University of Nottingham
UK | CHINA | MALAYSIA

Data-Driven and Physics-Informed Wind Prediction in Offshore Wind Farms

Thesis submitted to the University of Nottingham for the degree of **Doctor of Philosophy, March 2023.**

Basem Elshafei

20223460

Supervised by

Donald Giddings
Atanas Popov
Xuerui Mao

Signature _____

Date _____ / _____ / _____

Abstract

Wind resource assessments have become a very critical topic in the construction of wind farms. The potential of a location providing electricity can be assessed using measurements by lidars or masts, which are a source of high-fidelity data, but are expensive and scarce in space and time, particularly for offshore sites. Contrarily, numerical simulations use software such as the Weather Research and Forecasting model to generate temporally and spatially continuous data with relatively low-fidelity. A wind speed time series generated using either method is considered as signal with non-stationarity and non-linear characteristics. Therefore, pre-processing of signals using signal decomposition is proposed for dealing with uncertainty in wind data. However, due to the high frequency modes of the decomposed signals, single model decomposition is no longer sufficient to obtain accurate results, and secondary decomposition models are employed. Processed signals are then fed into a constructed forecasting model, which consists of relationship learning algorithms. Finally, parameters of the learning process are optimized to make optimal predictions.

In this work fusion of data is tested by employing machine learning algorithms to forecast wind speed data. Three models are proposed, two that consider data fusion of data from different sources, and one that only uses simulation data, used as control model to assess the significance on

performance, when using different sources of wind speed data of different fidelities. The numerical and measured wind speeds along the west coast of Denmark are used to evaluate the methods.

The first model, combines the merit of measurements and simulations for the assessment of offshore wind, following a two-step hybrid model. Firstly, a temporal data fusion of data using multi-fidelity Gaussian process regression is employed, which combines the intermittent measurements and the continuous simulation data of the onshore location. Results are then propagated from one onshore location to an offshore location, following a spatial data fusion of data using a non-linear autoregression with external input. The proposed data fusion technique using gappy onshore measurements results in accurate offshore wind resource assessment within a 2% margin error.

The second model proposes a hybrid approach which combines the merit of two decomposition algorithms, namely complete ensemble model decomposition with adaptive noise, and empirical wavelet transform. The wavelet transform approach is employed to further decompose the high frequency signal from the model decomposition algorithm, thus reducing forecasting complexity. Then, an improved bidirectional long-short term memory neural network is optimised using the grey wolf optimization algorithm to forecast all the decomposed signals. Results are then compared to lidar measurements to assess performance of models with only numerical simulations as input data. It is shown that forecasting models that consider multiple sources of data with different fidelities performs more efficiently when compared to the lidar measurements. Therefore, the presence of higher fidelity data, though scarce and limited, promises enhanced assessment of wind resources.

The third model proposes a hybrid approach that investigates using multiple data points for wind speed prediction both in spatial and temporal domains simultaneously. Firstly, the measurements and numerical datasets are fed onto a sparse matrix, where the columns represent the spatial lidar and WRF points, and the rows represent the time steps. Entries of the matrix reflect the wind speed at a given time and location, contrarily, empty entries reflect a time step where data is not observed. A non-linear probabilistic matrix factorization using Gaussian process model is used to train and test for matrix completion, which fills the missing data with predictions. The proposed data fusion model, using gappy measurements, results in accurate wind resource assessment with matrix completion results of higher accuracy than industrial and academic models, with 58% and 40% improvements, respectively.

Acknowledgements

Completing a PhD is a journey that cannot be accomplished alone, and I am grateful to the many people who have supported me throughout this journey. Their guidance, encouragement, and unwavering support have helped me reach this significant milestone.

Firstly, I would like to express my deep gratitude to my family. I am thankful to my parents for their unwavering support and for instilling in me the importance of education. Their love and encouragement have been instrumental in helping me navigate the ups and downs of graduate school. I also extend my heartfelt thanks to my father, Mr. Moha for his financial support, which made it possible for me to pursue this PhD. I am truly fortunate to have a family that has always believed in me and has supported me through thick and thin.

I also want to acknowledge the support of my supervisors, Professor Donald Giddings, Professor Atanas Popov, and Professor Xuerui Mao. They have been my source of motivation and inspiration throughout this journey. Their valuable insights, constructive feedback, and encouragement have played a critical role in shaping my research and enhancing my overall academic experience. I am grateful for the friendship and camaraderie we have built, and I look forward to many more years of collaboration and support.

I am also indebted to my collaborators. Their expertise and insight have helped me navigate the complexities of my research project. I am grateful for their generosity with their time and resources, which have been invaluable in helping me achieve my research goals. I have learned so much from them, and their contributions have undoubtedly enriched my academic experience.

Lastly, I want to express my deepest appreciation to my partner. Her unwavering love and support have been my constant throughout this journey.

She has been my sounding board, my cheerleader, and my confidante. Her presence in my life has made this journey enjoyable and memorable. I am lucky to have her by my side, and I look forward to many more years of adventure and growth together.

In conclusion, I would like to thank everyone who has been a part of this journey with me. Your support, encouragement, and guidance have been invaluable, and I could not have done this without you. I will always be grateful for the love, kindness, and generosity that you have shown me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

A major global challenge is climate change, which scientific evidence over the past decades proved related to human activity [1]. Unprecedented records of average sea surface temperatures were reported in the past 60 years, with continuous global sea levels rising and sea ice levels receding [2]. Since the industrial revolution, the increased consumption in fossil fuels has resulted in increased release of CO_2 and other green house gases in the atmosphere. As a result, the world is currently observing drastic natural disasters in every corner of the planet, and if nothing is done to reverse the release of greenhouse gases, more significant disasters will be observed, through extreme weather changes, and food and water supplies, causing significant impact on human life. It is impossible to fully predict the trajectory of consequences of climate change due to current electricity generation procedures, but a cut to reduce the harmful emissions is critical to reduce the impacts [3].

Historically, electricity is generated in central plants in locations near the available resources and industrial areas, and is then transported to customers and areas of need through distribution and transmission networks. An important advantage of wind energy is the feasibility of building wind farms at various locations, thousands of wind farms are being installed at various locations, changing the traditional centralised systems to a more broad and distributed power generation one [4]. Another factor that heavily defines electricity is its highly perishable commodity, as electricity must be used simultaneously as it is being generated. The fact electricity can not be easily stored directly makes it obligatory to maintain balance between power generation and consumption at all grids at any time. Through forecasting the capacity and capability of a wind farm location, and assessment of the wind resource, generated electricity can be linked to the surrounding demand, where power supply follows the demand required. In a relative manner, wind energy varies from other energy sources as it can not be scheduled and is merely controlled by the stochastic nature of wind. Thus, wind power generation must be managed where turbines are allowed to operate to generate enough electricity to meet foreseeable demands.

The aforementioned consequences and fear of continued damage to the planet have promoted the replacement of fossil fuels with renewable sources of energy for electricity generation. A strong interest is now visible in the renewable resources, with wind and solar leading the way. Renewable energy is on course to break global records, with unprecedented growth in capacity additions, to seek the advantages of energy security and climate benefits. Specifically, wind is becoming the major pillar of electricity generation in Europe, with estimated 35% of Europe's grid power coming from clean, renewable wind farms. Being one of the fastest growing energies, wind energy is becoming very economically viable, causing reductions in

operation costs. With further utilization in the wind power generation industry, the cost of electricity generation is approaching the conventional energy, giving it the ability to compete with traditional power plants [5].

However, a major challenge arises when integrating wind power into the grid, which is the intermittent nature of wind that makes it not possible to control the output power generated by the wind farms [6]. As wind can not be scheduled or delivered when required, wind power generation is considered a stochastic source of energy. This stochastic behaviour of the wind increases the complexity of regulating the power system and reserves standards that allow for stability and reliability [7]. It is therefore critical to to manage the intermittency of wind as it has greater impact on the security and impact of the grid, and market economies. In [8], the study presents that a variability or uncertainty of 20% in wind, would have a significant increase on the operating costs such as the commitment cost. The variability and uncertainty of wind make it hard to control the output power, and hence the costs. Therefore, more power system regulations are required to ensure stability and no additional costs. The most effective solution is to estimate the production of output power from operating wind farms. Wind power is controlled by multiple factors, of which wind speed is the most important. Hence, estimation of future wind power can be generated from wind speed forecasting. Once reliable forecasts of wind speeds are generated for future time horizons, utilization of wind power and electricity generation are obtained, thereby, electricity variability and uncertainty can be relieved [9].

1.2 Research Motivation

Due to the high stochasticity and low predictability of wind, wind energy is considered an intermittent source of energy [10]. This nature of wind influences the stability and safety of the large scale grid integrated power systems. Additionally, wind power is affected by distribution losses and transmission [11]. Therefore, for higher penetration of this source of energy, scheduling, management and optimisation are challenges that should be addressed. A number of suggestions can be considered to decrease the uncertainty in power systems, which include wind power forecasting [12]. Operators need to forecast future values of the energy, as wind power can not be controlled. In addition, since the output power from a wind farm's turbine and the wind speed have a cubic relation, simple speed forecast errors can lead to large cubic errors in output wind power. Therefore, accurate and consistent forecasts of wind speed, are crucial for reliable wind farms and electricity grids, the result is wind power being integrated into the dispatch and scheduling systems of different scales for decision making [13]. Subsequently, the regulation of wind systems is promoted and the operational efficiency of the industry is enhanced, which lead to reduced operating costs and enhanced competitiveness of wind power. The study of wind speed forecasting began immediately with the innovation of generating power from wind, as it has been a challenging yet essential requirement for the investments in more wind farms. The literature for wind speed forecasting techniques present major improvements being recorded, as different models and techniques are more appropriate for different applications and use cases where there exists different circumstances.

The demand on wind farms and wind power continues to rise as the need for more renewable energy increases [14]. Besides providing smoother in-

tegration of wind energy to the electricity system, wind power or speed forecasting provides additional benefits that make wind energy more viable in a cost effective manner. For example, specific to the UK, wind power forecasting limits the impact of excessive use of wind power in the grid, by scheduling the generation units effectively to meet the demand. In addition, the amount of electricity considered of excess capacity could be set to ensure a continued reliability of delivery of electricity. Additionally, with monitored power capacity generated from wind farms, operators have the capability to compete in the market and meet changing demands through planning ahead, which returns more investments and trust in the wind energy sector [15].

This research aims to develop multiple models for enhanced accuracy of wind speed predictions. Short-term wind speed forecasting is the main targeted time horizon of the developed models. Forecasts for this time horizon are typically acquired by modelling the trends and patterns of recent past measurements of the wind speed using numerical and statistical models. In the next chapter, numerous models are described and discussed with comparison to their merits and limitations for different applications. However, data fusion of wind speed measurements, where the observations are generated using different techniques as input, is under-developed and promises several benefits. The fusion of instrument measurements with numerically simulated data encourage more precise and rich predictions that are more accurate than simulations and continuously available at any given spatial and temporal point of observation at reduced cost with the addition of a trivial computational cost. In addition, pre-processing of the wind speed time series and taking into account additional side information such as magnitudes and derivatives of the wind speed time series is often overlooked. Moreover, the addition of spatial techniques captured hidden

spatial correlations that is shown to improve the accuracy of predictions. Furthermore, by studying different techniques of pre-processing, forecasting and optimizing the prediction models, along with providing additional sources of data generated using different techniques, it is demonstrated that there is a wide availability and scope for significant improvements in the short-term wind speed forecasting field.

However, wind speed forecasting does not provide direct solutions to the wind power sector challenges, they are instead used as a major contributing factor to the various decision making processes to integrating the wind power sector to the power generation market in a cost-effective and more reliable way. Therefore, improving the quality of predictions is set as a priority in wind energy research.

1.3 Thesis aims and objectives

1.3.1 Aims

The aim of this thesis is to use surrounding onshore measurements and offshore simulations to estimate accurate continuous offshore data without sending expensive equipment offshore. A range of combinations of signal processing, non-linear algorithms and neural networks are investigated to assess the influence of pre-processing time-series and multivariate data fusion and to further enhance the accuracy level of wind speed forecasts.

1.3.2 Objectives

1) Investigate the merits of data fusion on wind speed time-series predictions, by combining high-fidelity yet intermittent measurements from instrumental equipment, such as Light Detection And Ranging (Lidar) with low-fidelity but continuous Numerical Weather Prediction (NWP) simulations such as Weather Research Forecast (WRF).

2) Develop hybrid prediction models that consider additional side information often overlooked, such as wind speed magnitudes and derivatives.

3) Investigate the influence of pre-processing time-series using signal processing techniques, such as Empirical Wavelet Transform (EWT) on accuracy of wind speed predictions.

4) Reduce the cost of offshore wind resource assessment by investigating correlation between wind speeds at onshore and offshore locations, then carrying predictions from onshore to offshore locations depending on data fusion of simulations and onshore measurements without sending measuring instruments offshore.

5) Investigate the merits of employing deep machine learning models and the effect of tuning their hyper-parameters on accuracy of predictions by developing a novel model and assessing the predictions against statistical and persistence model results.

6) Investigate an advanced forecasting model with combined pre-processing of data and optimization algorithm using simulation and compare results with models that use both lidar observations and WRF simulations.

7) Develop a three dimensional space-space-time model that captures spatial and temporal features of wind speed at a wind farm location that

correlates latent spatial and temporal features to a single dimension.

1.4 The novelty of this research

1.4.1 Contribution

• A model of wind evaluation in offshore wind farms established through a data-based and physics-informed approach, where results are compared to previous industrial and academic standards.

