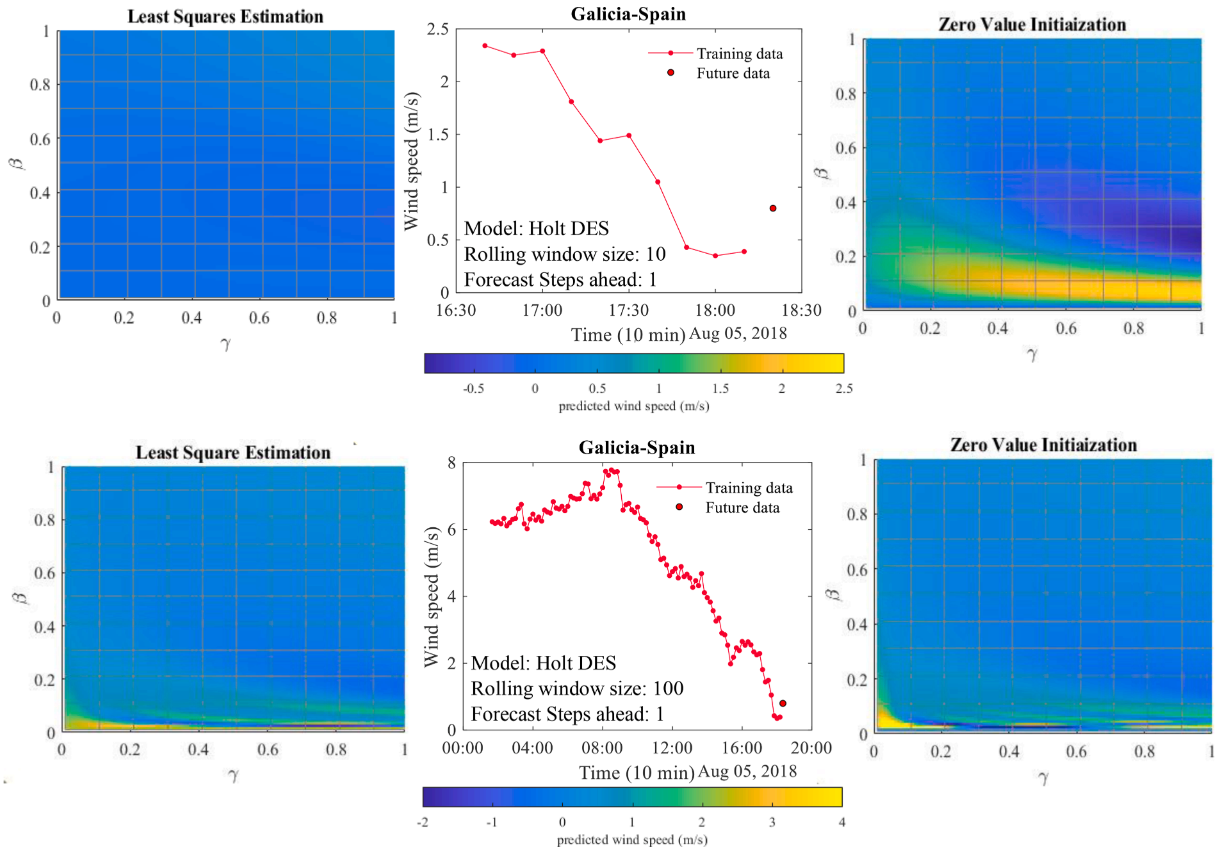


Fig. 5. Effect of initialization method on DES model.

$$C = \frac{\sigma_2}{\sigma_1} \tag{25}$$

where $\sigma_1 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(v^{(0)}(k) - \overline{v})^2}$, $\overline{v} = \frac{1}{n}\sum_{k=1}^{n} v^{(0)}(k)$, $\sigma_2 = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(\varepsilon(k) - \overline{\varepsilon})^2}$,

$$\overline{\varepsilon} = \frac{1}{n}\sum_{k=1}^{n}\varepsilon(k)$$

2.3. Problems associated with traditional models

In this section, problems associated with the traditional models are defined w.r.t. the case studies described in the previous section. Rolling window sizes (N) of 10, 50, and 100 are considered for multistep ahead wind speed forecast in the present study.

The major problem with all the considered models except Markov Chain is erroneous wind speed forecast, as shown in Fig. 3. For the Markov chain, the minimum and maximum state boundaries are always defined as zero and v_{max} (See Step 1. State categorization). Therefore, erroneous wind speed prediction is not possible.

Two types of erroneous predictions are observed: exceeding and negative. Exceeding predictions are the unreasonable wind speeds that differ considerably from the rest of the observations, whereas negative predictions are impossible values.

There might be three possible reasons for erroneous forecasts: uncalibrated model parameters, incorrect selection of initial conditions, and truncation error.

Uncalibrated parameters are the most common reason for erroneous predictions. As shown in equations (2), (5), (6), and (19), (p, d, q), (β, γ) and (α) are the parameters for ARIMA, exponential smoothing, and GM (1,1) models.

From Fig. 4, it is observed that the traditional value of $\alpha = 0.5$ for GM (1,1) is not the best selection in every case. Therefore, an optimal parameter is required. However, special attention is still required when applying this approach. In general, researchers used software packages to identify the optimal parameters. It includes Minitab for the DES model [79] and auto.arima for ARIMA models [83]. It is possible that the parameters identified by the software based on the least error still predict negative wind speed. As an example, Holt's function of forecast package [83] and Minitab is identifying parameters of Holt DES model as $(0.001, 0.001)$ and $(0.18910.1314)$ respectively for a time series fragment of $v = \{2.1, 2.3, 1.7, 1.8, 1.5, 1.3, 0.6, 0.4, 0.6, 0.2\}$. However, in these cases, the predicted wind speed is -0.0632 m/s and -0.0369 m/s. Therefore, a positivity constraint is required for such models by applying Box-Cox transformation with $\lambda = 0$ or identifying multiple combinations of parameters for prediction. As mentioned in [14], if a particular combination predicts a negative wind speed, the following combination will be applied for prediction. Similarly, parsimony is very important for any model. In the case of ARIMA, simpler order must be used in contrast with high order models. Otherwise, it may be possible that autoregressive polynomial becomes unstable and moving average polynomial becomes non-invertible for small sample size dataset.

The second possible reason that is mainly related to DES models is the inappropriate initial value selection. As seen in equations (3)-(6), the initial value is needed in the recursive calculations of the DES model. Hence, inappropriate selection of initial value results in erroneous wind speed. For example, Fig. 5 shows two different selections of initial value: least square estimation (traditional approach) and zero value initialization. It is evident from Fig. 5 that the zero value initialization method is a better approach than the least square estimate for the given time series fragment. Other methods include, but are not limited to, convenient initial values, backcasting, first data point, and zero values. One may note that the initialization method has very low relevance to the

Table 5
Shortcomings of considered statistical models and suggested improvements.

Models	Parameters	Shortcomings	Improvements
ARIMA	p, d, q	-Chances of erroneous wind speed -Unstable autoregressive polynomial -Invertible moving average polynomial	-Considering low order autoregressive and moving average polynomial -Imposing positivity constraints using Box-Cox transformation with $\lambda = 0$
DES	β, γ	-Chances of erroneous wind speed -Failure of traditional constraint	-Modifying the initialization method -Applying the admissible constraint -Considering larger rolling window
Markov Chain	S, P_{ij}	-Poor state categorization -Failure of the ergodic theorem	- Considering adjusting the rolling window - Applying adaptive state boundary selection algorithm
GM (1,1)	α	-Singularity Phenomena -Failure of posterior error detection -Failure of traditional constraint -Developing Coefficient is out of range	-Applying L'Hopital's rule -Considering adjusting the moving window -Apply remnant model for residual time series

prediction for larger window sizes. Therefore, another solution is to increase the rolling window, as shown in Fig. 5.

The third cause of erroneous prediction is truncation error, mainly related to the GM (1,1) model. As shown in Fig. 3, GM(1,1) is predicting $16m/s$ for a time series fragment of $v = \{2.8, 2.5, 2.6, 3.1, 2.7, 2.6, 3.2, 2.9, 2.3, 2.7\}$. The problem of exceeding prediction is caused by singular phenomena that arise mainly because of the developing coefficient a. As seen in equation (22), the developing coefficient a is in the denominator. Therefore, a smaller value of a results in a larger value of the factor $\left[v^{(0)}(1) - \frac{b}{a}\right]$. Hence, due to truncation error, the factor a shows an extremely small value (for the considered time series fragment, $a = 2.77 \times 10^{-17}$). It causes the factor $\left[v^{(0)}(1) - \frac{b}{a}\right]$ approaches to be extremely large, which results in exceeding prediction. Although high accuracy variants of GM(1,1) model such as the Fractional Accumulated Grey Model (FAGM), Heuristic Fuzzy time series Grey Model (HFGM), and Data Grouping Grey Model (DGGM) is proposed in literature [84,85], but such issue has not been discussed in these models. As defined in [67], one solution is to apply L'Hopital's rule to overcome the singular phenomenon.

Besides erroneous predictions, the second major problem is how much data is required to fulfill the necessary conditions. Such conditions include the ergodic theorem for the Markov Chain model, the traditional constraint of $0 < (\beta, \gamma) \langle 1$ for DES model and posterior check for GM (1,1). According to some research articles, a forecast can only be achievable or reliable if the defined conditions are fulfilled. As considered in [86–88], the model can be used for prediction only if the small probability error is greater than 0.70 and the post-error ratio is less than 0.65. Similarly, Liu et al. [82] analyzed that the grey models should not be applied if the developing coefficient is greater than 1. However, the historical wind speed does not exhibit the same behavior for all horizons. Hence, accomplishing these conditions is not always possible. In such circumstances, traditional restrictions and/or a single rolling window size are not enough. Therefore, admissible constraints and/or adjusting rolling windows are required. It should also be noted that there is no rule of thumb for the minimum sample size for a certain data-driven model. Theoretically, the minimum limit is more observations than the number of parameters in the forecasting model [89].

Table 6
Choices of Mixed Initialization Combinations.

S. No.	First Initialization Method	Second Initialization Method
1	$L_o = 0, T_o = 0$	$L_o = v_1, T_o = 0$
2	$L_o = 0, T_o = \hat{\beta}_{1,T}$	$L_o = \hat{\beta}_{o,T}, T_o = 0$
3	$L_o = 0, T_o = \hat{\beta}_{1,T}$	$L_o = v_1, T_o = 0$
4	$L_o = 0, T_o = 0$	$L_o = \hat{\beta}_{o,T}, T_o = 0$

Based on the above discussions, the shortcomings of considered statistical models and suggested improvements are outlined in Table 5.

2.4. Improved models

To address the limitations of the traditional models, improved versions are introduced. This study considers four models: modDES [14], modMC [12], modGM [11], and WFGM [90].

2.4.1. Modified double exponential smoothing model (modDES)

In the modDES model, two optimizations are suggested to overcome the shortfalls of the conventional model: mixed initialization method and admissible constraints application.

As seen in equations (3)-(6), initialization is required in the recursive calculations of the DES model. Further, Fig. 5 suggests using a proper initialization method to avoid erroneous predictions. Therefore, a mixed initialization method is introduced in [14] to improve the model performance.

The first initialization method determines the best smoothing constants correspond to the least MSE, whereas the following method makes final predictions. Instead of selecting a specific (β, γ), the framework initially records a vector of 81 possible combinations considering a step size of 0.1. A combination is omitted from the principal vector if it predicts a negative wind speed. The final vector is then sorted according to the least MSE value. However, if the resultant is a null vector, then admissible constraints $(1 < \alpha < 2)$ [91] are applied to determine the best smoothing constants. Table 6 shows the four choices for the mixed initialization method.

2.4.2. Modified Markov chain model (modMC)

As mentioned earlier, it is quite challenging to achieve the ergodic properties of Markov processes for every fragment of wind speed. Therefore, a concept of adjusting the rolling window is introduced instead of a single rolling window. In this framework, the theoretical minimum value (N_t) of a particular model is considered as the first window. The k-step ahead wind speed is forecasted and is stored in the memory matrix along with the mean square error. In the next step, more previous training data is added to increase the size of the rolling window $(V_{N_t+d})_{d=1}^{N-N_t}$ and correspondingly, the predicted wind speed and MSE values are stored in the final prediction matrix. After obtaining a matrix of length $(N - N_t + 1)$ values, the best window size is chosen based on the lowest MSE. The optimal window size is modified in this manner for each time step.

2.4.3. Modified grey model (modGM)

In modGM, three improvements are suggested to overcome the shortcomings of the traditional model. Firstly, the singularity phenomenon is solved by applying L'Hopital's rule (see [11] for more details).

Secondly, as discussed in Table 3, the condition of the developing coefficient must be analyzed before making any prediction. Therefore, an additional grey model (termed as Remnant GM (1,1) model) is applied to the residuals when $0.8 < |a| \leq 1$. Based on the two improvements mentioned above, equation (23) is modified as:

Fig. 6. Qualitative Performance of the forecasting models in terms of the average quality of forecasts per a) case study b) model type c) forecasting horizon d) Rolling window.

$$\widehat{V}^{(0)} = \begin{cases} \widehat{v}^{(0)}(1) = v^{(0)}(1) & \\ \widehat{v}^{(0)}(k) = b & a = 0,\ \ k > 1 \\ \widehat{v}^{(0)}(k) = \widehat{v}^{(1)}(k) - \widehat{v}^{(1)}(k-1) = (1-e^{a})\left(v^{(0)}(1) - \dfrac{b}{a}\right)e^{-a(k-1)} & 0 < a \le 0.8,\ \ k > 1 \\ \widehat{v}^{(0)}(k) = (1-e^{a})\left(v^{(0)}(1) - \dfrac{b}{a}\right)e^{-a(k-1)} \pm (1-e^{a_\varepsilon})\left(\varepsilon^{(0)}(2) - \dfrac{b_\varepsilon}{a_\varepsilon}\right)e^{-a_\varepsilon(k_\varepsilon - 1)} & 0.8 < a \le 1,\ \ k > 1,\ \ k_\varepsilon > 2 \end{cases}$$

Similar to the MC model, it is quite difficult to achieve the required levels of posterior error mentioned in Table 4 for a single rolling window. Therefore, the adjusting rolling window is also applied as a third modification to the modGM.

2.4.4. *Weighted fractional grey model (WFGM)*

Another improved GM model considered in this study is Weighted Fractional Grey Model (WFGM). Instead of considering first-order AGO as in equation (18), a weighting parameter (w) and fractional order-AGO (f) are introduced,

$$V^{(wf)} = \left\{v^{(wf)}(1), v^{(wf)}(2), v^{(wf)}(3), \cdots, v^{(wf)}(n)\right\}, \tag{26}$$

where $v^{(wf)}(k) = \sum_{i=1}^{k}\left[\begin{matrix} f \\ k-i \end{matrix}\right] w^{(k-i)} v^{(0)}(i),\ k = 1, 2, 3, \cdots, n$

The corresponding whitenization and time response equations in (22) are modified accordingly.

2.5. *Machine learning models*

Further, to compare the accuracy of the improved models, four machine learning models are also considered. These include Gaussian Process Regression (GPR), Decision Trees (DT), Ensemble Learning, and Support Vector Machines (SVM). The brief description of these models is as follows.

2.5.1. *Gaussian process regression*

Gaussian process regression (GPR) is a nonparametric kernel-based probabilistic machine learning model. Consider a training data $(x_i, y_i)_{i=1}^{n}$ with x as input data and y as response. The linear regression function is defined as:

$$y = f(x) + \varepsilon \tag{27}$$

where ε is assumed to be a Gaussian distribution with a mean and variance of zero, i.e., $\varepsilon N(0, \sigma^2)$. A GPR model describes the response by incorporating latent variables from a Gaussian process and explicit basis functions. The smoothness of the response is captured by the covariance function of the latent variables, and basis functions project the inputs × into a p-dimensional feature space. A Gaussian process is entirely specified by its mean and covariance function [92]. A set of kernel parameters is frequently used to parameterize the covariance function.

2.5.2. *Decision tree*

A decision tree is a symbolic learning approach that uses nodes and consequences to organize information gathered from a training dataset in a hierarchical structure [93]. The objective is to create a model that can predict a target value by learning simple decision rules derived from data attributes. Binary recursive partitioning is commonly used to fit trees. The word binary refers to the fact that the parent node is always divided into two child nodes. Unless it is a terminal node, the term recursive refers to the fact that each child node will eventually become a parent node [94]. The method begins with a root node that branches out into probable consequences. Then, each of those resultant nodes leads to more, which are connected to other possibilities [23]. Eventually, a terminal node or leaf is reached, and the predicted response is stored. The output of a decision tree can be structured in the form of a tree or rules, making the outcomes of decision trees simple to comprehend.

2.5.3. *Ensemble learning*

An ensemble learning is a predictive model that combines many regression trees in a weighted combination to improve the prediction performance. The typical representatives of ensemble learning include Bagging and Boosting, and the same is considered in this study. These iterative approaches generate a committee of expert tree models by resampling with replacement from the initial data set. Next, the expert tree models are averaged using simple averaging [94]. Bagging models average the output of numerous independent estimators trained on bootstraps of the original dataset. On the other hand, boosting models employ a series of decision trees, each of which aims to compensate for the errors of the prior model [23].

2.5.4. *Support vector machine*

Support Vector Machine (SVM) is a commonly applied machine learning model based on Structural Risk Minimization (SRM) from Statistical Learning Theory (SLT) [95]. In this study, linear epsilon-insensitive SVM (ε-SVM) regression is applied. SVM uses a nonlinear kernel function to map the input data into a high-dimensional feature space and then constructs a linear regression function in this hyperspace. Consider a training data x_n and response y_n. The linear regression function is defined as:

$$f(x) = w_i\varphi_i(x) + b' \tag{28}$$

where w, b', and $\varphi(x)$ represents the associative weight, bias term, and mapping function. For each training point x, the aim is to find a function $f(x)$ that deviates from y_n by no more than ε and is as flat as possible. The function is formulated as a convex optimization problem to minimize the regularized function with constraints of equations. Slack variables are also introduced to deal with infeasible constraints. The optimization problem is computationally simpler to solve in Lagrange dual formulation. Finally, new values can be predicted from the following function:

$$f(x, \delta, \delta^*) = \sum_{i=1}^{n} (\delta - \delta^*) k(x_i, x) + b' \tag{29}$$

where δ and δ^* are the Lagrange multipliers and $k(x_i, x)$ is the kernel function.

3. Results and discussions

The performance of the forecasting models is assessed based on normalized Mean Absolute Error (*nMAE*) and normalized Root Mean Square Error (*nRMSE*) and is defined as:

$$nMAE = \frac{MAE}{v_{max} - v_{min}}, \text{where } MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{v}_i - v_i| \tag{30}$$

$$nRMSE = \frac{RMSE}{v_{max} - v_{min}}, \text{where } RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{v}_i - v_i)^2} \tag{31}$$

The normalized performance indicators are feasible for comparing accuracy results, especially when the datasets are non-identical.

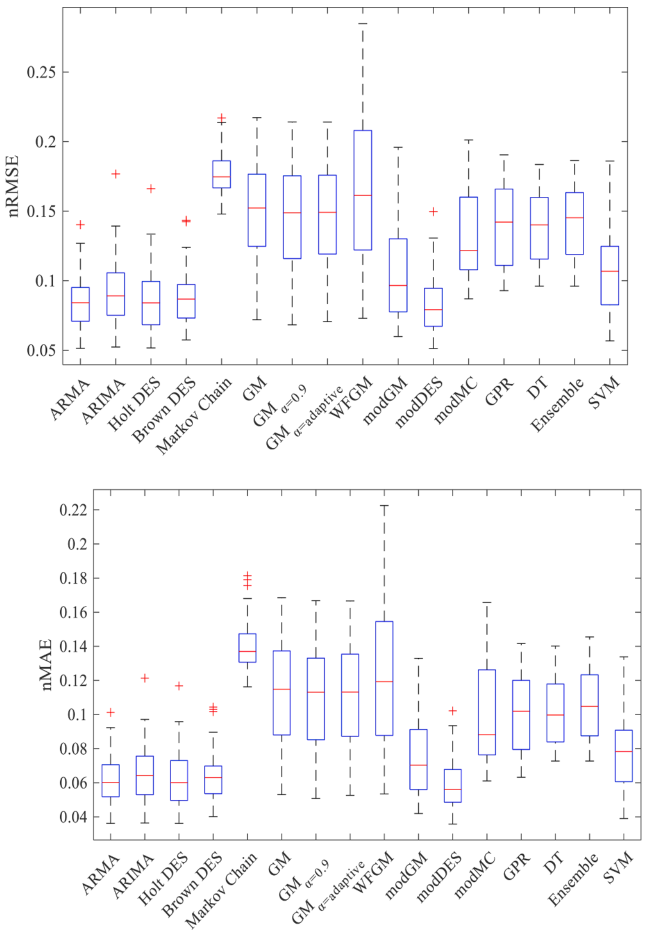


Fig. 7. Variations of errors in terms of boxplot.

3.1. *Qualitative results*

Figure 6 highlights the performance of the forecasting models in terms of the average quality of forecasts. The analysis presented in the figure is illustrative rather than quantitative. With four case studies, sixteen models, three forecasting horizons, and three rolling windows, the final quality-controlled database contains 576 entries. The averages, presented in terms of filled circles, are generated using all nMAE and nRMSE results of that specific category. For instance, sixteen filled circles present the averaged accuracy of sixteen models in Fig. 6b. In contrast, three filled circles show the average quality of forecasts for three forecasting horizons in Fig. 6c. It shows that (4 case studies $\times$ 3 forecasting horizon $\times$ 3 rolling window $=$) 36 data entries are considered to plot Fig. 6b, whereas (4 case studies $\times$ 16 models $\times$ 3 rolling windows $=$)192 data entries are averaged out to plot Fig. 6c.

Figure 6a shows the prediction accuracy per case study. The least errors are observed for Galicia, followed by Palmerston North, Gwadar, and Mpika. On average, Mpika has 30% more errors than Galicia. The main reason for larger errors is the failure of certain model conditions for different time fragments. Even some rare cases of erroneous predictions are also happening, such as the failure of the grey model with a rolling window of 50. As per our analysis, the model's failure is mainly due to frequent flat line data, i.e., no change in the mean value throughout a certain time period in the Mpika case study.