• Develop a novel algorithm to merge limited measured data and continuous simulation results, then propagate predictions from onshore to offshore sites.

• Develop a novel hybrid wind speed forecasting model, using a preprocessing algorithm, a forecasting algorithm, and an optimization model using advanced hyper-tuned recurrent neural networks.

• Develop a recommender system for spatio-temporal data fusion of several spatial points and several time steps to estimate missing data in a space-time dimension.

1.4.2 Outline of the dissertation

This thesis consists of 6 chapters. This chapter provided an overview of the background of the work, followed by the challenges in the field, suggested solutions, significance and motivation of the research, aims and objectives of the research, and finally the structure of the thesis.

Chapter 2 aims to provide an extensive review of the literature related to

the concepts of the research, starting with the state-of-the-art in the power forecasting field. It discusses the basis of the forecasting problem, the nature of wind power forecasting, types of predictions, and the most recent wind forecasting methods. This is followed by an assessment of previous models and the novelty and contributions they provided.

In Chapter 3, a presentation of the first novel hybrid model is provided. This model investigates data fusion of multi-fidelity data, pre-processing of time-series using Empirical Wavelet Transform (EWT), and the addition of side information such as magnitudes and derivatives of the wind speed time-series. Furthermore, it assesses the use of onshore lidar measurements in addition to offshore simulations to produce a continuous offshore time-series that is more accurate than simulations and less scarce than measurements.

Chapter 4 aims to assess the significance of data fusion by proposing a hybrid model that uses a secondary decomposition model to pre-process the data. The generated sub-series are then fed into an optimized neural network to generate predictions. The performance of the proposed model is assessed and compared to the results from models in Chapter 3.

In Chapter 5, the use of a recommender system is demonstrated to build a model that relates the latent features of wind speed in both space and time. This model takes into consideration pre-processed data and several further distant lidars for the prediction of wind speed. It generates predictions at multiple locations and time steps in a single iteration.

Chapter 6 summarizes the research conducted with a recapitulation of what has been achieved. It provides an overall discussion of the findings, explores the research implications and limitations, and offers suggestions and scope for future work.

Chapter 2

State-of-the-art in Wind Power Forecasting

This section aims to give a brief overview of the current research results in the field of wind power forecasting. The main objective of this chapter is to introduce the different aspects that establish wind power forecasting and demonstrate different research paths that have been explored. Thus, this chapter, can be regarded as a short summary of the base knowledge that this thesis was built on. The illustrations made will help contribute to a better understanding of the methods and approaches presented in this work.

Wind power forecasting is traced back to the early 1980s, it was perceived as a key element in operating wind power plants. Of the very first attempts, a discussion group at the pacific Northwest Laboratory carried out tests to clarify the importance and advantages of short term predictions to the electricity industry. The conclusion indicated that following a sufficiently reliable forecast can have three primary applications: weekly forecasts (daily rates) of winds for maintenance scheduling; daily forecasts

(hourly rates) of wind levels are factored into the load scheduling strategy; and hourly forecasts of likely winds for dispatching decisions [16]. Another good example of early wind forecasting is [17] where a wind turbine power curve and wind speed forecasts were used to understand how wind can contribute to the future of energy networks. At this early stage, it was recognised that "once a wind power generator is supplying power to an energy system, a method of forecasting wind power a few hours in advance is required to ensure efficient utilization of the power." Over the following 40 years, the research and development area in this sector expanded drastically, most significantly since the 2000s, due to recent and continued increase in wind power demand around the world.

Hence, wind forecast has been regarded as a high priority research area, where the running costs and reliability of the implemented systems are the main targets of improvement. Wind power plays a pivotal part in global energy production as it is clean, emission-free, and relatively cheap in the production cost; however, the nature of wind is stochastic and intermittent, making economic scheduling and dispatching, and planning the unit commitment very challenging. Here comes the paramount demand for advanced wind forecasting techniques over broader time horizons.

The basis of the forecasting problem will be outlined in section 1, while section 2 will discuss 2D interval predictions for a time series with different time horizons and the difference between point predictions, interval predictions and probabilistic predictions. Section 3, addresses the different classifications of wind speed forecasting. Additionally, section 4 will give a brief summary of the pre-processing approaches used in wind speed forecasting followed by a summary of the wide range of wind forecasting methods such as: persistence, statistical methods, machine learning methods, physical models and hybrid models in section 5. Then, in section 6, a

discussion on the different algorithms used in optimizing forecasting models is provided. Finally, section 7 concludes the current state of wind speed predictions and provides a brief overview of where the advances in this field are required.

2.1 Basis of the Forecasting Problem

2.1.1 Basic concepts

Wind, or in other words, the air flow around us, is more than just wind particles. It can be divided into three subcategories: mean wind, turbulence and air waves. The three categories can occur both separately or infused onto each other (super-imposed). The movement of moisture, heat and momentum quantities is dominated by mean wind horizontally and turbulence vertically. Multiple highly complex processes exist in the atmosphere, which drive many observable phenomena, thus exists theories, which describes it extensively with complexity. The behaviour of the atmosphere is governed by seven variables as functions of both time and space: temperature, pressure, moisture, density, two horizontal velocity components, and a vertical velocity component. Additionally, these variables are governed by seven equations: the first law of thermodynamics, the equations of state, the continuity equations for mass and water substance and three components of Newton's second law [18]. The mentioned equations of motion contain both space and time derivatives and require initial and boundary conditions for their solution. Since motion in the atmosphere is considered relatively slow, (velocities are considerably lower compared with the velocity of light); the Galilean/Newtonian paradigm of classical physics could be applied, where instantaneous propagation of interactions is assumed and

the concepts of absolute time and absolute space are embodied. However, the complete set of equations is too complex for an analytical solution, hence, in the meteorological field, i.e numerical weather prediction models, the equations are simplified with utilization of the parameterizations and approximations.

Turbulence

A critical weather phenomenon and one that is significantly unpredictable is turbulence. The generation of turbulence originates due to forces at the ground, where fractional drag on the flowing air develops wind shears causing irregularity in motion of air particles. As a result of irregular motions of the air, vertical currents and irregular swirls of motion called eddies are created. The eddies of a turbulence can be of different sizes, super-imposed on each other, and influence the degree of stability of the air, where turbulence can be classed as light, moderate, severe or extreme. In severe or extreme turbulence, the eddies are as large as hundreds of kilometers in diameter, and may last for months to years. They are very intense and may cause structural damage. Other moderate and light turbulence cases consist of smaller eddies, i.e the swirls of leaves or wavy motions, and are in the order of few millimeters and may last for a matter of seconds [19].

The scales of motion in the atmosphere influence the application of the aforementioned equations of motion, as the range varies for large eddies of thousands of kilometers down to small ones of millimeters, the direct application of equations is impractical. As it would require millimeter spatial and second fractions of a second temporal resolution observations, which is not possible. A cut-off scale where the influence of turbulence is treated statistically can be selected based on the application. For numerical

weather prediction model applications, the cut-off is on the order of 10 to 100 kilometers, while other applications such as boundary layer models, or large eddy simulation models, the cut-off drops significantly to be in the order of a few hundred meters [18].

The boundary layer and vertical profiles

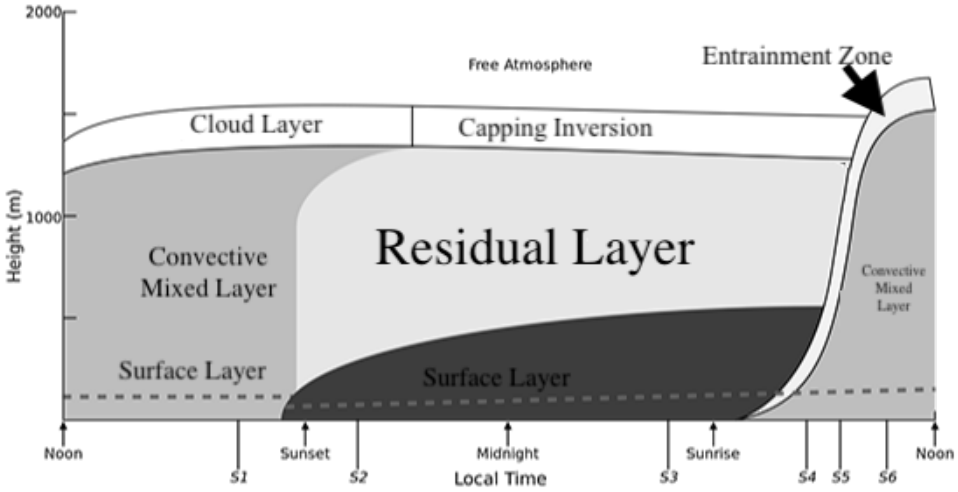

Figure 2.1: Illustration of the boundary layer w.r.t time and height [20].

Figure 2.1 captures the layers and characteristics of the boundary-layer over a high-pressure region of land during the day. There are three main layers in this well-defined structure of a boundary layer: the stable layer, the residual layer, and the mixed layer. At the bottom of the boundary layer, exists the surface layer, which is a region where turbulence varies by less that 10% of the total magnitude. For the mixed layer, the turbulence is generated by driven convection and wind shear coming from the top of this layer. In addition, the less turbulent air at the top also mixes down with the mixed layer, increasing the layer's height. The stable layer acts as a lid to the rising thermals on top of the mixed layer, where the absolute temperature increases with height. Moreover, around sunset time, the turbulence intensity reduces in the well mixed layer, where the mean state

variables form the initial variables of the residual layer. At night, the contact between the residual layer and the ground, converts the residual layer into a stable boundary layer, which is characterised by weak turbulence and stable air [20].

Turbulence is what causes the movement of atmospheric constituents, so accurate parametrizations of their vertical profiles dependent on turbulence intensity are essential for closure methods in governing equations. As a result of atmospheric stability, which determines vertical mixing within the layer, there can be significant variations in wind speed with respect to height. Large differences in atmospheric state at different heights cause low wind speeds close to the ground and high wind speeds above. Low wind speeds across all vertical layers are required to maintain atmospheric stability because the shear between these wind speed layers causes turbulence. Increasing turbulence, on the other hand, improves vertical mixing and lessens the dependence of atmospheric state on height [21].

Wind speed predictions are significantly affected by changes in boundary layer turbulence. By introducing vertical variations in wind speed and altering the overall wind profile, turbulence speeds up the mixing process. It results in wind shear, which alters speed and direction as altitude changes. Microscale variability is another aspect of turbulence that is introduced, and it affects regional flow patterns and turbulence intensity. Turbulence changes have an impact on the vertical extent and structure of the boundary layer by altering the distribution and vertical profiles of wind speed. Taking into account atmospheric stability, surface characteristics, and other pertinent factors, turbulence parameterizations must be incorporated for precise wind speed prediction. To verify and improve wind speed predictions, especially when there is turbulence, measurements from meteorological instruments, such as anemometers or remote sensing devices, are

frequently used.

2.1.2 Nature of wind power generation

After briefly describing the nature of wind, an introduction on how energy can be generated from it is covered. Wind farms use wind turbines to capture the wind and convert kinetic energy into electric energy, in other words, power. The wind speed directly regulates and controls the amount of power the turbine produces.

The power curve is given by the following power equation:

$$P = \frac{1}{2}\rho\pi R^2 v^3 C_p \tag{2.1}$$

where ρ is the air density, R is the radius of the rotor, v is for velocity of the wind speed, and since it is impossible to extract energy from the moving air without dissipation, C_p is the Betz's limit, equivalent to 16/27 (approx. 0.593). In practice, the efficacy rate reaches 0.52 - 0.55, but after taking into consideration the mechanical and electrical losses in the gear and generator, the rate drops to 0.46- 0.5 [22].

As previously stated, wind is the motion of particles that comprise the atmosphere, hence, the speed of wind can be measured by calculating the speed of the particles. However, this is not possible in practice, as it is very difficult to measure particles per second interacting with a wind turbine rotor blade. As interest is only in the power that could be generated, the speed seen by the rotor becomes our only concern. Wind speed can be measured by cup anemometers, a device present at most weather stations that is robust and cheap, which was used to generate most of the historic

data sets of wind speeds. But this device suffers delays in responding to changes in wind speed and have a higher precision in measuring increases in the wind speed compared with decreases, which results in over-estimated wind speeds. Other devices such as Light Detection And Ranging (Lidar) and sonic anemometers are more accurate but come at an expensive price and unavailability in space and time domains [22].

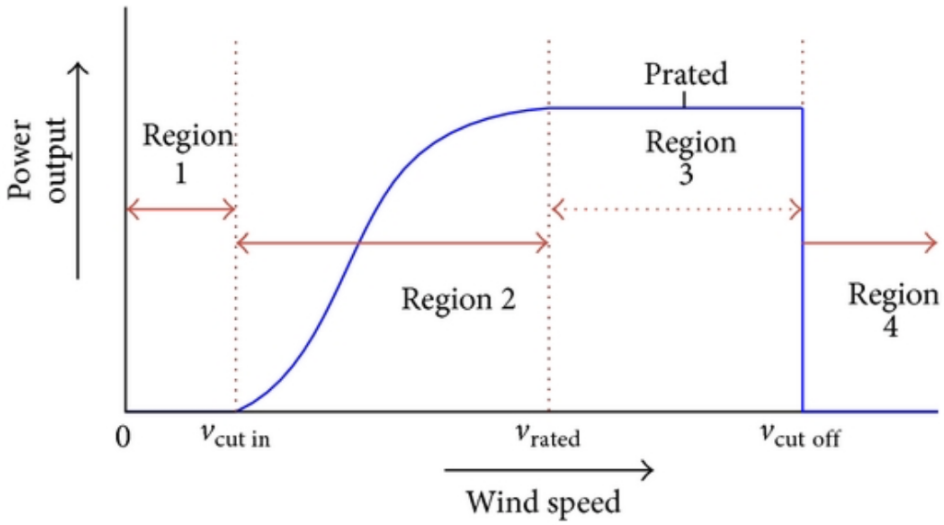

Figure 2.2: Wind power - speed curve [23].