Figure 6b shows prediction accuracy per model type. Visual inspection reveals that the DES models have the least nMAE and nRMSE on average, followed by ARIMA and SVM models. The significantly large errors are witnessed for the Markov Chain model, almost double the errors of the modDES model. However, the accuracy is increased for the

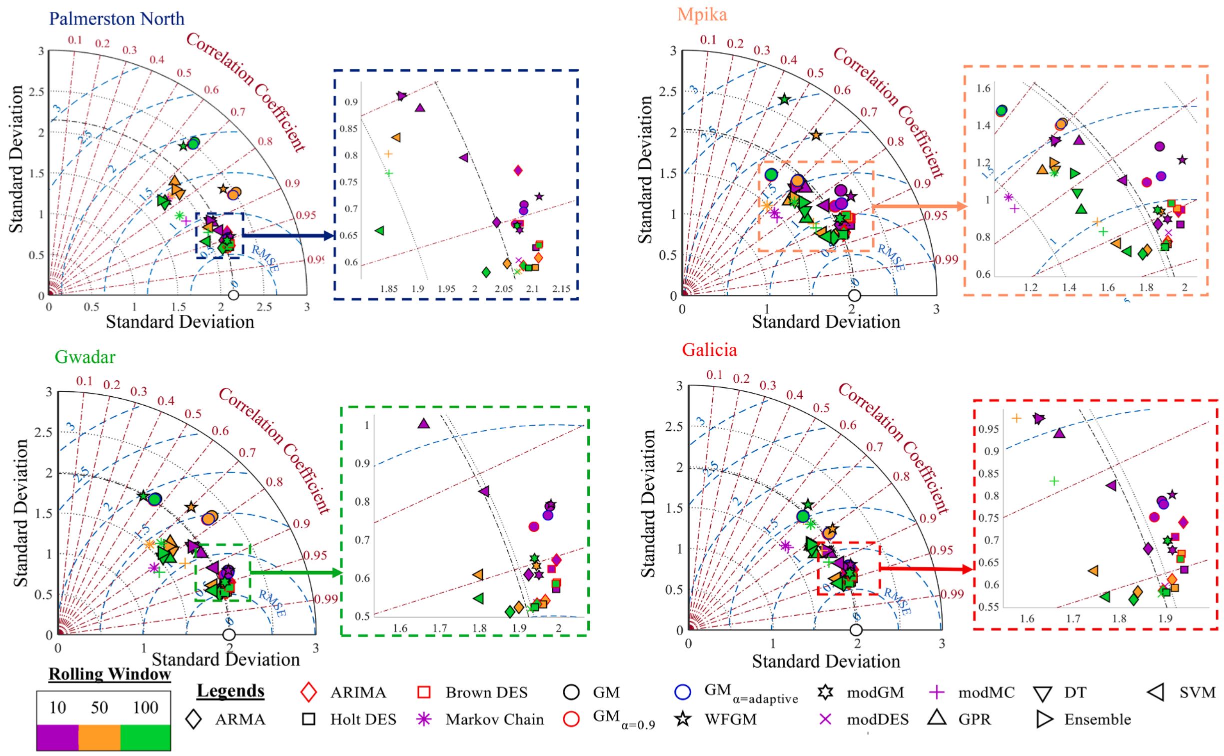


Fig. 8. Taylor Diagram for the comparison of 48 models (16 × 3 rolling windows) for 1-step ahead wind speed forecast.

modMC model. It concludes that the adaptive rolling window is superior to the single rolling window. It is also observed that the modified models showed substantial improvements in the traditional forecasting models. On average, modDES, modMC, and modGM show 4%, 25%, and 28% better performance than the counter traditional models. Further, it seems that grey models have high systematic biases; however, to gain a deeper understanding, one must consider the relationship between model type and rolling window size.

As per expectations, the accuracy of the forecasting models tends to decrease with increasing forecasting horizons, as shown in Fig. 6c. The tendency for errors is 18% and 32% more in 2- and 3-steps ahead. Surprisingly, the accuracy of the forecasting models is also decreasing with increasing rolling window size, as shown in Fig. 6d. However, it does not exclude the possibility that any one kind of model is performing satisfactorily only for a lower rolling window. Therefore, the relationship between model type and rolling window size must be considered before drawing a conclusion.

3.2. Quantitative results

Figure 7 shows the variations of errors in terms of the boxplot. The highest volatility is observed for the WFGM model ($\sigma_{nMAE} = 0.0445$, $\sigma_{nRMSE} = 0.0566$). Hence, proper justification is required while selecting the rolling window for WFGM. Further, the boxplots displayed that Markov Chain models have larger mean errors ($\mu_{nMAE} = 0.1409, \mu_{nRMSE}= 0.178$) with small volatility ($\sigma_{nMAE} = 0.0165$, $\sigma_{nRMSE} = 0.017$). It indicates that the model has generated systematic errors rather than stochastic. Hence, considering a higher window might improve the model performance. Overall, machine learning models have higher volatility than statistical models.

Besides the statistical indicators, the Taylor diagram is a valuable visual tool for comparing several models, as shown in Fig. 8. In terms of their correlation (R), root mean square error (RMSE), and standard deviation, the Taylor Diagram gives a concise statistical description of how well the models fit the measured data. A single point on a 2D polar plot represents the correlation coefficient, root mean square error, and standard deviation as:

$$E^{'2} = \sigma_m^2 + \sigma_f^2 + 2\sigma_m\sigma_f R \tag{32}$$

where $E^{'}$ is the centered RMSE, σ_m is the standard deviation of measured values and σ_f is the standard deviation of the model field. The correlation coefficient and the standard deviation are defined as,

$$R = \frac{\sum_{i=1}^{n}(v_i - v_{mean})(\widehat{v}_i - \widehat{v}_{mean})}{\sqrt{\sum_{i=1}^{n}(v_i - v_{mean})^2(\widehat{v}_i - \widehat{v}_{mean})^2}} \tag{33}$$

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(v_i - v_{mean})^2} \tag{34}$$

The standard deviation is related to the radial distance from the origin, and the azimuthal position is proportional to a correlation between the two models. The measured field's point is shown on a horizontal axis corresponding to the standard deviation. The accuracy of the model is measured by the distance between the reference and the estimated model.

Figure 8 shows that the performance of the forecasting models is almost following the same trend in every case study. Therefore, the best forecasting model is modDES for the considered three rolling windows. Two important conclusions are also drawn from Fig. 8. First, in every case study, the accuracy of WFGM and GM(1,1) decreases with increasing rolling window. It is because when the sample size is large, the perturbation bound of the parameters of GM(1, 1) will change larger [96]. Hence, in terms of model stability, the smaller the sample size of GM(1, 1), the more stable the solution of GM(1, 1) is, and vice versa. However, if a certain time series fragment satisfies quasi-smooth and quasi-exponential checking conditions, then the grey models can make well predictions even with a large number of samples [96]. For this reason, the modGM model has higher accuracy than all the other grey models as it has an adjusting moving window.

Secondly, the size of the rolling window cannot be chosen arbitrarily.

Ranking of models according to Global Performance Indicator (GPI).

		1 step						2 steps						3 steps					
		10		50		100		10		50		100		10		50		100	
Case Study	Models	GPI	Rank	GPI	Rank	GPI	Rank	GPI	Rank	GPI	Rank	GPI	Rank	GPI	Rank	GPI	Rank	GPI	Rank
Palmerston North	ARMA	0.12	6	1.62	3	1.2	3	0.54	4	1.71	5	1.57	4	0.91	3	1.76	5	1.65	5
	ARIMA	−0.22	10	1.61	4	1.2	4	−0.45	10	1.72	3	1.58	3	−0.71	13	1.79	4	1.68	3
	Holt DES	0.34	2	1.67	2	1.22	2	0.57	3	1.79	2	1.59	2	0.62	4	1.86	2	1.68	4
	Brown DES	0.15	4	1.53	5	1.12	5	0.63	2	1.72	4	1.55	5	1.07	2	1.84	3	1.7	2
	Markov Chain	−3.59	16	−2.03	16	−1.19	11	−3.14	16	−1.52	14	−1.12	9	−2.78	16	−1.19	13	−0.77	9
	GM	0.02	8	−0.74	11	−2.71	16	0.13	7	−0.39	11	−2.34	16	−0.06	9	−0.23	12	−2.07	16
	GM with 0.9	0.15	5	−0.63	10	−2.66	14	0.46	5	−0.26	10	−2.28	15	0.47	5	−0.1	10	−2.01	15
	GM with 0.1 to 0.9	0.06	7	−0.61	9	−2.65	13	0.24	6	−0.25	9	−2.27	14	0.13	6	−0.08	9	−2	14
	WFGM	−0.02	9	−0.91	12	−2.68	15	0.07	8	−0.46	12	−2.22	13	−0.13	11	−0.23	11	−1.86	13
	Mod GM	0.20	3	1.4	6	1.01	6	−0.07	9	1.14	6	1.08	6	−0.96	14	0.7	7	0.77	7
	Mod DES	0.41	1	1.69	1	1.23	1	0.86	1	1.82	1	1.62	1	1.22	1	1.92	1	1.74	1
	Mod MC	−2.29	15	0.65	8	0.58	8	−2.33	15	0.97	7	1.04	7	−1.93	15	1.13	6	1.2	6
	GPR	−0.83	12	−1.98	15	−1.05	10	−0.59	13	−2.14	16	−1.46	12	−0.11	10	−2.08	16	−1.52	12
	DT	−0.97	13	−1.4	13	−1.04	9	−0.58	11	−1.36	13	−1.13	10	0.02	7	−1.36	14	−1.04	10
	Ensemble	−0.97	14	−1.71	14	−1.24	12	−0.58	12	−1.66	15	−1.29	11	0.02	8	−1.59	15	−1.24	11
	SVM	−0.38	11	0.66	7	0.86	7	−0.8	14	0.22	8	0.84	8	−0.64	12	−0.07	8	0.63	8
Gwadar	ARMA	0.64	4	1.98	3	1.48	2	1	3	2.05	2	1.44	1	1.49	3	2.05	2	1.58	1
	ARIMA	0.54	6	1.97	4	1.47	4	0.31	7	2.01	4	1.42	3	0.13	6	1.97	5	1.54	3
	Holt DES	0.81	2	1.99	2	1.48	3	0.92	4	2.03	3	1.41	4	0.98	4	2	4	1.52	4
	Brown DES	0.62	5	1.83	5	1.36	6	1.11	2	1.97	5	1.34	5	1.73	1	2.04	3	1.5	5
	Markov Chain	−3.07	16	−1.83	16	−1.01	12	−2.7	16	−1.5	16	−0.97	12	−2.03	16	−1.21	13	−0.75	9
	GM	−0.05	9	−1.06	14	−2.08	15	−0.08	10	−0.88	12	−2.14	15	−0.29	9	−0.77	11	−2.04	15
	GM with 0.9	0.15	7	−0.98	13	−2.07	13	0.35	6	−0.78	9	−2.12	13	0.42	5	−0.64	9	−2.01	13
	GM with 0.1 to 0.9	0.04	8	−0.98	12	−2.08	14	0.07	8	−0.78	10	−2.14	14	−0.1	8	−0.65	10	−2.02	14
	WFGM	−0.06	10	−1.64	15	−2.51	16	−0.07	9	−1.38	14	−2.55	16	−0.3	10	−1.16	12	−2.41	16
	Mod GM	0.66	3	1.7	6	1.22	7	0.64	5	1.51	6	0.97	7	0.12	7	0.96	6	0.67	8
	Mod DES	0.93	1	2.01	1	1.49	1	1.3	1	2.08	1	1.43	2	1.71	2	2.1	1	1.57	2
	Mod MC	−2.44	15	0.32	8	−0.62	11	−1.96	15	0.59	8	0.6	8	−1.24	15	0.78	7	0.8	7
	GPR	−1.29	12	−0.61	11	0.02	8	−1.11	12	−1.47	15	−0.89	10	−0.43	13	−1.81	16	−1.32	12
	DT	−1.89	13	−0.33	9	−0.2	9	−1.27	13	−0.84	11	−0.74	9	−0.36	11	−1.23	14	−1.08	10
	Ensemble	−1.89	14	−0.61	10	−0.36	10	−1.27	14	−1.03	13	−0.94	11	−0.36	12	−1.35	15	−1.29	11
	SVM	−0.31	11	1.66	7	1.37	5	−1.08	11	1.08	7	1.18	6	−1.02	14	0.53	8	1.11	6
Mpika	ARMA	1.06	3	1.31	2	0.55	2	1.77	1	1.57	2	0.86	2	1.95	1	1.86	1	1.11	1
	ARIMA	0.88	5	1.27	4	0.53	4	0.79	6	1.47	4	0.81	4	0.26	7	1.61	4	0.99	5
	Holt DES	1.15	2	1.3	3	0.55	3	1.27	4	1.52	3	0.85	3	0.81	4	1.69	3	1.07	3
	Brown DES	0.84	6	0.8	7	0.13	8	1.64	3	1.06	5	0.47	7	1.89	2	1.32	6	0.7	6
	Markov Chain	−2.44	16	−2.08	15	−0.9	12	−1.63	16	−0.51	9	−0.34	10	−1.01	13	0.08	8	0.08	8
	GM	−0.55	11	−1.26	14	−1.75	15	−0.74	13	−0.84	12	−1.21	15	−1.35	16	−0.26	11	−0.81	15
	GM with 0.9	0.1	7	−1.23	12	−1.74	13	0.12	7	−0.8	10	−1.2	13	−0.36	10	−0.2	9	−0.79	13
	GM with 0.1 to 0.9	0.03	8	−1.25	13	−1.75	14	−0.29	8	−0.81	11	−1.21	14	−1.09	14	−0.22	10	−0.8	14
	WFGM	−0.2	10	−2.57	16	−3.44	16	−0.31	9	−2.39	16	−3.12	16	−1.09	15	−1.87	16	−2.86	16
	Mod GM	0.99	4	0.82	6	0.19	6	0.87	5	0.49	8	0.15	8	−0.63	11	−0.5	12	−0.42	10
	Mod DES	1.29	1	1.32	1	0.56	1	1.73	2	1.57	1	0.86	1	1.62	3	1.79	2	1.1	2
	Mod MC	−1.8	15	0.53	8	0.16	7	−0.92	14	1.05	6	0.66	5	−0.26	9	1.57	5	1.03	4
	GPR	−1.44	12	−0.62	10	−0.12	9	−0.62	12	−1.05	14	−0.26	9	0.19	8	−0.88	13	−0.35	9
	DT	−1.79	13	−0.56	9	−0.35	10	−0.55	10	−0.93	13	−0.43	11	0.45	5	−0.96	14	−0.52	11
	Ensemble	−1.79	14	−0.71	11	−0.57	11	−0.55	11	−1.12	15	−0.67	12	0.45	6	−1.13	15	−0.73	12
	SVM	−0.11	9	1.03	5	0.47	5	−1.27	15	0.74	7	0.56	6	−0.81	12	0.55	7	0.56	7

(*continued on next page*)

Table 7 (*continued*)

Case Study	Models	1 step 10 GPI	Rank	1 step 50 GPI	Rank	1 step 100 GPI	Rank	2 steps 10 GPI	Rank	2 steps 50 GPI	Rank	2 steps 100 GPI	Rank	3 steps 10 GPI	Rank	3 steps 50 GPI	Rank	3 steps 100 GPI	Rank
Galicia	ARMA	0.33	4	1.86	3	1.35	2	0.6	2	1.94	2	1.76	1	0.9	3	1.93	2	0.33	4
	ARIMA	0.18	6	1.83	4	1.33	4	0.04	8	1.84	4	1.71	4	−0.25	12	1.82	4	0.18	6
	Holt DES	0.52	2	1.89	2	1.35	3	0.51	4	1.94	3	1.72	3	0.52	4	1.87	3	0.52	2
	Brown DES	0.27	5	1.65	5	1.13	6	0.56	3	1.76	5	1.52	5	0.92	2	1.76	5	0.27	5
	Markov Chain	−3.38	16	−2.09	16	−2.11	12	−3.17	16	−2.01	16	−1.69	12	−2.87	16	−1.84	16	−3.38	16
	GM	−0.01	9	−0.48	14	−2.3	15	−0.02	9	−0.32	11	−1.88	15	−0.19	11	−0.25	11	−0.01	9
	GM with 0.9	0.1	7	−0.44	12	−2.29	14	0.21	5	−0.26	9	−1.86	14	0.18	7	−0.17	9	0.1	7
	GM with 0.1 to 0.9	0.01	8	−0.45	13	−2.28	13	0.04	7	−0.28	10	−1.86	13	−0.09	9	−0.2	10	0.01	8
	WFGM	−0.04	10	−0.57	15	−2.64	16	−0.03	10	−0.41	12	−2.24	16	−0.18	10	−0.33	12	−0.04	10
	Mod GM	0.36	3	1.54	7	0.98	7	0.07	6	1.11	7	0.86	7	−0.65	14	0.25	8	0.36	3
	Mod DES	0.62	1	1.91	1	1.36	1	0.83	1	1.99	1	1.74	2	1.13	1	1.97	1	0.62	1
	Mod MC	−3.01	15	0.07	8	0.21	8	−2.8	15	0.26	8	0.69	8	−2.46	15	0.4	7	−3.01	15
	GPR	−0.75	12	−0.16	10	−0.21	9	−0.34	11	−1.12	15	−0.7	9	0.18	8	−1.45	15	−0.75	12
	DT	−0.96	13	−0.12	9	−0.55	10	−0.37	12	−0.59	13	−0.78	10	0.25	5	−1.01	13	−0.96	13
	Ensemble	−0.96	14	−0.38	11	−0.79	11	−0.37	13	−0.82	14	−0.97	11	0.25	6	−1.12	14	−0.96	14
	SVM	−0.21	11	1.6	6	1.26	5	−0.66	14	1.21	6	1.45	6	−0.5	13	0.59	6	−0.21	11

This is because the historical data does not have an equivalent length of behavior throughout. As a result, the selection of incorrect window sizes increased the prediction errors. Therefore, a dynamic moving window is necessary for these models. For better understanding, individual Taylor diagrams are also provided in the Supplementary Material.

The analyses of Fig. 8 concluded that the modDES model has the best overall performance of all the models studied. However, comparing a large number of models in this manner is quite challenging. Therefore, another popular method for assessing model performance is to rank the model. Instead of analyzing individual accuracy measures, the global performance indicator (GPI) scales all statistical indicators between 0 and 1 to give them equal weight [97]. GPI is defined mathematically as,

$$GPI_i = \sum_{j=1}^{n} \xi_j \left(\widetilde{I}_i - I_{ij}\right) \tag{35}$$

where ξ_j equals −1 for Pearson's Correlation Coefficient and equals 1 for all the other indicators. $\widetilde{I}_i$ is the median of scaled values of indicator j and I_{ij} is the scaled value of indicator j for model i. The interpretation of the GPI shows that if the value of a statistical indicator is smaller than the median, the greater the distance between the value and the median value of all other models, the more accurate the model is compared to other models. Similarly, if a value of a statistical indicator is greater than the median, the higher the value goes from the median, the model is less accurate than other models. Therefore, a greater GPI score indicates that a model is more accurate.

Table 7 shows the ranking of the investigated wind speed forecasting models according to GPI. Based on tabulated results, it is concluded that exponential smoothing models have greater GPI in most of the cases, whereas Markov Chain models are among the poorest. Although the modMC model is far better than the traditional MC model, however, a rolling window of $n - 100$ is not enough to address the uncertainty of data. It is again evident from the ranking that the grey models are more suitable for lesser sample data.

3.3. Comparison of the best models

Table 7 and Fig. 8 illustrate that modDES and SVM models are the best choice from the categories of improved and machine learning models. To further exhibit the forecasting performance of the best models, a more explicit comparison is provided in Table 8. As seen in the Table, the nMAE, nRMSE, and R of the modDES model are higher than the SVM model in most cases. Considering the Gwadar case study ($N = 10$, $k = 1$ step ahead), for instance, the nRMSE of the modDES and SVM model is 0.053 and 0.083, respectively. It is also observed that the difference in error decreases with increasing the rolling window. For the Mpika case study ($N = 100$, $k = 2$ steps ahead), the nRMSE of the modDES and SVM model is 0.107 and 0.117. However, this is not a generalized conclusion. For the Palmerston North case study ($N = 100$, $k = 3$ steps ahead), the difference is still higher, with nRMSE values of 0.086 and 0.124 for modDES and SVM models. These conclusions show that wind speed time series contain linear, nonlinear, or both patterns in different fragments of wind time series. Hence a single prediction model is not adequate for forecasting wind speed. Therefore, an ensemble of SVM with modDES would be the best choice for small sample datasets to handle both linear and nonlinear characteristics of wind speed equally.

3.4. Future directions

3.4.1. Unaddressed issues of GM (1,1)

The problem of exceeding prediction in GM(1,1) is due to the singular phenomena that arise mainly because of the developing coefficient a. However, as shown in Fig. 9, there are still areas that need to explore for the erroneous prediction. For the given time series fragment, $|a| =$

Table 8
Comparison of modDES and SVM models for selected case studies.

Case study	Rolling Window (N)	Models	k steps ahead								
			1-step			2-steps			3-steps		
			nMAE	nRMSE	*R*	nMAE	nRMSE	*R*	nMAE	nRMSE	*R*
Palmerston North	10	modDES	0.046	0.061	0.960	0.058	0.079	0.935	0.068	0.091	0.915
		SVM	0.061	0.082	0.928	0.079	0.106	0.882	0.087	0.117	0.857
	50	modDES	0.044	0.059	0.963	0.056	0.075	0.940	0.065	0.086	0.923
		SVM	0.064	0.089	0.913	0.087	0.121	0.840	0.103	0.142	0.781
	100	modDES	0.044	0.059	0.963	0.056	0.075	0.940	0.065	0.086	0.923
		SVM	0.054	0.074	0.941	0.076	0.102	0.883	0.092	0.124	0.823
Gwadar	10	modDES	0.037	0.053	0.964	0.055	0.080	0.921	0.070	0.101	0.876
		SVM	0.058	0.083	0.910	0.086	0.122	0.814	0.102	0.142	0.753
	50	modDES	0.036	0.052	0.966	0.054	0.076	0.925	0.068	0.096	0.884
		SVM	0.045	0.063	0.947	0.076	0.103	0.854	0.101	0.134	0.753
	100	modDES	0.036	0.051	0.966	0.054	0.076	0.925	0.068	0.095	0.885
		SVM	0.039	0.057	0.957	0.062	0.084	0.903	0.081	0.108	0.836
Mpika	10	modDES	0.055	0.080	0.919	0.083	0.120	0.826	0.102	0.150	0.745
		SVM	0.078	0.112	0.836	0.120	0.169	0.653	0.134	0.186	0.595
	50	modDES	0.052	0.074	0.930	0.077	0.108	0.849	0.093	0.131	0.782
		SVM	0.060	0.083	0.907	0.092	0.126	0.772	0.114	0.153	0.650
	100	modDES	0.052	0.073	0.931	0.076	0.107	0.852	0.092	0.128	0.787
		SVM	0.056	0.077	0.921	0.086	0.117	0.807	0.107	0.142	0.700
Galicia	10	modDES	0.039	0.057	0.954	0.056	0.080	0.913	0.066	0.094	0.878
		SVM	0.056	0.081	0.908	0.080	0.113	0.827	0.090	0.125	0.788
	50	modDES	0.038	0.056	0.956	0.055	0.078	0.917	0.065	0.091	0.885
		SVM	0.046	0.064	0.941	0.070	0.094	0.870	0.088	0.118	0.791
	100	modDES	0.038	0.056	0.956	0.054	0.077	0.918	0.064	0.090	0.888
		SVM	0.042	0.058	0.952	0.061	0.082	0.900	0.076	0.101	0.848

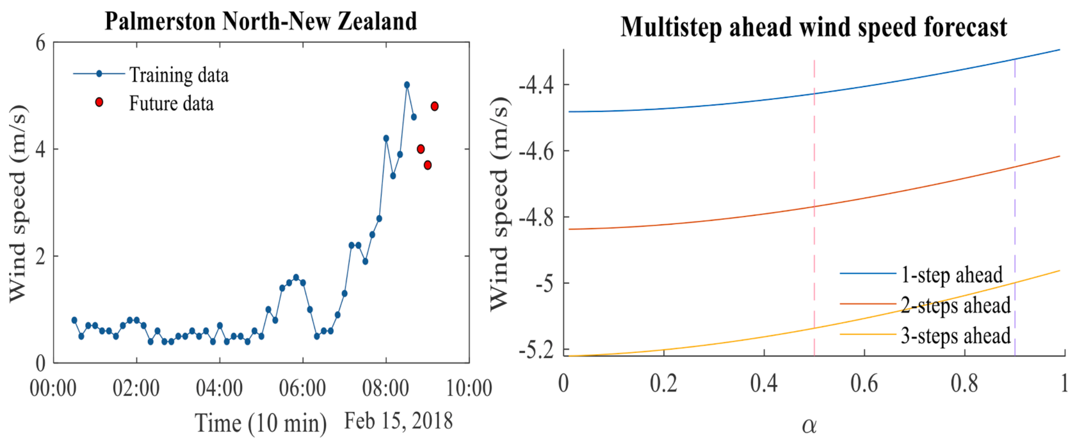


Fig. 9. Failure of GM(1,1) at a rolling window of 50.