Once the wind speed time-series is measured or estimated, a wind turbine power curve can interpolate the output power. The power curve describes the direct dependency between speed and power, since power is the cube of speed. The amount of power generated by the turbine depends on multiple considerations: the generator, installed power electronics of the turbine, power of the air flow, and the efficiency of the conversion process. Despite that, most power curves are very similar, as they are governed by the same laws of physics. The power curve of a modern, variable speed turbine can be split into four distinctive parts: 1) Below cut-in speed: wind speed is too low to operate turbine and no power is produced, 2) Between cut-in and rated speed: speed reaches cut-in value and power productions starts, 3) Rated to cut-off speed: a constant power production dependent

on the turbine's rated value, and 4) Above cut-off speed: no production occurs as turbine is switched off for safety, to prevent any damages to the rotating parts. For the wind between the cut-in and cut-off speeds, the physics of the energy conversion process can explain the variation in power production. However, uncertainty exists within the curve, as the conversion process is affected by multiple external factors, including the blade erosion, mechanical wear and quality of the measured wind. In addition, the curve drawn in Figure 2.2 is for an idealised situation, typically produced by a manufacturer, where the turbine was not exposed to any turbulence, obstacles, or abnormal air pressures. The reality of air behaviour is much more complex due to more external factors such as shadowing effect, terrain and topography of the wind farm which all have significant implications and influence on the curve [23].

2.2 2D-Interval Predictions for Time Series

2.2.1 Time series

In prediction, it is of great importance to understand the characteristics of the dependent variables, and the different scales at which the prediction process could be tackled. In wind speed predictions and specifically for our application, the effect of spatial and temporal parameters and their scale is considered of significant importance. Wind speed may be estimated at a specific given moment of time, or averaged over a time-period (10 minute and 1 hour intervals), the spatial coordination of the point of interest such as height, longitude and latitude, all have a considerable influence on the predictability of the resulting time-series. In a time-series, the notion of order (time of observation) distinguishes the analysis and relates the pro-

vided variable (X) data points (value at time t: x_t) to one another in a sequence of measurements collected at a constant time interval (t) [24].

$$X = x_t, x_{t+1}, x_{t+2}, ...|t \in N \tag{2.2}$$

Once patterns in the time-series are identified and correlated, underlying phenomena with respect to the variable of interest are explained, and missing values or future values of the time-series can be found (interpolation and extrapolation forecasting).

Adapting and adjusting the spatial and temporal scales helps accomplish the variety of industrial interests in wind speed prediction. Figure 2.3 provides the forecast time horizons along their common temporal resolutions and their industrial use (informed decisions).

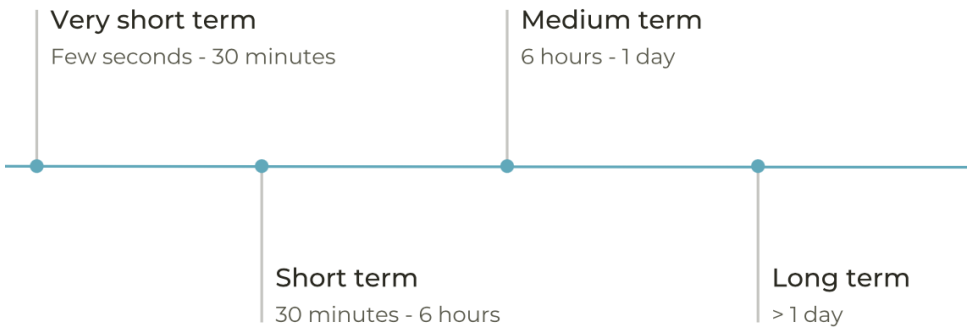

Figure 2.3: Time horizons for wind speed forecasting.

The prediction horizons discussed in Table 2.1 are a paramount factor in the choice of the forecasting technique. The proper classification of the prediction horizon is specially important for various operating systems in wind forecasting applications, such as: electricity market, dispatch and schedules of the economic load, and ancillary support. Once the classification of the prediction horizon is decided based on the application type, the most accurate and applicable method of forecasting can then be selected. Due

Table 2.1: Categorization of forecasting methods based on prediction horizon [25].

	Forecast Horizon	**Resolution Interest**	
Ultra-short term	$<$ 1 Minute	Seconds	Turbine control and load tracking in real-time.
Very-short term	$<$ 1 Hour	10 Minutes	
Short term	1-48 Hours	30 Minutes, 1 Hour	Load dispatch planning
Medium term	1-10 Days	1 Hour, 3 Hours	Utilized for energy trading and power system management
Long-term	Months-Years	Days-Months	Guide optimal maintenance scheduling

to their accuracy and robustness, short-term and medium-term predictions are the most engaged methods in the literature. To have accurate forecasting of future values, it is important to evaluate the performance of the method at sites with different atmospheric conditions.

In the very short-term prediction horizon methods, the future predictions of values is in the span of few seconds to half an hour. This method is used extensively in market clearing and regulatory actions. Methods of this criterion include the spatial correlation method [26], and the artificial neural network-Markov chain model [27], where predictions are made for 1-second ahead and 7.5-seconds ahead, respectively. Moreover, the Bayesian structural break model [28], and data mining approach [29] can forecast predictions of 1-minute to 1-hour ahead and 10-seconds to 5-minutes ahead, respectively. In addition to these methods, the use of intelligent machine learning methods, such as Deep Neural Network (DNN) is rapidly growing and becoming more approved with recent advances. Another method which

form the core of the very short-term methods include, Support Vector Machine (SVM), and hybrid methods that combine neural networks with preprocessing techniques, such as Empirical Wavelet Transform (EWT). The Decomposition Forecasting Algorithms (DFA), which breaks the time-series of the targeted variable, wind speed / power, to analyze individual units and generate a resultant series. One of these decomposition algorithms [30] broke down / decomposed the time-series into multiple units, then feature construction is performed at each unit to select the best features for predictions, which can be carried by the previously mentioned methods, such as Artificial Neural Network (ANN), Support Vector Machine (SVM) or an Auto-regressive Moving Average (ARMA) model. In addition, decomposition techniques are also used for non-stationary and non-linear forecasting models, i.e [31] which is based on Hilbert-Huang Transform (HHT).

In [32], a neural network was trained with 12 hours previous data (15-minute time-step) as input, Wavelet Transform (WT) was employed to decompose the series and the resultant was also fed to the neural network. The model provided predictions for a 3-hour ahead time span. Another short-term forecast method involves the Empirical Wavelet Transform (EWT) and Nonlinear Autoregression With External Input (NARX) to forecast, the result showed full superiority of machine learning and neural networks compared to other forecasting methods [33]. In [34], the concept of genetic algorithms was used to optimize the input parameters for and Support Vector Machine (SVM) model with wavelet transform preprocessing. The results of these predictions showed that the Mean Absolute Percentage Error (MAPE) obtained is around 15% while that with persistence is 23%.

Contrarily, long-term predictions are also essential for major applications such as unit commitment decisions and maintenance scheduling. In [35],

1-day forecasts are obtained by combining the merits of Particle Swarm Optimisation (PSO) and Adaptive Network Based Fuzzy Interference System (ANFIS), where the PSO algorithm estimates the best parameters for neuro-fuzzy systems. A different method was approached by [36], where they used Relevance Vector Machine (RVM)), Wavelet Transform (WT), and Artificial Bee Colony Optimization (ABCO), the wind time-series is decomposed into different frequency ranges, and ABCO uses a meta-heuristic algorithm to estimate the most suitable parameters for the RVM kernel, which then carries the predictions.

2.2.2 Point Prediction

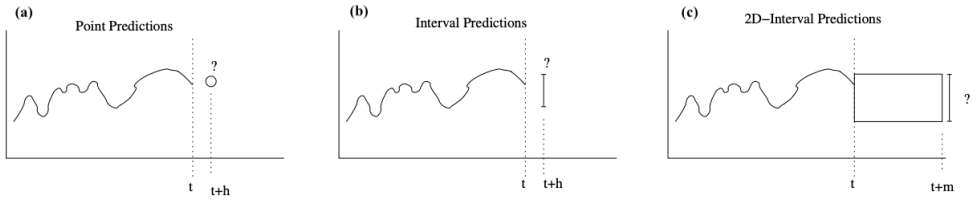

Figure 2.4: Types of time series prediction, namely (a) point, (b) interval and (c) 2D interval prediction, respectively. [37]

Point prediction is the simplest yet most common form of forecast in time-series analysis, they are easy to handle during decision making times. They target a specific single instant of time and predict the observation at it exclusively. Due to its simplicity, many practitioners favour point predictions, i.e a non-expert weather forecaster can easily generate, communicate and interpret weather point forecasts because of point predictions. The approach for point prediction is mainly mentioned in the time-series analysis, the previous (past) relations between variable points (i.e x_{t-1}, x_t), which occur at different times, is studied and used to estimate the future unobserved point (x_{t+1}), where 1 here represents the forecast time horizon one step ahead of the current time, check Figure 2.4 panel (a) for presentation

of point prediction.

However, multiple limitations exist for point prediction, which limit its use with real-world problems. As computer powers are limited, weather grids are limited too, and hence important local features between two points are overlooked, which can lead to over or under estimation of wind speed. But this can be resolved with expensive and highly computational equipment and models if necessary [37].

When focusing on a single location point, i.e a wind farm, point forecasts can be issued in an inexpensive way, depending on a linear time-series and only the local measurements. This was first introduced in 1984, using Auto-regressive Moving Average (ARMA), where the lead time was between hours and a few days. Focusing solely on prediction of wind power at short forecasting horizons, i.e 2 hours, simple Auto-regressive (AR) models perform very well. Contrarily, auto-regressive models require a fairly large data set to perform adequate estimations, which is not usually easily accessible to a new location to construct a wind farm, but is becoming more and more available these days [38]. In addition, as more approaches are being developed to predict wind power based on both measurements and meteorological forecasts, such as neural networks, multi-fidelity gaussian processes and machine learning models, the wealth of data is not a huge problem. The aforementioned models use Least Squared (LS) and Maximum Likelihood (ML) approaches, which allow smooth changes in the parameters of the employed model, also while lowering the computational expense. This was implemented by [39], where point forecasts for 5 zones of western Denmark were carried. The models had 43 hours ahead hourly resolution, and resulted in a well-captured pattern for the early lead times generated from the meteorological forecasts and AR based component based on local observational data.

As point prediction only focuses on the estimation of a single point at specified time or space horizon, the output is in the form of a single value that provides no information on the uncertainty around that estimated point value or confidence bounds. This is a key limitation of point prediction and is solved by introducing interval prediction.

2.2.3 Interval Prediction

Interval predictions can manage the key limitation of point prediction by providing an estimate of the likelihood over a range of outcomes, hence producing an interval supported by upper and lower limits, and a probability. There comes a variety of forms for this type of prediction: quantile prediction, can estimate that the wind speed will be exceeding a certain value for a given horizon with a confidence bound, i.e the wind speed for the next hour will be between 5-10 meters per second with an 80% confidence in that information. Similarly, an interval prediction will forecast the probability of an observation occurring within a specified interval. Information regarding the full range of outcome is embodied in a predictive distribution, which can take either parametric or non-parametric forms, a more detailed review can be found in Zhang's technique [37]. But in short, the aim is to obtain a prediction at time t with a certain confidence of an interval, given the expected time of the future horizon $t + h$, lies within a certain probability.

There is far less scientific research and contributions with respect to interval prediction when compared to point prediction, mainly due to how economically viable point predictions provide information with ease and specific targets. In the field of wind power manufacturing, a point predictor can make predictions of the future consumption, contrarily, an interval

prediction will yield the estimated future consumption and prove lower and upper confidence limits with probability of data being valid. In addition, other applications include predictions of connected forecasts, i.e relation between generated power at different wind farms in the same region. Here, the inner dependencies between observations are very important and interval prediction will provide the probability and confidence of each link, where strongly related dependencies can be used to establish the relations between wind farms in both spatial and temporal structures [40].

In wind predictions, sharp fluctuations and changes in the wind speed can occur at any given time, this is usually called ramp events. Ramp events heavily affect the accuracy of the prediction, and are responsible for a large proportion of errors in the prediction process [41]. It is possible to forecast the magnitude of a major change, however, the more important factor is time and when they occur, which is very challenging. The preliminary results show an increase in the magnitude of ramp events by at least 5 folds by the year 2025.

In addition to sudden changes in wind speed, which are not a rare occurrence specifically at offshore sites, other properties of wind vary drastically in the offshore environment compared with those onshore. Due to diurnal heating of the surface and topography effects such as roughness of the surrounding area, the wind exhibits different atmospheric boundary layer properties, thus some techniques have been assigned specifically for offshore predictions. A comparison between onshore and offshore prediction accuracy was conducted in [42], which concluded that the complexity of the terrain of the onshore area, directly correlates to the performance of the offshore prediction. In which simple smooth terrains yield more accurate predictions than rough complex onshore sites.

2.2.4 2D Interval Prediction

Despite point prediction and interval prediction covering an extensive area in the forecasting of a time-series, it is not sufficient for multiple domains. In the energy supply sector, it is not only the wind speed at a specific point that is required, but an estimation of the wind for a time frame supported by a confidence interval. In other words, it is important to have a clear estimation of the wind speed for time intervals to estimate the minimum power output for a time period, this way, wind operators can agree on electricity and power deals, knowing they can meet their agreements without unforeseen fluctuations in the output power and that supply and demand is met. Consequently, a time interval is when predictions occurs between $t + m$ or $t + h + k$, where h is the forecasting horizon and k is a measure of the size of the time interval, together they form a 2D array m, Check Figure 2.4 panel (c) for a presentation of 2D interval prediction.[37].

2.3 Classification of wind speed forecasting

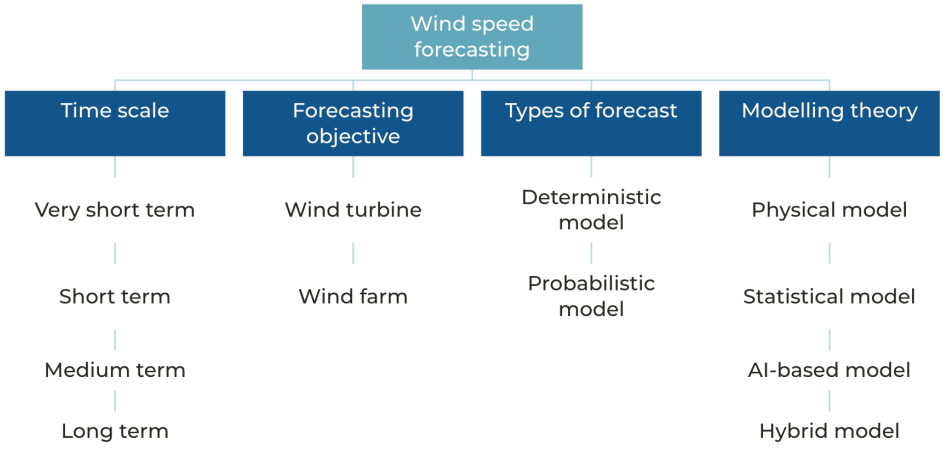

Figure 2.5: Classification of wind speed forecasting.

In the quest for more precise and accurate wind speed and wind power pre-

dictions, several approaches were proposed and implemented. The different approaches can be classified into sub-classes according to the standards and parameters of the forecasting task as per the presented in Figure 2.5.

Relative to the forecasting objectives, forecasting techniques can be further categorized into two classes: wind turbine and wind farm forecasting. The wind turbine forecasting techniques aid forecasts for a single wind turbine [43, 44, 45]. Contrarily, wind farm techniques investigate the wind information from several wind turbines and targets wind speed and wind power forecasts of entire wind farms, hence is more difficult [46-49].

The type of forecast provided also contributes to classifying the wind speed and wind power models into deterministic and probabilistic models [50]. The former can only provide point predictions, ignoring complexities in the environment, therefore are often inaccurate with unsatisfactory results [51-54]. The latter provides more information than point predictions and can represent uncertainty in terms of prediction intervals [53-55]. Prediction of uncertainty could be either parametric, which assumes the uncertainty can be described by a probability distribution; or non-parametric, which generate prediction intervals to characterize uncertainty without a distribution [56]. The most common parametric models are Mean Variance Estimations (MVE), the delta method, and Gaussian Process (GP) [57]. In the MVE forecasting models, errors obey additive Gaussian distribution with a varying variance [59]. Alternately, the error in the delta method is assumed to follow a Gaussian distribution [58]. For non-parametric models, Lower Upper Bound Estimation (LUBE) [60], which measures the quality of prediction intervals, and Quantile Regression (QR) [59], which follows an asymmetric Laplace loss function are amongst the most popular approaches.

2.4 Wind data pre-processing

Uncertainty in wind speed and wind power time-series had been dealt with using several proposed algorithms. Pre-processing of the wind data can be either using signal decomposing methods or outlier detection methods [61,62]. In the first approach, signal decomposing methods consider the series of the wind data to be a signal that can be decomposed to multiple sub-series. Then, predictions are generated for each sub-series, and results are aggregated to produce a final forecast for the original wind data series [63]. Alternately, in some literature [64, 65], signal decomposing is used as an algorithm to remove noise in the signal, the denoised signal is then fed to the forecasting models for predictions. The signal decomposing algorithm is divided into three categories: wavelet-based, mode decomposition-based, and singular spectrum analysis-based approaches.

The wavelet-based approaches include Wavelet Decomposition (WD) [66], Wavelet Packet Decomposition (WPD) [67–69], and Empirical Wavelet Transform (EWT) [70,71]. The mode decomposition-based methods include Variational Mode Decomposition (VMD) [72,73], Adaptive Variational Mode Decomposition (AVMD) [74], Optimal Variational Mode Decomposition (OVMD) [75], Empirical Mode Decomposition (EMD) [76], Ensemble Empirical Mode Decomposition (EEMD) [77], Fast Ensemble Empirical Mode Decomposition (FEEMD) [67], Complete Empirical Mode Decomposition (CEEMD) [68], Complete Empirical Mode Decomposition With Adaptive Noise (CEEMDAN) [79], and Improved Complete Empirical Mode Decomposition With Adaptive Noise (ICEEMDAN) [80]. Also, Singular Spectrum Analysis (SSA) [81-83] and Improved Singular Spectrum Analysis (ISSA) [84] have been employed to process original wind speed data before forecasting. For wavelet-based approaches, the algorithm de-

composes the original signal to generate different levels of the wind speed series, while the other two approaches can only produce sub-series at a single level. Moreover, Liu et al. [85] proposed a combined signal processing method, which combined FEEMD and WPD to process wind speed data. The latter was employed to decompose the original signal into different sub-series at different frequencies, while FEEMD was employed to further decompose the high frequency sub-series. Results from combined signal processing methods have shown higher efficiency in wind speed forecasting [85, 86].

The pre-processing phase in the model focuses either on decomposing the signal or data denoising, but the algorithm processes all the data in the signal indiscriminately [61]. The decomposed sub-series usually have a stronger predictability than that of the original signal. However, removing the noise from a signal can still generate a denoised signal with outliers, for this reason, outlier detection-based processing methods exist, but they do not completely remove outliers [62]. In few studies, researchers combined both pre-processing methods to reduce the uncertainty of the signal.

2.5 Wind Forecasting Methods

Wind speed forecasting methods involve using various techniques to predict the future wind speed at a particular location. Some common methods include persistence method, numerical weather prediction or physical models, statistical models, machine learning algorithms, and hybrid models that combine multiple techniques, illustrated in Figure 2.6. These methods use historical wind data, atmospheric conditions, and other relevant data to make predictions. The accuracy of the forecasts can be affected by factors

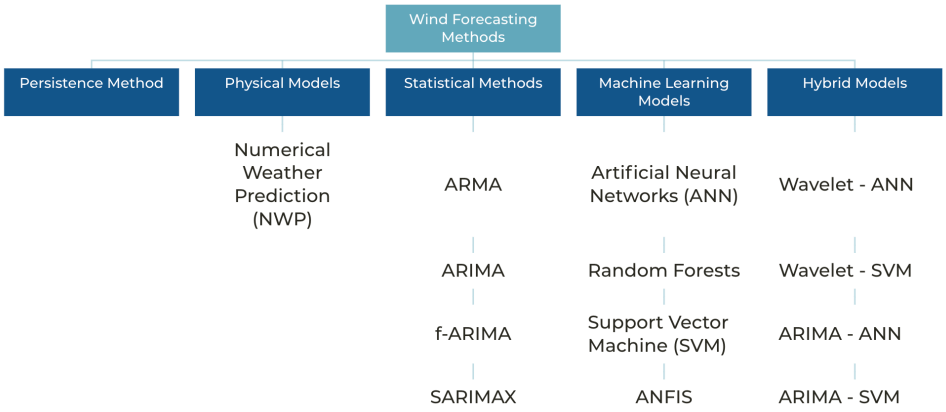

Figure 2.6: Categorization of the different methods for wind forecasting techniques.

such as the complexity of the terrain, the season, and the time of day.

2.5.1 Persistence method

In meteorology, stating that the weather condition in the near future is unchanged to the current condition, is called persistence forecasting. The assumption for this method is that conditions around the wind speed at the time of forecasting will be constant and unchanged to present ones when the forecasts are made. The method disregards any changes in the boundary conditions, and hence spontaneous turbulence effects, thus it is often used as a standard model for comparisons and bench-marking. Despite that, persistence models can yield very accurate results, more accurate than other complex wind forecasting models, when specific parameters are met, such as a balanced state in the boundary layer with no sudden or abrupt changes, and when the forecasts are made within the ultra to very short term forecasts. However, degree of accuracy will deteriorate rapidly as the time-scale is increased. Hence, the persistence method is mostly used in electricity utility for ultra-short term time horizons, or as test result benchmark for newly developing models [87].

The persistence method is considered the simplest and one of the most economical wind speed forecasting methods, as it is based on performing simple linear mathematical equations. If the measured wind speed is v_t at time t, then the change in speed until $t+1$ is equivalent to the change from $t-1$ to t [87].

2.5.2 Physical models (NWP)

Development of numerical models

Numerical Weather Prediction (NWP) are computer simulations that consider meteorological factors (air pressure, humidity, temperature) and have been in place since the 1950s. Charney, Fjortoft, and Von Neuman [88], used a low resolution, highly filtered version of the dynamic equations. The early models of NWP established consistent low-resolution models, by relying on the theories of quasi-geostrophic. At that time, it was nearly impossible to include relevant physical processes, due to lack of knowledge or computer resources. However, quasi-geostrophic models were very popular and were mainly used for process studies and hemispheric forecasts (mostly short range of up to 3 days ahead). The models focused on the development of mid-latitude weather systems, which were characterized by large horizontal scales on the order of hundreds of kilometers spatially, days temporally, and thorough tropospheric structures [88]. As these models overlooked some external forcing processes, such as the important atmospheric processes (radiation and phase transition) and the evolution of the temperature on surface, their predictions had very limited capabilities. On the other hand, modern NWP models form the foundation of most of the current meteorological forecasts.

Current models simulate processes occurring in the atmosphere on computers, to estimate the future state from available current states. Assessment of the current state of the atmosphere is called data assimilation, a paramount process for accurate predictions, which estimates the state from measurements collected from stations and satellites all over the globe. This process also validates the measurements and fills in the gaps between the stations in locations such as oceans, deserts, and big forests. A set of linearised equations are initialised, which describe the physics of the atmosphere, namely, the Navier-Stokes equation and the ideal gas law (solved on 3-Dimensional grids). Both the initialization and linearisation steps of the atmospheric parameters and governing equations, respectively, contribute to a meaningful data assimilation of the atmosphere and accuracy of predictions [89].

The systems in old NWP models contained relatively simple algorithms. The model interpolates observations to the grid after a preprocessing step, where filters and constraints are imposed on the balance between different fields, which reduces the noise in the initial state. Recently, the initial state is optimized by variational approaches to assimilate observations along a time frame in modern NWP models [90]. This is typically appropriate for remote sensing observations that are becoming very popular. These observations are updated frequently and are global datasets.

Available data for boundary conditions

Furthermore, direct 3D atmospheric monitoring of observations of the atmosphere are only available from radiosonde networks. With addition to dense surface station networks, the synoptic network is constituted, which provides worldwide synchronous observations at appointed times. Contrarily, these networks are considered heterogeneous in space, as the major

setback of such models include a wide unavailability of observations at large areas, such as oceans. This brought major attention to non-synoptic data, where satellite data became the main source. These are considered to be large-scale data products, however, other sources of smaller-scale models are available, such as lightning retrievals and radar images, which have great potential to be incorporated in NWP [91].

Despite the advanced technologies available and powerful supercomputers being used for NWP simulations, spatial and temporal resolutions are still limited. Several NWPs are run in multiple places around the globe, covering different regions using measurements generated by weather satellites and radiosondes. The forecast horizons of the weather forecasts are typically between 7 and 10 days, where the spatial resolution ranges from 3 to 30 kilometers, and the temporal resolution is either 1 or 3 hours. It is possible to increase the duration of the forecast, however, this results in a lower resolution result. Moreover, forecasts are usually issued every 6 or 12 hours due to the high computational cost of running NWP [92]. In Europe, multiple countries have national weather service providers that run NWPs, for example the UK has the Met office, France uses the Meteo, and Deutscher Wetterdienst, Germany. In addition, the European Centre For Medium-Range Weather Forecasts (ECMWF), runs many NWPs, which ranges over days, months, and seasons ahead [93]. Due to the increased demand for high-resolution forecasts, several other organisations were formed, producing high-resolution data covering different regions (ALADIN, COSMO, HIRLAM) [94]. The final step in NWP is completed by multivariate statistics. In [95] the 3 dimensional data assimilation system that takes place at HIRLAM is described. The equations used to simulate the atmosphere take into account the effect of turbulence and its variation during the day. A thorough description is given in [96], where they

described how the boundary conditions are applied and integration of the model by numerical methods.