0.0723, hence the problem is not due to a singular phenomenon. The rolling window is 50, and optimizing the hyperparameters does not solve the erroneous prediction problem.

3.4.2. Prediction lag

Two main inconsistencies might be present in comparing wind speed forecasting models with measured data: amplitude and phase errors (see Fig. 10) [18]. Vertical deviations are defined by amplitude errors after the values are under/overestimated, whereas phase errors are induced by horizontal mismatches that lag the predicted wind speeds. One solution is to apply multilevel decomposition [98] with proper methods of selecting length and level [99].

3.4.3. Imputation techniques

Boxplots and PCHIP are applied to identify and impute the outliers in this study. However, caution is needed when the repository shows flat data over time. Hence, appropriate anomaly detection methods are required to identify the outliers in small sample datasets.

4. Conclusions

This study analyzed statistical wind speed forecasting models for small sample datasets. With four case studies, sixteen models, three rolling windows, and three forecasting horizons, the final quality-controlled database contains 576 entries. The generalized conclusions are as follows:

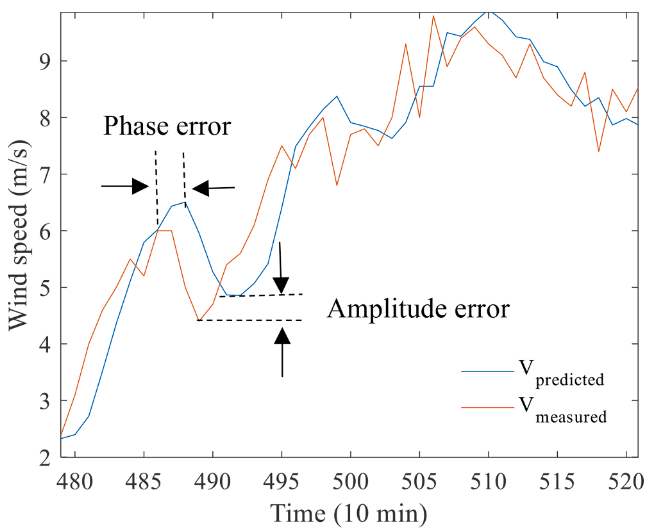


Fig. 10. Two main inconsistencies: Amplitude and Phase errors.

- The primary problem with the ARIMA, DES, and GM(1,1) is erroneous wind speed forecast. Two types of erroneous predictions were observed: exceeding and negative. Exceeding predictions are the unreasonable wind speeds that differ considerably from the rest of the observations, whereas negative predictions are impossible values.
- Three possible reasons for erroneous forecasts are uncalibrated model parameters, incorrect selection of initial conditions, and truncation error.
- Specific optimizations can overcome the shortcomings of the traditional models. Such optimizations include the Mixed initialization method, admissible constraints application, residual time series model, and adjusting rolling window.
- The highest errors were observed for Mpika. The main reason for larger errors is the failure of certain model conditions for different time fragments. The failure of the model is mainly due to frequent flat line data, i.e., no change in the mean value throughout a certain time period.
- Visual inspection revealed that the DES models have the least nMAE and nRMSE on average, followed by ARIMA and SVM models. However, the significantly larger errors are witnessed for the Markov Chain model, almost double the errors of the modDES model.
- On average, modDES, modMC, and modGM show 4%, 25%, and 28% better performance than the counter traditional models.
- The tendency of errors is 18% and 32% more in 2- and 3-steps ahead.
- The highest volatility is observed for the WFGM model. Hence, proper justification is required while selecting the rolling window for WFGM.
- Analysis showed that Markov Chain models have larger mean errors with small volatility. It indicates that the model has generated systematic errors rather than stochastic. Hence, considering a higher window might improve the model performance.
- Grey models perform the best with small sample datasets. It revealed that increasing the rolling window is not always helpful to increase the model accuracy.
- From the category of machine learning models, SVM is proven to be the best choice.

Although the improved models showed higher accuracy than the conventional models, there are still erroneous prediction problems that need further research to enhance the model performance.

5. Data availability

The raw data is retrieved from Energy Sector Management Assistance Program (https://energydata.info/), National Institute of Water and Atmospheric Research (https://cliflo.niwa.co.nz/), and Sotavento wind farm (https://www.sotaventogalicia.com/en/technical-area/real-time-data/historical/) databases.

CRediT authorship contribution statement

Muhammad Uzair Yousuf: Conceptualization, Methodology, Software, Validation, Investigation, Writing – original draft. **Ibrahim Al-Bahadly:** Resources, Supervision, Writing – review & editing. **Ebubekir Avci:** Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to acknowledge the NIWA, ESMAP, and Sotavento wind farm for their public data.

Funding

This research was supported by Tilt Renewables Tararua Wind Farm Research Bursary, Department of Mechanical and Electrical Engineering, Massey University, and Higher Education Commission (HEC) Pakistan Grant 5-1/HRD/UESTPI(Batch VI)/6082/2019/HEC.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enconman.2022.115658.

References

[1] "Statistical Review of World Energy," 2021. [Online]. Available: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2021-full-report.pdf.
[2] Hodge B, Florita A, Sharp J, Margulis M, Mcreavy D. The value of improved short-term wind power forecasting. National Renewable Energy Laboratory (NREL), Technical Report No. NREL/TP-5D00–63175, 2015.
[3] Mc Garrigle EV, Leahy PG. Quantifying the value of improved wind energy forecasts in a pool-based electricity market. Renewable Energy 2015;80:517–24.
[4] Xu A, Yang T, Ji J, Gao Y, Gu C. Calculating reserve power requirements from wind–power forecasts. J Eng 2019;2019(9):5427–31.
[5] Zhang F, Qiao Y, Lu Z. Extreme wind power forecast error analysis considering its application in day-ahead reserve capacity planning. IET Renew Power Gener 2018;12(16):1923–30.
[6] Kosana V, Teeparthi K, Madasthu S, Kumar S. A novel reinforced online model selection using Q-learning technique for wind speed prediction. Sustainable Energy Technol Assess 2022;49:101780.
[7] Yousuf MU, Al-Bahadly I, Avci E. Current perspective on the accuracy of deterministic wind speed and power forecasting. IEEE Access 2019;7:159547–64.
[8] Donadio L, Fang J, Porté-Agel F. Numerical weather prediction and artificial neural network coupling for wind energy forecast. Energies 2021;14(2):338.
[9] Jung J, Broadwater RP. Current status and future advances for wind speed and power forecasting. Renew Sustain Energy Rev 2014;31:762–77.
[10] Sui A, Qian W. Forecasting the wind power generation in China by seasonal grey forecasting model based on collaborative optimization. *RAIRO–Operat Res* 2021;55(5).
[11] Yousuf MU, Al-Bahadly I, Avci E. A modified GM (1, 1) model to accurately predict wind speed. Sustainable Energy Technol Assess 2021;43:100905.
[12] Yousuf MU, Al-Bahadly I, Avci E. Short-term wind speed forecasting based on hybrid modwt-arima-markov model. IEEE Access 2021;9:79695–711.
[13] Li W, Jia X, Li X, Wang Y, Lee J. A Markov model for short term wind speed prediction by integrating the wind acceleration information. Renewable Energy 2021;164:242–53.

[14] Yousuf MU, Al-Bahadly I, Avci E. Wind speed prediction for small sample dataset using hybrid first-order accumulated generating operation-based double exponential smoothing model. *Energy Sci Eng* 2022.
[15] Priya CB, Arulanand N. Univariate and multivariate models for Short-term wind speed forecasting. *Mater Today: Proc* 2021.
[16] Wang Y, Wang D, Tang Y. Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods. IEEE Access 2020;8:17071–9.
[17] Zaman U, Teimourzadeh H, Sangani EH, Liang X, Chung CY. Wind Speed Forecasting Using ARMA and Neural Network Models, in *2021 IEEE Electrical Power and Energy Conference (EPEC)*, 2021: IEEE, pp. 243-248.
[18] Liu X, Lin Z, Feng Z. Short-term offshore wind speed forecast by seasonal ARIMA-A comparison against GRU and LSTM. Energy 2021;227:120492.
[19] Elsaraiti M, Merabet A. A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed. Energies 2021;14(20): 6782.
[20] Ma Z, Guo S, Xu G, Aziz S. Meta learning-based hybrid ensemble approach for short-term wind speed forecasting. IEEE Access 2020;8:172859–68.
[21] He Y, Li H, Wang S, Yao X. Uncertainty analysis of wind power probability density forecasting based on cubic spline interpolation and support vector quantile regression. Neurocomputing 2021;430:121–37.
[22] Prasad SMM, Nguyen-Huy T, Deo R. Support vector machine model for multistep wind speed forecasting. In: Predictive Modelling for Energy Management and Power Systems Engineering. Elsevier; 2021. p. 335–89.
[23] Ahmadi A, Nabipour M, Mohammadi-Ivatloo B, Amani AM, Rho S, Piran MJ. Long-term wind power forecasting using tree-based learning algorithms. IEEE Access 2020;8:151511–22.
[24] Cai R, Xie S, Wang B, Yang R, Xu D, He Y. Wind speed forecasting based on extreme gradient boosting. IEEE Access 2020;8:175063–9.
[25] Zhang C, Peng T, Nazir MS. A novel hybrid approach based on variational heteroscedastic Gaussian process regression for multi-step ahead wind speed forecasting. Int J Electr Power Energy Syst 2022;136:107717.
[26] Wang H, Zhang Y-M, Mao J-X. Sparse Gaussian process regression for multi-step ahead forecasting of wind gusts combining numerical weather predictions and on-site measurements. J Wind Eng Ind Aerodyn 2022;220:104873.
[27] Hua L, Zhang C, Peng T, Ji C, Nazir MS. Integrated framework of extreme learning machine (ELM) based on improved atom search optimization for short-term wind speed prediction. Energy Convers Manage 2022;252:115102.
[28] Saoud LS, Al-Marzouqi H, Deriche M. Wind speed forecasting using the stationary wavelet transform and quaternion adaptive-gradient methods. IEEE Access 2021;9: 127356–67.
[29] Duan J, Wang P, Ma W, Fang S, Hou Z. A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting. Int J Electr Power Energy Syst 2022;134:107452.
[30] Liu X, Zhang H, Kong X, Lee KY. Wind speed forecasting using deep neural network with feature selection. Neurocomputing 2020;397:393–403.
[31] Xia X, Wang X. A novel hybrid model for short-term wind speed forecasting based on twice decomposition, PSR, and IMVO-ELM. Complexity 2022;2022.
[32] Liu H, Yang R, Wang T, Zhang L. A hybrid neural network model for short-term wind speed forecasting based on decomposition, multi-learner ensemble, and adaptive multiple error corrections. Renewable Energy 2021;165:573–94.
[33] Xiao L, Wang J, Dong Y, Wu J. Combined forecasting models for wind energy forecasting: a case study in China. Renew Sustain Energy Rev 2015;44:271–88.
[34] Liu H, Chen C, Lv X, Wu X, Liu M. Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods. Energy Convers Manage 2019;195: 328–45.
[35] Tawn R, Browell J. A review of very short-term wind and solar power forecasting. Renew Sustain Energy Rev 2022;153:111758.
[36] Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: concerns and ways forward. PLoS One 2018;13(3):e0194889.
[37] Dhiman HS, Deb D, Guerrero JM. Hybrid machine intelligent SVR variants for wind forecasting and ramp events. Renew Sustain Energy Rev 2019;108:369–79.
[38] Gupta D, Natarajan N, Berlin M. Short-term wind speed prediction using hybrid machine learning techniques. Environ Sci Pollut Res 2021:1–19.
[39] Ranganayaki V, Deepa S. Linear and non-linear proximal support vector machine classifiers for wind speed prediction. Cluster Comput 2019;22(1):379–90.
[40] Hazarika BB, Gupta D, Natarajan N. Wavelet kernel least square twin support vector regression for wind speed prediction. Environ Sci Pollut Res 2022:1–17.
[41] Kong X, Liu X, Shi R, Lee KY. Wind speed prediction using reduced support vector machines with feature selection. Neurocomputing 2015;169:449–56.
[42] Petropoulos F, et al. Forecasting: theory and practice. Int J Forecast 2022.
[43] Mi X-W, Liu H, Li Y-F. Wind speed forecasting method using wavelet, extreme learning machine and outlier correction algorithm. Energy Convers Manage 2017; 151:709–22.
[44] Zhang Y, Zhao Y, Kong C, Chen B. A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic. Energy Convers Manage 2020;203:112254.
[45] Tian Z, Wang G, Ren Y. Short-term wind speed forecasting based on autoregressive moving average with echo state network compensation. Wind Eng 2020;44(2): 152–67.
[46] Wang L, Li X, Bai Y. Short-term wind speed prediction using an extreme learning machine model with error correction. Energy Convers Manage 2018;162:239–50.
[47] Liu M-D, Ding L, Bai Y-L. Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. Energy Convers Manage 2021;233:113917.
[48] Bai Y, Tang L, Fan M, Ma X, Yang Y. Fuzzy first-order transition-rules-trained hybrid forecasting system for short-term wind speed forecasts. Energies 2020;13 (13):3332.
[49] Jiang P, Wang B, Li H, Lu H. Modeling for chaotic time series based on linear and nonlinear framework: application to wind speed forecasting. Energy 2019;173: 468–82.
[50] Niu M, Sun S, Wu J, Zhang Y. Short-term wind speed hybrid forecasting model based on bias correcting study and its application. Math Probl Eng 2015;2015.
[51] Huang X, Wang J, Huang B. Two novel hybrid linear and nonlinear models for wind speed forecasting. Energy Convers Manage 2021;238:114162.
[52] Qolipour M, Mostafaeipour A, Saidi-Mehrabad M, Arabnia HR. Prediction of wind speed using a new Grey-extreme learning machine hybrid algorithm: a case study. Energy Environ 2019;30(1):44–62.
[53] Zhang Y, Sun H, Guo Y. Wind power prediction based on PSO-SVR and grey combination model. IEEE Access 2019;7:136254–67.
[54] Wu D, Gao C. Short-term wind power generation forecasting based on the SVM-GM approach. Electr Power Compon Syst 2018;46(11–12):1250–64.
[55] Kani SP, Ardehali M. Very short-term wind speed prediction: a new artificial neural network–Markov chain model. Energy Convers Manage 2011;52(1):738–45.
[56] Zhang Y, Han J, Pan G, Xu Y, Wang F. A multi-stage predicting methodology based on data decomposition and error correction for ultra-short-term wind energy prediction. J Cleaner Prod 2021;292:125981.
[57] Wang Y, Wang J, Wei X. 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 2015;91:556–72.
[58] Wind Data obtained from World Bank via ENERGYDATA.info, under a project funded by the Energy Sector Management Assistance Program (ESMAP). Available: https://energydata.info/.
[59] *The National Climate Database*, National Institute of Water and Atmospheric Research (NIWA). [Online]. Available: https://cliflo.niwa.co.nz/.
[60] *Historical-Real Time data*. [Online]. Available: https://www.sotaventogalicia.com/en/technical-area/real-time-data/historical/.
[61] Qiu X, Ren Y, Suganthan PN, Amaratunga GA. Short-term wind power ramp forecasting with empirical mode decomposition based ensemble learning techniques, in *2017 IEEE Symposium Series on computational intelligence (SSCI)*, 2017: IEEE, pp. 1-8.
[62] Laudari R, Sapkota B, Banskota K. Validation of wind resource in 14 locations of Nepal. Renewable Energy 2018;119:777–86.
[63] Zhou Q, Wang C, Zhang G. Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems. Appl Energy 2019; 250:1559–80.
[64] Zhao Y, Ye L, Wang W, Sun H, Ju Y, Tang Y. Data-driven correction approach to refine power curve of wind farm under wind curtailment. IEEE Trans Sustainable Energy 2017;9(1):95–105.
[65] Jeong J, Park E, Han WS, Kim K, Choung S, Chung IM. Identifying outliers of non-Gaussian groundwater state data based on ensemble estimation for long-term trends. J Hydrol 2017;548:135–44.
[66] Schlechtingen M, Santos IF. Comparative analysis of neural network and regression based condition monitoring approaches for wind turbine fault detection. Mech Syst Sig Process 2011;25(5):1849–75.
[67] Wang J, Xiong S. A hybrid forecasting model based on outlier detection and fuzzy time series–A case study on Hainan wind farm of China. Energy 2014;76:526–41.
[68] Ouyang T, Huang H, He Y, Tang Z. Chaotic wind power time series prediction via switching data-driven modes. Renewable Energy 2020;145:270–81.
[69] Zhang Z, Fang H, Wang H. A new mi-based visualization aided validation index for mining big longitudinal web trial data. IEEE Access 2016;4:2272–80.
[70] Rabbath C, Corriveau D. A comparison of piecewise cubic Hermite interpolating polynomials, cubic splines and piecewise linear functions for the approximation of projectile aerodynamics. Defence Technol 2019;15(5):741–57.
[71] Maki Y, Monno Y, Yoshizaki K, Tanaka M, Okutomi M. Inter-Beat Interval Estimation from Facial Video Based on Reliability of BVP Signals, in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019: IEEE, pp. 6525-8.
[72] Quej VH, Almorox J, Arnaldo JA, Saito L. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J Atmos Sol Terr Phys 2017;155:62–70.
[73] Ur Rehman W, Khan MFN, Sajjad IA, Umar Afzaal M. Probabilistic generation model for grid connected wind DG. J Renew Sustain Energy 2019;11(4):045301.
[74] Azpúrua MA, et al. A review on the drawbacks and enhancement opportunities of the feature selective validation. IEEE Trans Electromagn Compat 2014;56(4): 800–7.
[75] Brown RG. Statistical forecasting for inventory control. McGraw/Hill; 1959.
[76] Holt, “Forecasting Seasonals and Trends by Exponentially Weighted Averages,” *(ONR Memorandum No. 52), Vol. 10. Carnegie Institute of Technology,* 1957.
[77] Gardner Jr ES. Exponential smoothing: the state of the art—Part II. Int J Forecast 2006;22(4):637–66.
[78] Montgomery DC, Jennings CL, Kulahci M. Introduction to time series analysis and forecasting. John Wiley & Sons; 2015.
[79] “Methods and formulas for Double Exponential Smoothing.” https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/time-series/how-to/double-exponential-smoothing/methods-and-formulas/methods-and-formulas/.
[80] Lakatos L, Szeidl L, Telek M. Introduction to queueing systems with telecommunication applications. Springer; 2019.
[81] Ju-Long D. Control problems of grey systems. J Syst Control Lett 1982;1(5): 288–94.

[82] Liu S, Yang Y, Forrest J. Grey data analysis. Singapore: Springer Singapore; 2017.
[83] Hyndman RJ *et al.* "Package 'forecast'." https://cran.r-project.org/web/packages/forecast/forecast.pdf.
[84] Qian W, Wang J. An improved seasonal GM (1, 1) model based on the HP filter for forecasting wind power generation in China. Energy 2020;209:118499.
[85] Wu W, Ma X, Zeng B, Wang Y, Cai W. Forecasting short-term renewable energy consumption of China using a novel fractional nonlinear grey Bernoulli model. Renewable Energy 2019;140:70–87.
[86] Zhou D, Hu F, Zhu Q, Wang Q. Regional allocation of renewable energy quota in China under the policy of renewable portfolio standards. Resour Conserv Recycl 2022;176:105904.
[87] Deng M, Dong Y. Application of Improved Grey GM (1, 1) Model in Power Prediction of Wind Farm, in *2019 Chinese Control And Decision Conference (CCDC)*, 2019: IEEE, pp. 3764-9.
[88] Xiao Y, Hua Z. Misalignment fault prediction of wind turbines based on combined forecasting model. Algorithms 2020;13(3):56.
[89] Hyndman RJ, Kostenko AV. Minimum sample size requirements for seasonal forecasting models. Foresight 2007;6(Spring):12–5.
[90] Shen Q-Q, Shi Q, Tang T-P, Yao L-Q. A novel weighted fractional GM (1, 1) model and its applications. Complexity 2020;2020.
[91] Hyndman R, Koehler AB, Ord JK, Snyder RD. Forecasting with exponential smoothing: the state space approach. Springer Science & Business Media; 2008.
[92] Bousquet O, von Luxburg U, Rätsch G. Advanced Lectures on Machine Learning: ML Summer Schools 2003, Canberra, Australia, February 2-14, 2003, Tübingen, Germany, August 4-16, 2003, Revised Lectures. Springer, 2011.
[93] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Routledge; 2017.
[94] Moisen GG. "Classification and regression trees," *In: Jørgensen, Sven Erik; Fath, Brian D. Encyclopedia of Ecology, volume 1. Oxford, UK: Elsevier. p. 582-588.,* pp. 582-588, 2008.
[95] Vapnik V. The nature of statistical learning theory. Springer Science & Business Media; 1999.
[96] Wu L, Liu S, Yao L, Yan S. The effect of sample size on the grey system model. Appl Math Model 2013;37(9):6577–83.
[97] Despotovic M, Nedic V, Despotovic D, Cvetanovic S. Review and statistical analysis of different global solar radiation sunshine models. Renew Sustain Energy Rev 2015;52:1869–80.
[98] Dixit P, Londhe S, Dandawate Y. Removing prediction lag in wave height forecasting using Neuro-Wavelet modeling technique. Ocean Eng 2015;93:74–83.
[99] Qian Z, Pei Y, Zareipour H, Chen N. A review and discussion of decomposition-based hybrid models for wind energy forecasting applications. Appl Energy 2019;235:939–53.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

7.1. Conclusions

This work has resulted in five journal articles. The primary aim of this thesis is to address the problems of traditional statistical models while enhancing the prediction performance. To answer the research questions, three objectives were defined. For the first objective, problems that limit forecasting models' applicability are highlighted. Such issues include negative wind speed predictions, predetermined accuracy levels failure, non-optimal estimates, and higher computational cost with limited performance. To address these concerns, improved models are developed as second objective. Several methodologies are established to enhance the forecasting performance and fulfill the necessary and sufficient conditions. These approaches include adjusting dynamic moving window, self-adaptive state categorization algorithm, similar approach to leave-one-out method, and mixed initialization method. As a third objective, multiple models are compared, each for four case studies, three rolling windows, and three forecasting horizons and discussed outcomes of comparisons explaining higher prediction accuracy of improved models.