Extracting wind power from the models

To derive wind power forecasts from NWP models, significant post-processing is required. The vast research carried out on post-processing of NWP lead to developments of calibration models, such as Model Output Statistics (MOS), which calibrates the forecasts at specific locations and variables when possible [97]. The modelling of wind power conversion process is essential, but can be a significant source of uncertainty. However, NWPs used for short term predictions provide accurate power estimates [97]. Physical and statistical modelling approaches should be used simultaneously in wind power forecasting systems, which is the case for most operating commercial models [98]. Prediktor is a wind power forecasting tool developed by Landberg and Troen [99], the tool takes NWP estimates of the wind speeds and directions to adapt them to the local site prior to applying a power curve model. Throughout the process, statistical improvements could be achieved through MOS. Another tool was developed at the Technical University Of Denmark (DTU), the Wind Power Prediction Tool (WPPT), which is managed and operated by an external industrial firm. To discover the best relation between the NWP predicted wind speeds at the site and the measured power for each forecast horizon, the WPPT employs adaptive neural recursive least squares estimation of the parameters of conditional parametric models [100]. Another model, the Wind Power Management System (WPMS), employs neural networks for the same task, this tool was developed in Germany and is currently being used by E.On, RWE and National Grid in the United Kingdom. DNV offers a forecasting model called GH forecaster, which is based on UK met office NWP projections. The tool converts NWP to local wind speeds using multiple input linear regression

algorithms [101].

An essential NWP technique is ensemble forecasting, which runs the NWP simulations multiple times with the estimates of the initial atmosphere conditions modified [102]. The ensembles should ideally be viewed as samples from a probability distribution function that reflects the uncertainty of the unperturbed forecast.

Physical models perform superior to other models in the medium and long time horizons, which made it possible to obtain wind power forecasts using power curve for wind turbine at given wind speeds. Physical models can be superior in wind speed forecasting because they are based on the fundamental physical principles that govern the atmosphere and the behavior of wind. These models use equations that describe the dynamics and thermodynamics of the atmosphere, and they take into account various factors that can affect wind speed, such as topography, land use, and atmospheric conditions. Additionally, physical models can also incorporate real-time data from weather observations and remote sensing technologies, which can improve the accuracy of the forecasts. Additionally, physical models can be used to simulate different scenarios and analyze the impact of various factors on wind speed, which can be useful for planning and decision-making. However, physical models can be computationally expensive and require a significant amount of data to be accurate. They may also require specialized expertise to develop and operate. As a result, different forecasting methods may be more appropriate depending on the specific application and available resources.

2.5.3 Statistical methods

The term statistics when used in forecasting, implies the usage of historical data to ascertain the probability of what might happen in the future. These are time-series models employed to characterize the linear fluctuation of wind speed at different locations. The process consists of a single step, where input variables are directly transformed into wind generation, in a statistical block. For this method, short-term wind forecasting is the most common category for time horizons [103]. Statistical methods have three main advantages when used for short-term predictions, which make it superior to physical methods. Firstly, NWPs require several hours to produce, additionally, they are issued in 6 and 12 hours; hence, when forecasts are issued, recent measurements that should serve as input will be relatively old. Secondly, the output from NWPs is distributed on spatial grids with varying resolutions, which might not provide data at points of interests such as a wind farm. This requires additional spatial interpolation to allow for forecasts at points of interest, which adds an extra layer of complexity, due to complex and different terrains. Thirdly, for very short-term forecasts where high temporal resolution of finer than 1 hour are required, operators will face the same interpolation problem as in the spatial domain. In conclusion, to have a strong performing wind forecasting method, a combination of statistical and physical methods will combine the merits of both techniques. The statistical method will power the predictions for the first few hours, then a smooth transition to physical models for the long time horizons. Wind power forecasting can be approached in two ways, either forecasting the wind power directly, or to forecast the wind speed with direction and combining it with a turbine power curve model [103].

The statistical block computes statistical linear and nonlinear models that

can be of different types. Among these are combinations of machine learning methods, such as Neural Networks (NN), Support Vector Machine (SVM) and kernel regression, and time-series models such as: AR, MA, ARMA, ARIMA, ARMAX and Fractionally Integrated Auto-regressive Moving Average (FARIMA) as described in the following. The Auto-regressive (AR) time-series models use previous time-step observations as input to a regression equation, which in return outputs a forecast for the next time-step. Contrarily, Moving Average (MA) depends on past errors for forecasting. Furthermore, Auto-regressive Exogenous Moving Average (ARMAX), is a non linear model, which captures uncertainties in the time-series associated with the stochastic nature of wind. The time-series models only take past values from the foretasted variable as inputs, simultaneously, they can also use past values of other related variables to improve the forecast accuracy. These models can outperform the persistence method in the very short and short time horizon between 3-6 hours by 15-20% [104]. However, greater time horizons would require NWPs as input data. The problem can be looked at from two approaches, a univariate approach and a multivariate one:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_0 Y_0 + \epsilon_t \tag{2.3}$$

this model only uses past values of the predicted variable, where ϵ_t is white noise.

For the multivariate model, past and present values of other variables can be used:

$$Y_1(t) = \beta_1(11) Y_{t-1} + \beta_1(12) Y_{t-2} + .. + \epsilon_t \tag{2.4}$$

$$Y_2(t) = \beta_2(21)Y_{t-1} + \beta_2(22)Y_{t-2} + .. + \epsilon_t \tag{2.5}$$

where, $\beta_1(11)$ and $\beta_2(21)$ are coefficients for both variables and $\epsilon_t(1)$ and $\epsilon_t(2)$ are white noise (error terms). The function is expressed in past values of the forecast variable, and past and present explanatory variables t.

The wind is commonly represented as an AR process, with the resulting forecast being a linear combination of previous data. However, a traditional AR process has zero mean and is homoscedastic, which opposes the nature of the wind speed. As a result, the wind speed time series must be completely altered or the AR model must be adjusted to match these needs. Wind time series are also non-stationary, which means that their statistical features fluctuate with time.

An improved statistical model is ARMA [105], which consists of combining the merits of autoregression with parameter p and moving average with parameter q, hence modelling the trends in the time-series. The combined models form the the final approach ARMA with combined parameters (p, q), here the parameters represent the lag between present and past values of the predicted variable. ARMA can be mathematically expressed as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \epsilon_t \tag{2.6}$$

$$y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + ... + \theta_q \epsilon_{t-q} \tag{2.7}$$

$$y_t = c + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \theta_q\epsilon_{t-q} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_p y_{t-p} \quad (2.8)$$

where p is the order, c is a constant and, ϵ is noise, and top line is the autoregressive model of order p and the second line is for the moving average model of order q. The combined parameters of the two preliminary models form the ARMA full equation seen in the third line. The model is usually assessed by calculating the error matrices such as, Root Mean Square Error (RMSE), Residual Squared (R^2), Mean Absolute Error (MAE), or the Mean Absolute Percentage Error (MAPE).

However, a great limitation exists, as a constant variance is still assumed in the ARMA model, which means they are only applied to stationary data with single mean value [105]. Thus, another model Auto-regressive Integrated Moving Average (ARIMA) [106], aims to eliminate the non-stationary part of the time-series (diurnal, meteorological, and seasonal variations) by using an initial differencing step, which introduces parameter d, a difference operator. The mathematical expression for ARIMA (p, d, q) is:

$$y_t = c + \phi_1 y_{t-1} + .. + \phi_p y_{t-p} + \theta_1\epsilon_{t-1} + .. + \theta_q\epsilon_{t-q} + \epsilon_t \quad (2.9)$$

where ϕ and θ are the coefficients of AR and MA. ARIMA is successful at outperforming ARMA for non-stationary data, the literature supports that ARIMA is often used as the benchmark for newly developed models, rather than ARMA. In [107], a fractional variant of ARIMA is introduced, f-ARIMA, where the fractional property is a differencing parameter. Results show that it outperforms the persistence method by at least 42%.

2.5.4 Machine learning methods

As wind flow has a complex non-linear nature, which depends on multiple atmospheric parameters such as pressure and humidity, several applications of machine learning algorithms for time-series forecasting were conducted to further improve the precision and accuracy of wind speed predictions. Machine Learning is a major section of a much larger subject called Artificial Intelligence. The intuition behind machine learning is to either classify, predict, cluster, or recognize a pattern. The learning occurs when the model is provided with labelled input and targeted output, once the model establishes a connection between the inputs and outputs, the training process is completed and testing may start by providing new inputs to estimate the desired outputs. This type of forecasting model is superior to traditional physical and statistical models when the time-series has complex nonlinear relations [108].

The previous example of machine learning is called supervised learning, where the inputs are labelled and the algorithm tries to converge for the best fit classifier, so independent features set X is linked to dependent distinctive class Y, when an external input is introduced the model can then identify and classify the data and generate an estimation of which class the input data belongs to. The second type of machine learning is unsupervised learning, in this type, input data is unlabelled and the algorithm relies on itself to learn the data and build the classifier, which occurs based on catching similarities between the input sets .

To have a robust and accurate machine learning algorithm, large historical data of the wind speed must be highly available to train supervised machine learning regression models. Regression models to work with numbers, where the input is a time-series of continuous measurements of variables.

The desired output from regression models is an estimation of the value at unobserved points in the times-series or future forecasts [108]. Common regression supervised models include K-nearest neighbors, decision trees, random forests, Support Vector Regression (SVR), extreme learning machine, and Neural Networks (NN).

However, the most common machine learning approach for wind speed predictions is neural networks, also known as deep learning, where the intuition behind the algorithms is to mimic the behaviour of human brain neurons by having multiple layers consisting of neurons all connected to each other and a huge amount of possible combinations is tested to generate one with optimum learning. Deep learning algorithms are extensively used in time-series forecasting including wind and wind power forecasting applications [108]. The duration of estimations can range between very-short terms such as 1 minute and short-terms such as 6 hours. Aghajani and Afshin in [109] used Multilayered Perceptron (MLP), which is a neural network that has multiple hidden layers (where the neurons connect to each other and do the learning), were used along with wavelet transform to predict wind power for wind farms. The time-series of the wind is pre-processed and broken down into sub-series using wavelet transform, which is later fed as inputs to train the neural network. After the training stage, an optimization stage follows, where learning of the network improves by minimizing the error between the training input and predicted values. A very common learning approach is back-propagation, in this approach once the desired value is estimated and the error is calculated, information flows back from the output layer to the hidden layers to re-calculate the parameters [110]. An alternative approach that replaces back-propagation is Levenberg-Marquardt, which is faster than back-propagation [111]. Results of neural network training with time-series pre-process by wavelet transform are superior to other ap-

proaches such as Linear method ARIMA. As powerful as neural networks can become, they are still considered a 'black box', as they offer almost no insights to the learning phase, it is hard to tell why a neural network performs better than another variant.

Table 2.2: AI-based models

Acronyms	**Abbreviation**	**Reference**
SVM	Support Vector Machine	[112]
LSSVM	Least Squared Support Vector Machine	[113]
ELM	Extreme Learning Machine,	[114]
BPNN	Back Propagation Neural Network	[115]
MLP	Multilayered Perceptron	[116]
WNN	Wavelet Neural Network,	[115]
LSTM	Long Short Term Memory	[114]
BiLSTM	Bidirectional Long Short Term Memory	[117]
CNN	Convolutional Neural Network,	[118]
GRU	Gated Recurrent Neural Network	[119]
BiGRU	Bidirectional Gated Recurrent Neural Network,	[120]
DBN	Deep Belief Network	[121]
AE	Auto Encoder	[122]

Deep neural networks have attracted great attention in wind speed forecasting due to their superiority in dealing with nonlinear complexes. In many previous studies, AI-based models outperformed traditional statistical models in wind speed forecasting. A few machine learning forecasting models are listed in Table 2.2.

2.5.5 Hybrid models

To improve the performance of algorithms and prediction, multiple approaches combined the merits of different techniques. Hybrid models are divided into two types; stacking based, which consider the forecasts of one or more base models as features and later combined with a high model, and weight based hybrid strategies, which follow a diversity in choosing

the forecasting models. An example of stacking-based hybrid models is taking the output features from a CNN and feeding them into a LSTM to extract spatio-temporal features of wind data [123,124]. In [125] and [126], multiple LSTM were combined to form a stack of neural networks that makes complete use of spatio-temporal data and multiple meteorological information, respectively. In [127], forecasts from different LSTMs were combined and fed into a SVM model to make wind speed predictions. Alternately, for weight-based hybrid models such as in [116], a combination of three different base forecasters, namely ARIMA, MLP, and LSTM, were investigated to forecast wind speed, while an intelligent optimization algorithm optimized their corresponding combination weights. In [115], an ensemble of four different NNs (BPNN, Elman, WNN and GRNN), were combined with certain combination weights, to make wind speed forecasts. Additionally, two hybrid models (ARIMA-SVM and ARIMA-ANN) were simultaneously used to extract linear and nonlinear fluctuations in a wind speed dataset, and outperformed single forecasting models [128]. Another construction of weight-based models is to use parameters from diverse homogeneous forecasters. In [127], six LSTM neural networks were used as base models to forecast wind speed.

A hybrid approach by Liu et al. [129] combined the merits of wavelet transform and support vector machine to predict short-term wind speed, which used the genetic algorithm for optimization. Another approach in [130] demonstrated that using ANN with statistical weight preprocessing techniques optimized by Bayesian optimization can predict the coefficient of performance of a Ground Source Heat Pump (GSHP) system. In addition, other studies concluded that applying optimization methods can yield better results due to improved selection of model parameters. Among the commonly used optimization algorithms is the genetic approach, which

tends to converge faster than other approaches. In [131], Su et al. employed particle swarm optimization to optimize the parameters of ARIMA along with Kalman filter for the forecasting process of the wind speed. As aforementioned, hybrid techniques are continuously used to study short-term wind speed and wind power time-series. The pre-processing techniques, wavelet transform, empirical wavelet decomposition, and empirical wavelet transform are good examples of signal processing methods that are used for studying the stochasticity of the wind time-series. Du et al. [132] used wavelet decomposition along with ANN and optimized the prediction algorithm using a multi-objective optimization of hyperparameters, results demonstrated an improved accuracy in predictions and higher reliability for short-term wind forecasting compared to other benchmarking models such as hybrid models that disregarded the optimization step, persistence method, ARIMA and a neural network.

The setting of the model configuration for a hybrid model is a great factor to enhance the performance, therefore, several intelligent optimization algorithms have been developed to configure the optimal model configuration. For example, in [133], three wind speed forecasting neural networks, BPNN, WNN, and GRNN, were tuned using Genetic Algorithm (GA), Cuckoo Search (CS), and Conjugate Gradient Bat Algorithm (CG-BA), respectively. In [134], Improved Atomic Search Algorithm (IASA) was used to select optimal SVM parameters, which outperformed SVM optimized using Particle Swarm Optimisation (PSO), GA and Atomic Search Algorithm (ASA). In [113], PSO is combined with gravitational search algorithm, resulting in hybrid optimizer PSOGSA, which tuned wind speed forecasting model LSSVM faster than PSO. In [135], a hybrid LSTM wind speed forecasting system is optimized using GA, while in [136], the Differential Evolution (DE) algorithm was used to balance the complexity of

LSTM models. The optimal configuration of a BiLSTM neural network was selected by using Generalized Normal Distribution (GND) for wind speed forecasting [117]. The mentioned optimization algorithms are considered single objective algorithms, used to enhance the performance of a forecasting deep neural network, despite observed improvements, they are unable to perform at high stability. Hence, multi-objective optimization methods were developed. For instance, the Multi Objective Sine Cosine Algorithm (MOSCA) optimized a WNN, where results reported stable and accurate wind speed predictions [137]. In [138], Elman and Volterra filters were optimized using the Multi Objective Cuckoo Search (MOCS) algorithm, while Multi Objective Grey Wolf Optimization (MOGWO) optimized ELM in [139].

2.6 Optimization of model configuration

A very important stage of building a wind speed or wind power forecasting model is to configure the model and choose an optimizing algorithm that appropriately optimizes aspects in the forecasting model. As performance of any prediction model is highly influenced by the configuration, it is critical to employ an intelligent optimizing algorithm that considers configurations such as the architecture and hyper-parameters of the model to achieve maximum performance. Over the past few decades many optimizing algorithms have been developed and adopted, which targeted various aspects of the model [118].

For DNN based forecasting models, model configuration parameters such as weights, number of hidden layer, batch size, learning rate, and optimizer type are important to determine for optimal performance [117, 140].

In [141], ELM was employed to estimate the wind power output of wind turbines, the weights and biases of the model were determined using Ant Colony Optimization (ACO). In [142], the crisscross optimization algorithm was adopted to determine the input weights and biases in the hidden layers of a variant ELM forecasting model. Another approach optimized the output weights of ELM using a Self Adaptive Differential Evolution (SADE) algorithm for wind power forecasts [143]. Additionally, Improved Hybrid Grey Wolf Optimizer And Sine Cosine Algorithm (IHGWOSCA) was also used to optimize ELM in [144]. In [145], short-term forecasts of wind speed using BPNN, WNN, GRNN, were optimized using GA, CS, and CG-BA, respectively. Moreover, in [135], GA was employed to optimize a LSTM neural network, while in [136], to balance the complexity of a LSTM, DE was used. And in [137], GND was adopted to tune the hyper-parameters of a BiLSTM. Subsequently, multi-objective optimization algorithms were adopted in several studies to achieve both strong stability and high accuracy [138]. Algorithms such as MOSCA, MOCS, and MOGWO, were employed to optimize WNN, Elman, and ELM models, respectively [138-140].

Optimization models are adopted significantly with weight-based models to tune the weights of each base model, which has a critical role in the final accuracy. In [146], GWO, an optimizing algorithm, was employed to tune the weights of several LSTMs, forecasting different intrinsic modes, generated by the signal decomposing algorithm ICEEMDAN. Similarly, GWO was adopted in [77] to determine the weights of a hybrid forecasting model, which consisted of BPNN, Elman, WNN, and GRNN forecasting models, for wind speed forecasting. In [147], four ANNs were optimized by multi-objective bat algorithm to tune the weights of the hybrid forecasting model. A Multi Objective Multi Universe Optimization (MOMUO) was

employed in [148] to determine the optimal weights for two base BiLSTM networks.

Past studies have shown that optimization models are often employed in wind speed and wind power forecasting models. However, more research is required for tuning DNN models with intelligent optimization algorithms. A possible reason for the lack of DNN models with intelligent optimization algorithms is that the model would require long time to try several different configurations as DNNs are complex models with numerous parameters.

2.7 Conclusion

Wind speed forecasting is a vital aspect of weather forecasting, especially in the field of renewable energy. As wind energy becomes a more prominent source of energy around the world, the accuracy of wind speed forecasting becomes increasingly important for efficient and effective operation of wind farms. There are various methods for predicting wind speed, and the choice of method depends on the specific application and available resources. Numerical Weather Prediction (NWP) models are based on mathematical models of the atmosphere and use real-time weather observations and other data to simulate atmospheric conditions and to make predictions about future weather patterns. These models are highly complex and require significant computational power, but they can be very accurate with detailed information about weather patterns. Physical models, on the other hand, are based on fundamental physical principles that govern the behaviour of wind. These models use equations that describe the dynamics and thermodynamics of the atmosphere, and take into account various factors that can affect wind speed, such as topography, land

Table 2.3: Pros and Cons of Wind Speed Forecasting Methods

Forecasting Method	**Pros**	**Cons**
Persistence model	Simple and easy to implement	Limited accuracy and no adaptation to changing conditions
Statistical methods	Can capture patterns and trends	Relies on historical data and may not account for dynamic changes
	Flexible and adaptable to different datasets and variables	May struggle with non-linear relationships and extreme events
Physical models	Based on fundamental physics principles	Complex and computationally intensive
	Can capture physical processes and interactions	Requires accurate input data and precise knowledge of model parameters
	Can be tailored to specific geographical locations	High computational requirements and longer processing times
Machine learning models	Ability to learn from data patterns	Requires large amounts of quality training data and model complexity
	Adaptability to changing conditions	Prone to overfitting if not properly regularized and validated
	Can handle high-dimensional data	Sensitivity to outliers and noise in the training data
Hybrid models	Combination of strengths from different methods	Increased complexity and potential integration challenges
	Improved accuracy and robustness	May require additional computational resources and expertise

use, and atmospheric conditions. Physical models can be computationally expensive and require a significant amount of data to be accurate, but they can be highly accurate and provide detailed information about the underlying physical processes. Statistical models are based on historical data and use statistical analysis techniques to make predictions about future wind speed. They are often simple and easy to implement, but their accuracy can be limited by changes in the environment or other external factors. Machine learning algorithms are another popular method for wind speed forecasting. These models use data-driven techniques to identify patterns in the data and make predictions about future wind speed. Machine learning models are becoming increasingly sophisticated, and they can be trained on large amounts of data to improve their accuracy. Overall, wind speed forecasting is an important tool for a variety of industries that rely on accurate weather information. The accuracy of wind speed forecasts can be affected by many factors, including the complexity of the terrain, the season, and the time of day. Therefore, it is important to choose the appropriate forecasting method for the specific application and available resources. Despite the challenges, continued research and development in wind speed forecasting techniques will undoubtedly help to improve the efficiency and effectiveness of wind energy systems and that of wind resources assessment prior to wind farm installations. In the recent years, hybrid models had shown superior performance in comparison with traditional models, specifically at describing complex relations. Hence, more hybrid models of different DNNs continue to be widely investigated, where characteristics of advanced models are combined to accurately forecast the wind speed and power. However, there is significantly low research on hybrid models that combine DNNs with physical approaches, this hybrid combines investigating the air parameters, such as air pressure, temperature, and humidity using physical models, and characterises the non-linear

complex behaviours using DNNs to generate powerful forecasts.

Hence, in the following chapters, investigation on combinations of WRF data and other forecasting models such as neural networks, statistical methods, and combining them with pre-processing techniques for learning patterns will conducted and multiple models following this scheme will be developed to test the concepts of propagating predictions both temporally and spatially for a wind farm location, and how nearby and further distant wind speeds each affect the accuracy of predictions and model performance. The experiments aim to test different combinations of data fusion and modelling of datasets from different sources such as LiDAR observations and WRF simulations. Pre-processing of the data will be carried out using both single and double signal decomposition techniques, comparing the benefits of a second decomposition method. Additionally, novel algorithms of neural network combinations and optimisation using different optimisation algorithms will be used to test the merit of multiple optimisation techniques. The algorithms aim to optimise different parameters in the model and the configuration of the neural networks and prediction algorithms. Moreover, the work will present a comparison test between the performance of machine learning models and hybrid models of both machine learning algorithms, statistical models, and physical models.

Chapter 3

Offshore Wind Resource Assessment from Limited Onshore Measurements

In wind resource assessments, which are critical to the pre-construction of wind farms, measurements by lidars or masts are a source of high-fidelity data, but are expensive and scarce in space and time, particularly for offshore sites. On the other hand, numerical simulations, using for example the Weather Research and Forecasting (WRF) model, generate temporally and spatially continuous data with relatively low-fidelity. A hybrid approach is proposed here to combine the merit of measurements and simulations for the assessment of offshore wind. Firstly a temporal data fusion using deep Multi Fidelity Gaussian Process Regression is performed to combine the intermittent measurement and the continuous simulation data at an onshore location. Then a spatial data fusion using a neural network with Non-linear Autoregression (NAR) and Non-linear Autoregression with external input (NARX) are conducted to project the wind from onshore to

offshore. The numerical and measured wind speeds along the west coast of Denmark are used to evaluate the method. The proposed data fusion technique using a gappy onshore measurement results in accurate offshore wind resource assessment within a 2% margin error. This chapter aims to investigate the first objective, which is to understand and employ the merits of data fusion on wind speed time-series predictions. Additionally, the experiment will consider the addition of side information, such as velocity vector components and derivatives of wind speed time series, and how this addition affects the performance of the forecasting model.

3.1 Introduction and Literature Review

In the past decades, there has been worldwide demand for renewable energy, leading to a dramatic expansion in all its sectors, with a significant fraction generated by wind. There are over 230 GW of installed wind capacity in Europe as of 2020, consisting of 190 GW onshore and 40 GW offshore. Additionally, Europe intends to further the rise in demand for wind energy and its capacity by 35% [149]. Before the construction of a wind farm, it is critical to evaluate the wind speed condition of the location. Since the power is the cube function of wind speed, minor speed changes can cause large deviations in the output power. Moreover, the wind varies both geographically and temporally over a wide range of scales. Therefore, an accurate assessment of the wind resource for a proposed site is highly essential and is considered of paramount significance for a wind project to be successful [150],[151]. The assessment also helps to the selection of wind turbines and their layouts [152].

Physical wind measuring devices include e.g. lidars, meteorological mast

towers, Satellite Synthetic Aperture Radars (SARs), and so forth. Those equipment yield accurate results but are expensive and the data is commonly sparse in space and time. For example, the SARs measured wind is at 10 m above the sea surface with low temporal resolution and only applies to offshore assessments; lidars measure the Line-Of-Sight (LOS) velocity by computing the Doppler shift of the signal of an infrared laser beam based on the movement of aerosols and the output is usually intermittent at a fixed location; buoy systems are expensive and require regular maintenance, redundant systems for power, measurements, and communication for the measurement at a given location. On the contrary, numerical simulations result in wind predictions that cover large geographical areas and long time horizons continuously, but with relatively low fidelity. The numerical models include e.g. Weather Research and Forecasting (WRF), Global Forecast System (GFS), and European Centre for Medium-Range Weather Forecasts (ECMWF) [153].

Such a clear complement of physical measurements and numerical information suggests data fusion or a hybrid technique to combine the merit of both. It would be very desirable to extend the information from coastal online vertical lidars for the reconstruction of offshore time series as they are easier to maintain [154]. This technique can be used to numerically extend the information from coastal measurements to the offshore time series with low cost and high accuracy. Such techniques have been widely used in the prediction of future developments based on various inputs [155]. For example, Hu and Wang [156] used Empirical Wavelet Transforms (EWT), Partial Auto-Correlation Function (PACF), and GPR for wind speed assessments. EWT was employed to extract the meaningful data from the wind speed series through a customized wavelet filter bank, and PACF provided the input parameters for the GPR to simulate dynamic features and internal

uncertainties. An alternative combination, i.e. Auto-Regressive (AR) and GPR, was followed by Zhang and Wei [154]. AR was employed to capture the structure of the wind speed series, and GPR to extract the local structures. As a supplement, Automatic Relevance Determination (ARD) considered the importance of using different inputs; thus, various types of covariance functions were combined to comprehend the characteristics of the data. This hybrid method outperformed others including Support Vector Machine (SVM), Artificial Neural Network (ANN), and the persistence approach. Meanwhile, an improved near-surface wind speed prediction experiment, which considered the atmospheric stability using GPR combined with Numerical Weather Prediction (NWP), for a time horizon of 72 hours, showed that the consideration of atmospheric stability was able to reduce the estimated errors, thus improving power predictions [150].