The novelty of the work and the original contributions are outlined in each chapter. Further, the summary of the novel research contributions is provided below:

- The first model enhanced in this study is GM (1,1). A comprehensive modified GM(1,1) model is based on the remnant model, L' Hopital's rule, optimal moving window, and the adaptive weighting factor.
- The second model considered in this study is Markov Chain. The enhanced model integrates MCs with an adjusting dynamic moving window, a similar approach to the leave-one-out method, and a self-adaptive state categorization algorithm.
- The third model improved in this study is Holt's Double Exponential Smoothing model. Again, the traditional model is enhanced by considering optimal parameters matrix, admissible constraints, and mixed initialization method.
- The hybrid models such as MODWT-ARIMA-Markov and AGO-HDES further improved the forecasting performance.
- Visual inspection revealed that the DES models have the least nMAE and nRMSE on average, followed by ARIMA and SVM models. However, the significantly larger

errors are witnessed for the Markov Chain model, almost double the errors of the modDES model.

- On average, modDES, modMC, and modGM show 4%, 25%, and 28% better performance than the counter traditional models. Also, errors are 18% and 32% more in 2- and 3-steps ahead.

Although the improved models showed higher accuracy than the conventional models, other challenges still need further research to enhance the model performance.

7.2. Future Directions

This research opened up several directions for future work. A few of them are discussed below.

The problem of exceeding prediction in GM(1,1) addressed in this thesis was due to the singular phenomena. However, another issue that causes the wind speed prediction to be negative is highlighted in chapter 6, where the larger rolling window and optimizing the hyperparameters do not solve the erroneous prediction problem.

The highest volatility is observed for the WFGM model. Four parameters need to evaluate the WFGM model. Applying the loss minimization function for a dynamic rolling window will significantly increase the computational time. Therefore, metaheuristic algorithms may be applied to minimize the computational cost.

The problem of trend lag is not eliminated entirely, and a complete study is needed in this regard. Proper utilization of decomposition models would address the issue of 'prediction lag' or 'timing error'.

The cyclic component is not considered in this study as we have a resolution of 10 minutes for small samples. Therefore, an adaptive higher order exponential smoothing model may also be considered in the future for low-resolution data to address the diurnal effects of wind speed.

We have considered three models in this study. However, it is recommended to further explore the other statistical models for forecasting performance enhancement.

APPENDICES

SUPPLEMENTARY MATERIAL

Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows

This file contains individual Taylor diagrams illustrated in Figure 8 of original article. Taylor diagram is a valuable visual tool for comparing several models. In terms of their correlation (R), root mean square error (RMSE), and standard deviation, the Taylor Diagram gives a concise statistical description of how well the models fit the measured data. A single point on a 2D polar plot represents the correlation coefficient, root mean square error, and standard deviation as:

$$E'^2 = \sigma_m^2 + \sigma_f^2 + 2\sigma_m\sigma_f R$$

where E' is the centered RMSE, σ_m is the standard deviation of measured values and σ_f is the standard deviation of the model field. The standard deviation is related to the radial distance from the origin, and the azimuthal position is proportional to a correlation between the two models. The measured field's point is shown on a horizontal axis corresponding to the standard deviation. The accuracy of the model is measured by the distance between the reference and the estimated model.

Readers are requested to consider the following legends for analyzing the Taylor diagrams

Models

○ Measured	◇ ARIMA	□ Brown DES	○ GM	○ $GM_{\alpha=adaptive}$	✡ modGM	+ modMC	▽ DT	◁ SVM
◇ ARMA	□ Holt DES	✳ Markov Chain	○ $GM_{\alpha=0.9}$	☆ WFGM	× modDES	△ GPR	▷ Ensemble	

Rolling Window

10	50	100
(purple)	(orange)	(green)

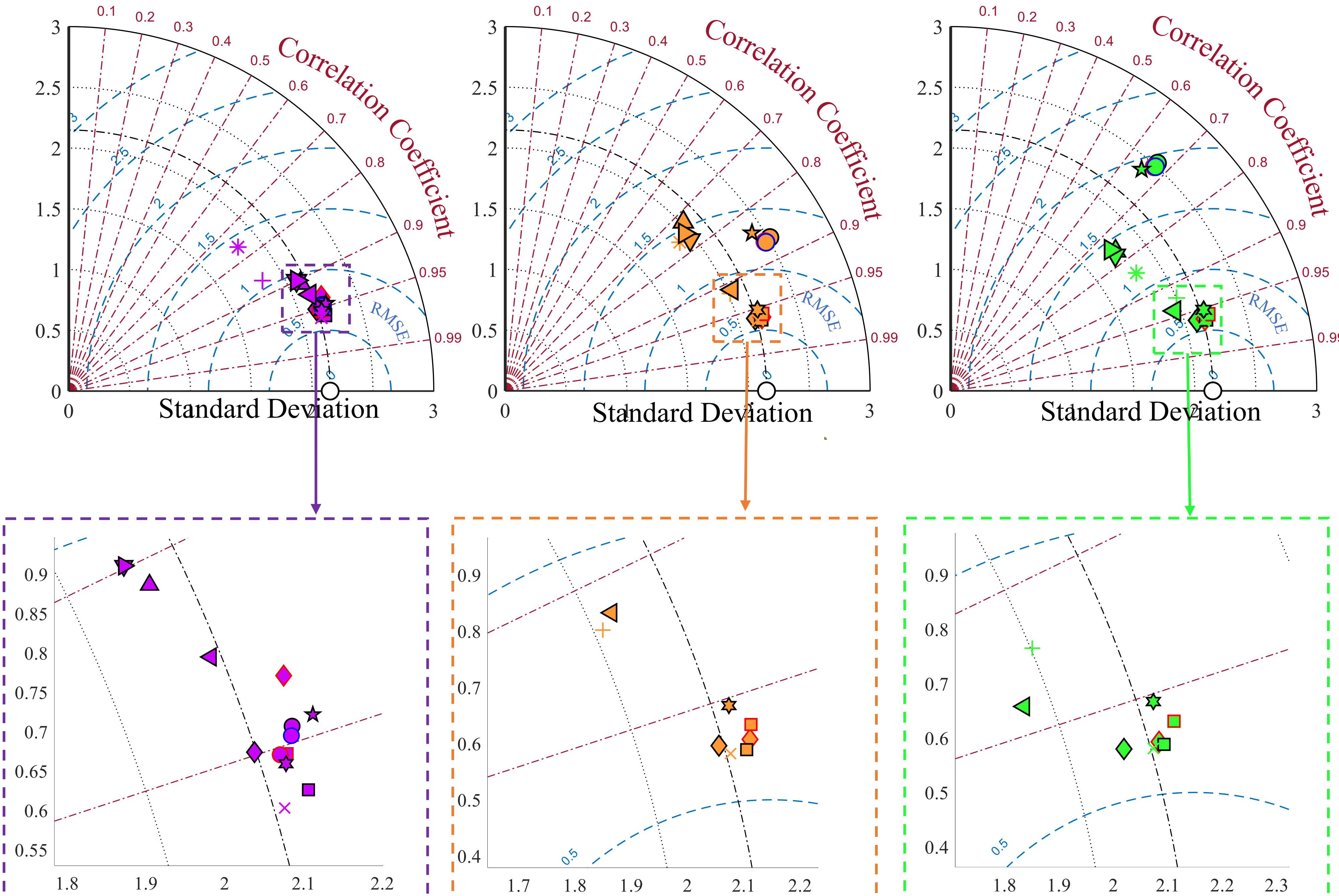


Figure A.1. Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows: Case study of **Palmerston North**.

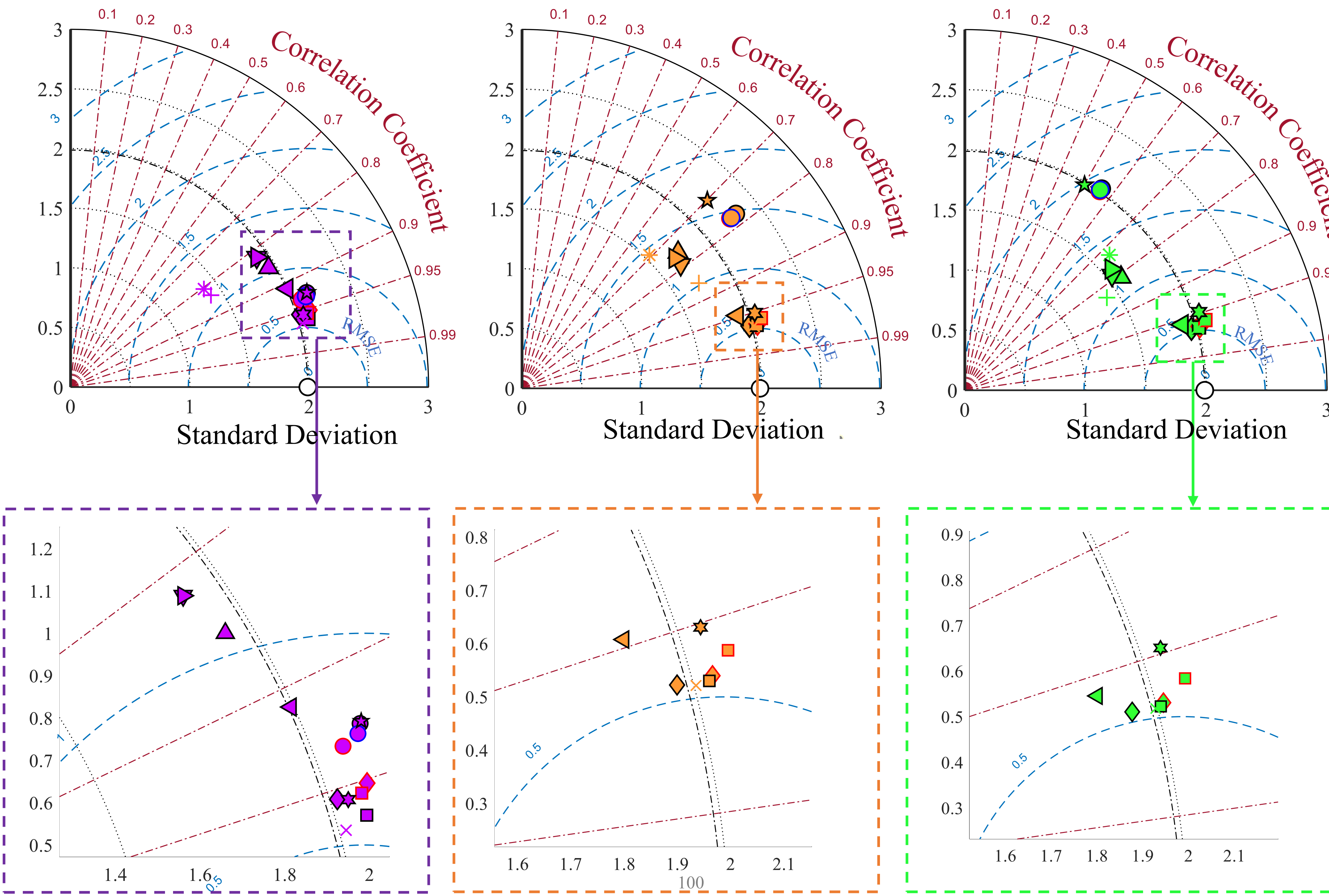


Figure A.2. Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows: Case study of **Gwadar**.