Most recently, the Multi-Fidelity Gaussian Process Regression (MF-GPR) has been demonstrated to significantly outperform the regular single fidelity model. The strategy of MF-GPR is to go beyond the regular AR kriging scheme and introduce more than one dataset at different fidelity levels. The first is a high-fidelity, scarce dataset which can be the physically measured one; the second is a low-fidelity, continuous dataset which can be generated from numerical simulations. The literature in this area further discussed various developments; for example, the Deep MF-GPR with additional datasets, e.g. first and second derivatives, phase-shifted oscillations, and different periodicity datasets leading to drastically improved approximations [159].

Further, wind resource assessments are commonly requested over a long-time-interval (e.g. a few months or years) and to cover large areas, therefore requiring spatial-temporal fusion of numerically and physically measured wind [160]. Apart from the temporal prediction reviewed above, ANNs

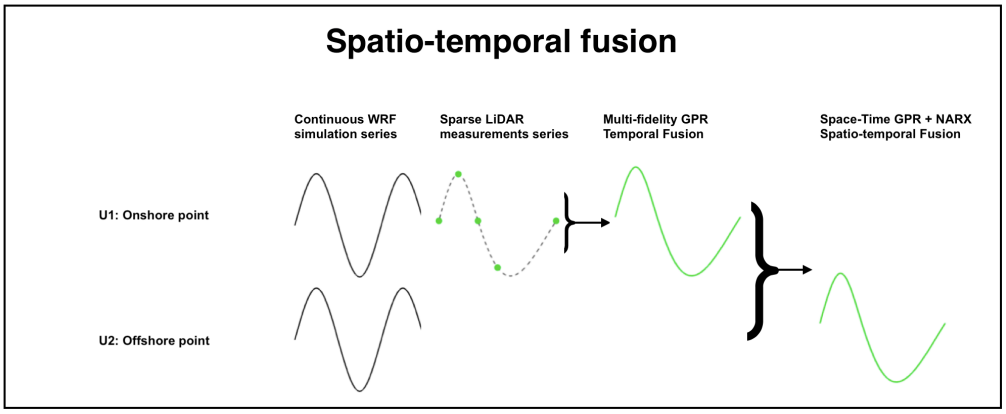

Figure 3.1: Flow chart for spatiotemporal fusion. U1 and U2 represent the wind speed at an onshore and offshore positions, respectively.

are trained and tested on datasets from two different locations. Cadenas and Rivera, considered the problem of non-linearity in the time-series using Nonlinear Auto-Regressive with Exogenous inputs (NARX) [158]. The method was compared with both the persistence approach and Nonlinear Auto-Regressive (NAR). The results demonstrated that the NARX model was the most precise of the three and justified the extra input, suggesting that it could be suitable for spatial data fusion.

Spatiotemporal models which combine the aforementioned temporal and spatial fusion, have been widely used in the geostatistics field, where temperature and wind speed were the main variables of concern. In these models, the temporal extrapolation is performed to predict the values out of the measured interval at a fixed spatial point [161], followed by spatial extrapolation to project the estimation to a different point [162]. This sequential extrapolation in time and space is developed in the present work for wind resource assessment. Temporal data fusion of low and high-fidelity data from simulations and measurements at a given location is performed using deep MF-GPR, and spatial data fusion using a customised nonlinear autoregressive ANN with exogenous inputs is conducted thereafter. As illustrated in Figure 3.1, the low fidelity results (e.g from numerical sim-

ulations) are assumed to be available across a continuous domain $X \times T$, where X and T are the spatial and temporal domains, respectively. On the other hand, high-fidelity results (e.g from the lidar measurements) are available in a reduced domain $X_{re} \times T_{re}$ where X_{re} is a subset of X and can be discontinuous and T_{re} is a subset of T and can be discontinuous. Thereafter, the objective is to combine the low and high-fidelity results to reach a data fusion on the full domain $X \times T$.

The novelty of this work lies on the development of a hybrid algorithm for the accurate assessment of offshore wind resources with reduced cost. It combines the generally continuous but low-fidelity numerical data and high-fidelity but limited physical measurements. Efforts are also devoted to pre-processing the time series and taking into account additional information not considered in existing methods to lift the accuracy of the fusion. This algorithm enables the projection of limited onshore measurements to offshore locations in light of numerical simulations with significantly higher accuracy than the industry standard approach.

This chapter is organized as follows. In Section 2, details for methods for the pre-processing technique used to smooth the given wind speed time-series, the temporal extrapolation using multi-fidelity GPR, and spatial extrapolation using NARX algorithms. Section 3 describes the case to be studied and the collection of the high and low fidelity data. Section 4 presents the main results and compare the performance of our methods against the industrial standard before drawing conclusions in Section 6.

3.2 Methods for Offshore Wind Resource Assessment from Limited Onshore Measurements

3.2.1 Data Preprocessing

EWT is a three-model decomposition algorithm used in forecasting. Several studies demonstrated that it could be used to achieve good forecasting results for non-stationary time series such as wind speed series [156]. EWT can extract meaningful information from the series by designing appropriate wavelet filter banks. In our work, pre-processing of the WRF time series generated by ERA-5 can adaptively represent the processed signal by generating the adaptive wavelet, and then decomposing the signal into a finite number of modes as per previous literature [156]. The algorithm is based on identifying and extracting the different intrinsic modes of a time-series by relying on robust pre-processing for peak detection, and then performing spectrum segmentation based on detecting maxima to construct a corresponding wavelet filter bank.

The WRF series that are pre-processed using EWT are 1) the WRF series in Hybrid (3) of the Multi-fidelity GPR for the most onshore point in Figure 5.4, panel (c) , 2) the WRF series for the 36 points generated by the dual-Doppler scans performed using the algorithm of Hybrid (3) in Figure 5.5 and 3) the WRF series of the second most onshore point in the temporal data fusion part of the spatiotemporal data fusion section.

The process can be divided into five steps [156]:

1. Extending the signal.

2. Fourier transforms.

3. Extracting boundaries.

4. Building a filter bank.

5. Extracting the sub band.

The original wind speed signal had considerable high-frequency fluctuations. The three-level decomposition attained by the EWT algorithm describes the wind speed series in a meaningful way. Three uncorrelated filter modes are extracted from the wind speed series and a residual is also obtained from the extraction. The reconstructed wind speed series shows a significant decrease in fluctuations and will be served as input to the third GPR model.

3.2.2 Temporal data fusion

This section introduces the algorithms for temporal data fusion by combining the low-fidelity continuous time series and the high-fidelity intermittent one. The prototype of the Gaussian process regression is briefly introduced and then multi-fidelity GPR is presented. The use of different co-variance functions such as constant, linear, squared exponential, Matern and rational quadratic, defines the method of prediction for the Gaussian process.

Gaussian Process Regression (GPR)

GPR is a non-parametric, stochastic process that follows the Bayesian approach for regression, working well on small data sets and having the ability to provide uncertainty measurements on predictions. Predictions are

derived using a probability distribution over all possible values of a time-series using prior functions w of training points f at observed points t, and targeted values f^* at unobserved points t^* are calculated from a predictive distribution, $p(f^*|t^*, f, t)$, by considering all possible predictions using their calculated posterior distribution [165]:

$$p\left(f^*|t^*, f, t\right) \;=\; \int p\left(f^*|t^*, w\right) p\left(w|f, t\right) dw. \tag{3.1}$$

To trace the integration process of equation (3.1), all terms of the equation are assumed Gaussian. The prior function defines the Gaussian distribution [165]:

$$f(t) \sim GP(m, k(t, t')), \tag{3.2}$$

where m is the mean function, which represents the trend of the function, and the covariance function (kernel), $k(t, t')$, represents the dependence of the structure, defined by the hyperparameters [163].

Multi-fidelity Gaussian Process Regression

This section discusses advanced temporal data fusion using data with multiple fidelities to enhance the accuracy of prediction. The data sets are obtained using different techniques mathematically, the multi-fidelity technique considers the high-fidelity model as a function of two variables (t, s) and then uses the low-fidelity data as the s variable [165]:

$$f_h(t) = g(t, f_1(t)), \tag{3.3}$$

where in the present work $f_h(t)$ and $f_l(t)$ are the high-fidelity lidar measurements and low fidelity WRF simulations, respectively. Such non-linear

auto-regressive Gaussian process (NARGP) has been observed to produce highly accurate prediction when $f_h(t)$ is non-linearly dependent on $f_l(t)$, and GPR is then performed in a two-dimensional space.

To implement this, the co-kriging model is adopted, which uses multivariate functions with respect to different levels of fidelities to reflect different accuracies. The additional data set is later introduced to the Gaussian distribution and the terms of the first data set (t, s) and the second data set (t', s') are added, while the mean function is zero, through [165]:

$$f(t) \sim GP(m, k((t, s), (t', s'))). \tag{3.4}$$

Merging of two or more sets that are approximately linearly dependent by scaling and shifting parameters is approached by Kennedy and O'Hagan [164]. However, due to the presence of nonlinear dependencies generally between the datasets, the quality of results degraded as a major issue for linear data fusion algorithms. To overcome and resolve the nonlinear dependencies, space-dependent scaling factor $\rho(x)$ [165] or alternatively, deep multi-fidelity GP [166] is introduced. Yet the improvement brings further optimizations of additional hyperparameters. Here the NARGP algorithm, an implicit automatic relevance determination (ARD) weight, is employed in the extended space, parameterized by t and s, which counts as a different scaling of the existing hyperparameters for each dimension in the kernel [167].

Additionally, the formulation can be extended through functions of the low-fidelity data set. The high-fidelity data can be further considered as a function t, $f_l(t)$ and the derivatives of $f_l(t)$, exploiting that $f_l(t)$ has a

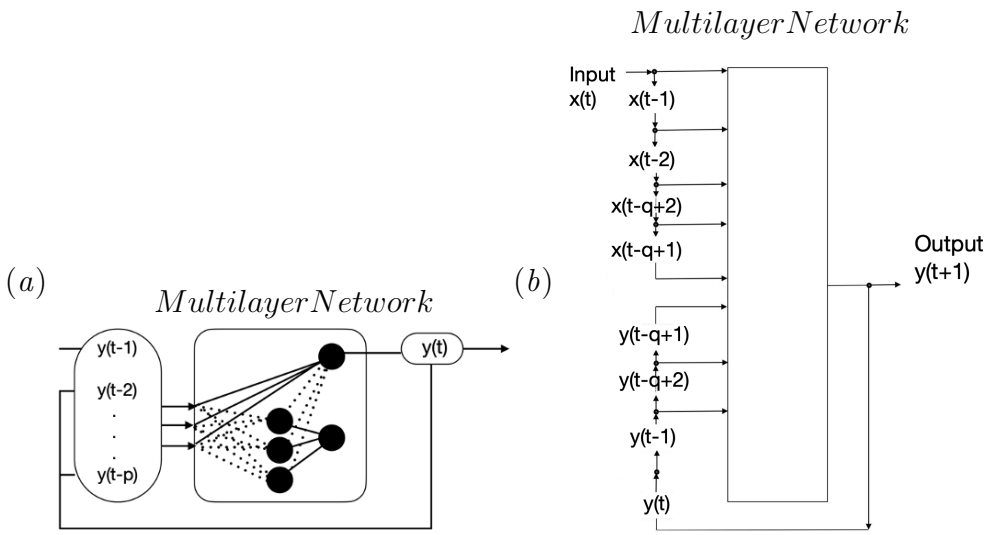

Figure 3.2: (a) Architecture of the NAR (nonlinear autoregressive) model with a multi layer perceptron. (b) Framework for NARX model with an exogenous variable x as the input. q past values of x and y are considered for the prediction of $y(t+1)$.

similar trend with $f_h(t)$ [165]:

$$f_h(t) = g(t, f_l(t), f_l^1(t), ...f_l^i(t)), \tag{3.5}$$

where $f_l^i(t)$ is the i-th derivative of the low fidelity data.

3.2.3 Spatial extrapolation

As aforementioned the target is to predict offshore wind indirectly from onshore measured wind. Therefore apart from the temporal fusion algorithms presented above, a spatial extrapolation is required. Here a time series neural network is adopted to link the wind speed at the two points using a single variable nonlinear network NAR along with NARX to have a fair performance comparison. In practice, low-fidelity WRF data will be used to train the network and onshore hybrid solution (obtained by temporal fusion) is served as input to estimate the offshore wind.

Nonlinear Autoregression (NAR)

The NAR model is most suitable for time-series predictions where the main source of training data is only past values of the time series itself, and this process is called feedback delays. The network is trained in an open loop, which uses the real target values as a response. Following, the network becomes a closed-loop, and the predicted values are used as new response inputs to the network. The framework of this model is seen in Figure 3.2(a), a multi-layer network where the left hand side is the past delayed input values $y(t-1), y(t-2)$, and $y(t-p)$ is used to obtain the independent variable $y(t)$. Optimisation of the network aims to reduce the number of synapses (weights) and neurons, and subsequently reducing the complexity of the network, and maintaining the generalisation capabilities.

Nonlinear Autoregression with External Input (NARX)

NARX is a dynamically guided type of recurrent ANN containing one or more feedback loops. The loops can be either local or regional, and the use of regional loops enables a significant reduction of memory requirements. Recurrent networks are being used for two main functional tasks: first, for associative memory tasks and second, for input-output mapping networks. The applications of input-output mapping networks include modelling and signal predictions for time-series [157].