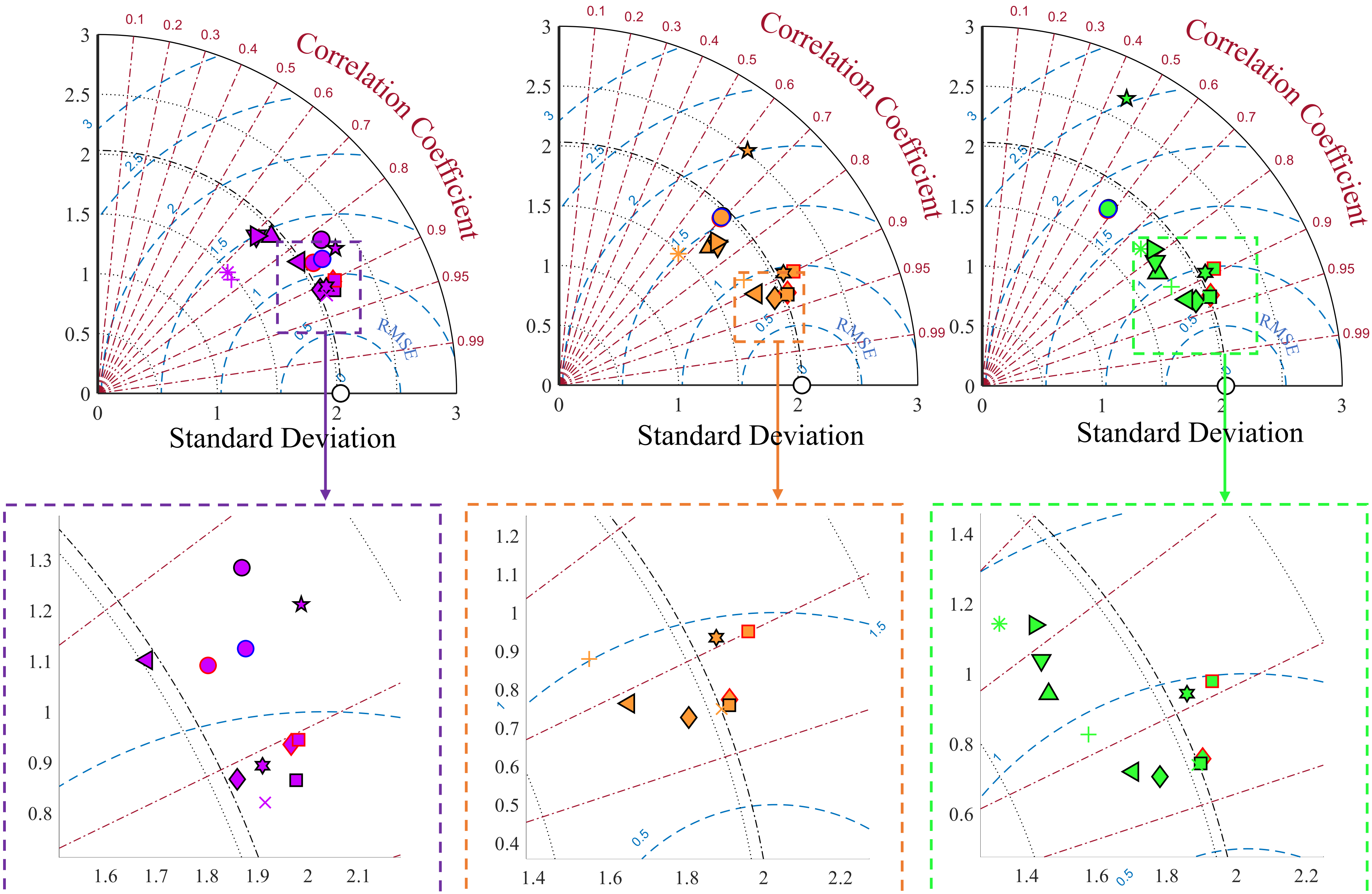


Figure A.3. Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows: Case study of **Mpika**.

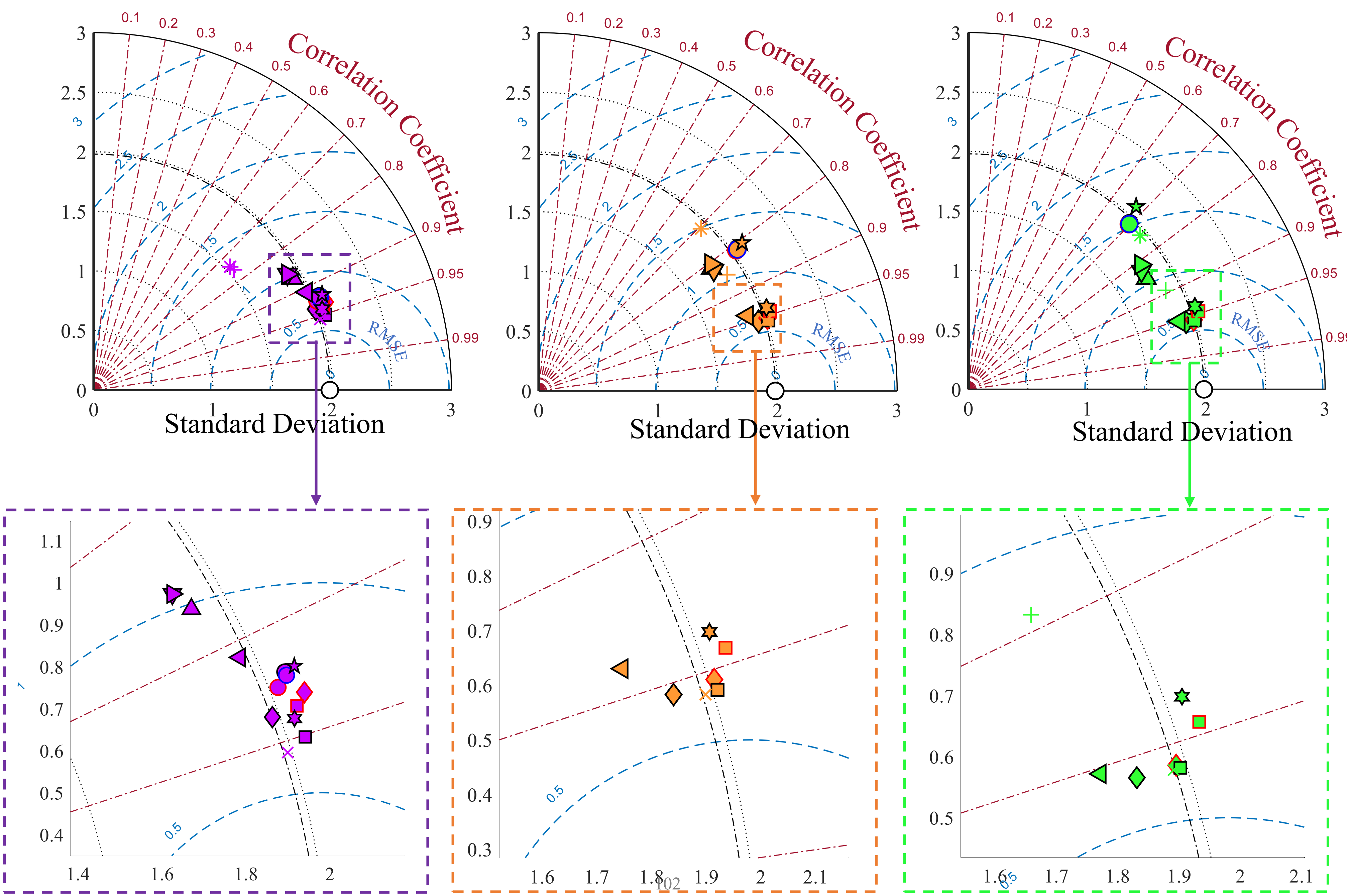


Figure A.4. Comparison of 16 models for one step ahead wind speed forecasting for three rolling windows: Case study of **Galicia**.

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Muhammad Uzair Yousuf
Name/title of Primary Supervisor:	Associate Professor Ibrahim Al-Bahadly
In which chapter is the manuscript /published work:	Chapter 2

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

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

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Muhammad Uzair Yousuf Digitally signed by Muhammad Uzair Yousuf Date: 2022.03.07 08:30:16 +13'00'
Date:	07-Mar-2022
Primary Supervisor's Signature:	Ibrahim Al-Bahadly Digitally signed by Ibrahim Al-Bahadly Date: 2022.03.08 09:12:55 +13'00'
Date:	8-Mar-2022

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10

Figure A.5. *DRC 16 - Chapter 2*

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Muhammad Uzair Yousuf
Name/title of Primary Supervisor:	Associate Professor Ibrahim Al-Bahadly
In which chapter is the manuscript /published work:	Chapter 3

Please select one of the following three options:

(◉) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

M. U. Yousuf, I. Al-Bahadly, and E. Avci, "A modified GM (1,1) model to accurately predict wind speed," Sustainable Energy Technologies and Assessments, vol. 43, p. 100905, 2021.

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Muhammad Uzair Yousuf Digitally signed by Muhammad Uzair Yousuf Date: 2022.03.07 08:32:38 +13'00'
Date:	07-Mar-2022
Primary Supervisor's Signature:	Ibrahim Al-Bahadly Digitally signed by Ibrahim Al-Bahadly Date: 2022.03.08 09:13:31 +13'00'
Date:	8-Mar-2022

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10

Figure A.6. *DRC 16 - Chapter 3*

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Muhammad Uzair Yousuf
Name/title of Primary Supervisor:	Associate Professor Ibrahim Al-Bahadly

In which chapter is the manuscript /published work: Chapter 4

Please select one of the following three options:

(◉) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Short-term wind speed forecasting based on hybrid MODWT-ARIMA-Markov model," IEEE Access, vol. 9, pp. 79695-79711, 2021

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Muhammad Uzair Yousuf — Digitally signed by Muhammad Uzair Yousuf Date: 2022.03.07 08:34:16 +13'00'
Date:	07-Mar-2022
Primary Supervisor's Signature:	Ibrahim Al-Bahadly — Digitally signed by Ibrahim Al-Bahadly Date: 2022.03.08 09:14:05 +13'00'
Date:	8-Mar-2022

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10

Figure A.7. *DRC 16 - Chapter 4*

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Muhammad Uzair Yousuf
Name/title of Primary Supervisor:	Associate Professor Ibrahim Al-Bahadly
In which chapter is the manuscript /published work:	Chapter 5

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Wind speed prediction for small sample dataset using hybrid first-order accumulated generating operation-based double exponential smoothing model," Energy Science and Engineering, 2022 (Early View)

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Muhammad Uzair Yousuf Digitally signed by Muhammad Uzair Yousuf Date: 2022.03.07 08:35:38 +13'00'
Date:	07-Mar-2022
Primary Supervisor's Signature:	Ibrahim Al-Bahadly Digitally signed by Ibrahim Al-Bahadly Date: 2022.03.08 09:14:37 +13'00'
Date:	8-Mar-2022

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10

Figure A.8. *DRC 16 - Chapter 5*

DRC 16

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Muhammad Uzair Yousuf
Name/title of Primary Supervisor:	Associate Professor Ibrahim Al-Bahadly
In which chapter is the manuscript /published work:	Chapter 6

Please select one of the following three options:

(●) The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

M. U. Yousuf, I. Al-Bahadly, and E. Avci, "Statistical wind speed forecasting models for small sample datasets: Problems, Improvements, and Prospects," Energy Conversion and Management, vol. 261, pp. 115658, 2022

( ) The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

( ) It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Muhammad Uzair Yousuf (Digitally signed by Muhammad Uzair Yousuf DN: cn=Muhammad Uzair Yousuf, o=Massey University, ou=Department of Mechanical and ELectrical Engineering, email=M.U.Yousuf@massey.ac.nz, c=NZ Date: 2022.06.13 20:43:22 +05'00')
Date:	13-Jun-2022
Primary Supervisor's Signature:	Ibrahim Al-Bahadly (Digitally signed by Ibrahim Al-Bahadly Date: 2022.06.14 09:22:45 +12'00')
Date:	14-Jun-2022

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

GRS Version 5 – 13 December 2019
DRC 19/09/10

Figure A.9. *DRC 16 - Chapter 6*