Meanwhile, different models of NARX networks are comprised of the same structure and thus have a reasonable cost from a computational point of view. The previous argument allows a NARX network to gain degrees of freedom when it includes a time-frame forecast. Input data in subsequent periods compared to a feed-forward network, allowing the synthesis

of information from exogenous variables and less inclusion of their remains, reducing the number of parameters to be calculated.

Learning in NARX networks is more efficient and effective than in other neural networks since it has a better descending gradient, which leads to faster convergence and better generalisation than other networks. The model predicts a series $y(t)$ given specific past values of $y(t)$ and an additional input series $x(t)$ as shown in Figure 2(b), where the NARX architecture is explained. The model has only one input, the exogenous variable $x(t)$, providing feed forward to a q delayed memory neurons, and one output, $y(t+1)$, which is the value of the predicted variable one step ahead. In summary, the input is one step behind the output with respect to time. Hence, the output can provide feedback to the network through the delayed memory neurons, which in turn makes up the input neural layer of a multi-layer network. Dynamically, this is expressed by[158]:

$$y(n+1) = F(y(n), y(n-q+1), u(n), u(n-q+1)), \tag{3.6}$$

where q is the total number of delayed memory neurons, the output $y(n+1)$ is the predicted value representing the one step ahead variable and F is a nonlinear function.

The learning algorithm for the NARX network is based on the performance function used in the training of ANN, which is the mean squared error (MSE). NARX neurons are sigmoid and the performance function is derived to include a mean squared weight function with a performance ratio. As a result, the performance function operates under smaller weights and biases, thus causing the network response to be much smoother and less likely to over fit.

As it was proven essential to have at least two-thirds of the data for training, the training, validation, and test divisions, are partitioned to meet 60 percent training, 20 percent validation, and 20 percent testing, respectively. 10 hidden neurons are constructed with 2 delays, and adopted the Bayesian regularisation as the training algorithm trades off more computational costs for better accuracy [158].

3.3 Case description

To test the methods, a case associated with the RUNE project is considered, which was a near-shore experiment conducted at the west coast of Denmark (see Figure 3.3 (b)). The surrounding area is nearly flat coastal farmland and moving northwards from position 1 to 3, the sand embankment separating the North Sea and the grasslands transforms into cliffs covered by grass. In this work, dual-Doppler scans performed nearly perpendicular to the coast from about 5 km offshore to 2 km onshore are used. These scans are performed by synchronising measurements from two scanning lidars, which are modified versions of WLS200S Leosphere units, one located at position 1 and the other at position 3. Here, the dual-Doppler scans performed at 50 m above mean sea level (amsl) during the period 2015-12-08 to 2016-02-17. Due to filtering of high noise/low signal strength and system availability, only 114 10-min are available at all the dual-Doppler positions shown as black markers in Figure 3.3 (a). Further details with regards to the experimental campaign and the instrumentation can be found in [168].

Here a numerical experiment, which was part of a number of numerical simulations performed using the WRF model v3.6 to supplement the mea-

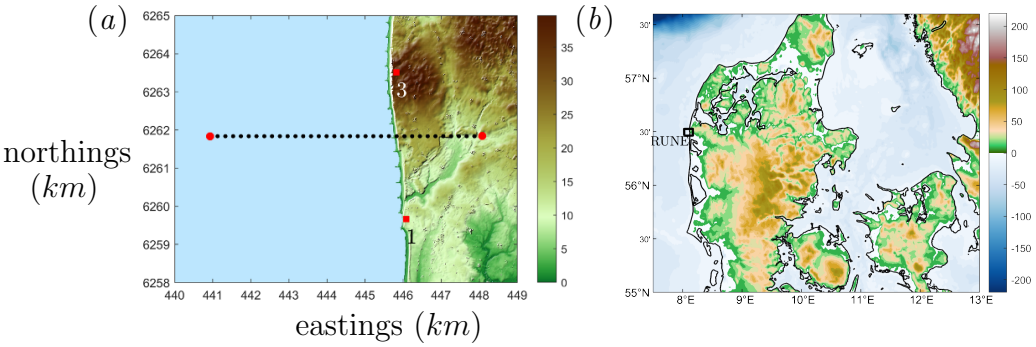

Figure 3.3: (a) RUNE experimental area. Positions of the lidars are shown in red square markers and the dual-Doppler scans in black and red dot markers. The colour bar indicates the terrain elevation in meters above mean sea level. (b) The location of the RUNE experiment (black rectangle) in Denmark.

surements of RUNE [170]. This particular experiment was setup with 4 nested domains, the outermost covering northwestern Europe and a 2-km horizontal resolution innermost domain covering the west coast of Denmark. Spectral nudging to the ERA5 reanalysis is used in the upper model levels of the outermost domain. The simulation had 8 vertical levels within the first 100 m and instantaneous output was produced every 10 min. The experiment also used the Mellor-Yamada Janjic planetery boundary layer scheme, a sea surface temperature product from the Danish Meteorological Institute [169], and the CORINE land cover description. An illustration of the low fidelity WRF simulation data set of the most onshore point can be seen in Figure 3.4(a), while Figure 3.4(b) shows the high fidelity lidar measurements for the same point.

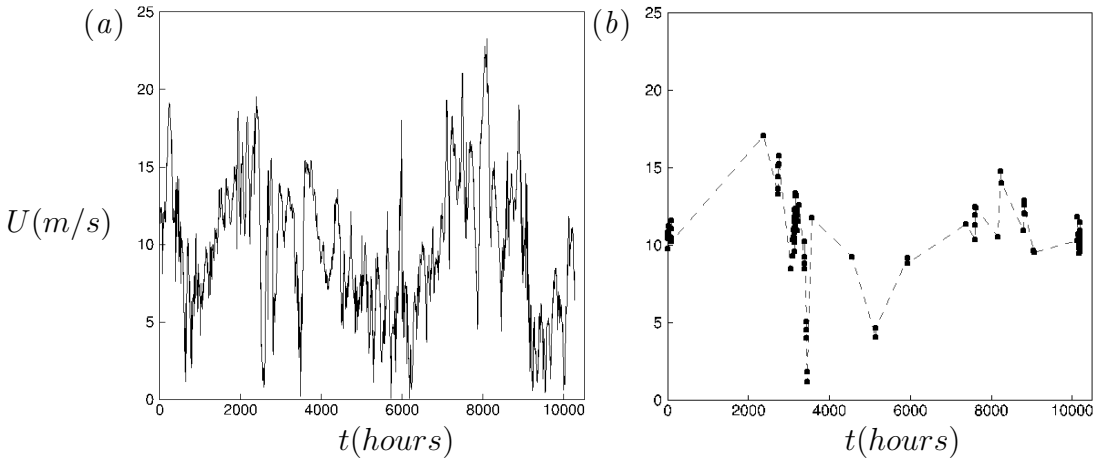

Figure 3.4: (a) Low-fidelity data from numerical simulation (WRF) at the most onshore point.(b) High-fidelity data from the dual-Doppler lidar setup at the most onshore point, (see Figure 3.3 a).

3.4 Results and Discussion

3.4.1 Temporal data fusion

The different data sets at the furthest onshore point of the dual-Doppler line are merged by applying MF-GPR. The performances of the three models are optimized, through the hyperparameters, by applying 30 iterations of basis and kernel function combinations, including: Zero, Linear, Constant and Matern 5/2 and 3/2; Rational Quadratic, squared Exponential, etc. Firstly, the original model is explored, represented as Hybrid (1), where the input information is low and high-fidelity data sets. Then, introducing additional information sets, which are functions of the low fidelity data set (first and second derivatives) can enhance the accuracy of predictions, hence Hybrid (2). The third model, Hybrid (3), involved pre-processing of the training set using the EWT reconstruction algorithm, the regenerated first and second derivative sets of the low fidelity data, and finally the North and East decomposed wind speed vector components. A higher drop in RMSE is noted from the Original WRF data.

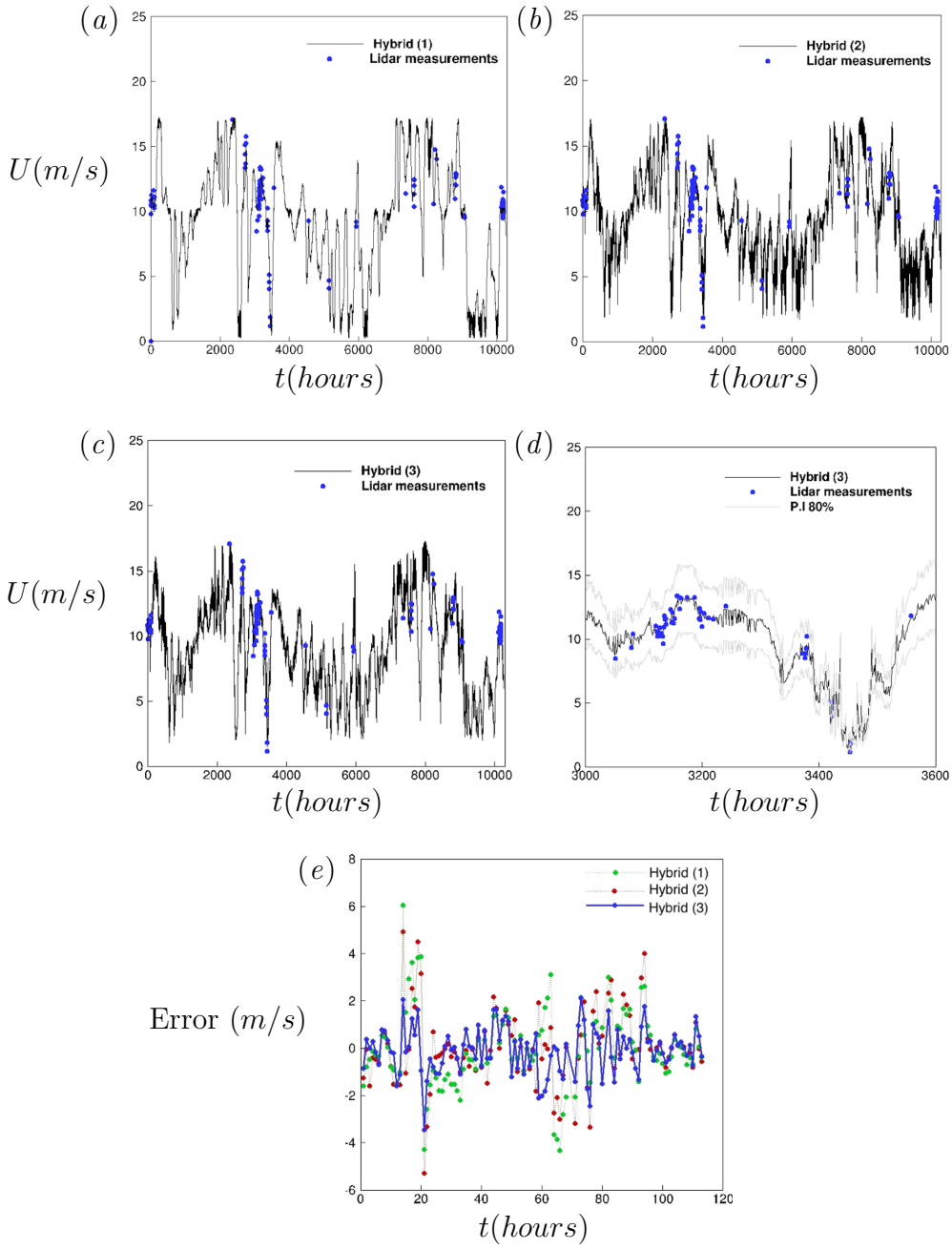

Figure 3.5: (a), (b) and (c) represent temporal data fusion results for Hybrids (1), (2) and (3), respectively. (d) Cut-off panel for Hybrid (3) from 500 to 600 hours of the experiment and (e) Error in all three hybrid models against 114 lidar measurements.

Table 3.1 shows the RMSE for the best performing GPRs for the three hybrid models and WRF data using the lidar data as the ground truth, for each model as well as their respective basis and kernel function configuration. It can be seen that the third model outperformed the other two in terms of RMSE superbly.

Figure 3.5(a), (b) and (c) show the time series of the results from the three hybrid models, respectively. Hybrid (1) showed a 9% decrease in RMSE from WRF data, in Hybrid (2) the RMSE is reduced by about 18%, and finally Hybrid (3) showed the largest drop compared to other hybrid models, about 31%. Besides, panel (d) reflected a cut off to the interval between the hours 500-600 since the start of the experiment, with an 80% confidence interval, showing a better visualization of the performance. Thus, carrying a sensitivity analysis to test the different approaches and additional features, it is concluded that the addition of data from different data sources adds 9% improvement. The addition of input information leads to more improvement of 10%, and the pre-processing of signals results in further improvement of 14%.

Finally, panel (e) presents the deviation error between the provided high-fidelity time series of 114 points with their predicted counterparts by the hybrid techniques. It can be seen that for some points, Hybrids (1) and (2) outperform each other with deviations ranging from 6 to 2 m/s. On the other hand, the performance of Hybrid (3) outperformed the other two at almost all measured points with the deviation capped at 2 m/s and mean deviation 0.6 m/s. These results substantiate that increasing the number of additional sets and pre-processing the data enhanced the accuracy of the Gaussian process [165].

Furthermore, the work also compared the three algorithms developed for

Table 3.1: Configurations and accuracy of the GPR models.

Model	Basis and Kernel functions	RMSE [m/s]
WRF	-	1.24
Hybrid (1)	Zero, Matern 5/2	1.15
Hybrid (2)	Constant, Rational quadratic	1.01
Hybrid (3)	Zero, Matern 3/2	0.86

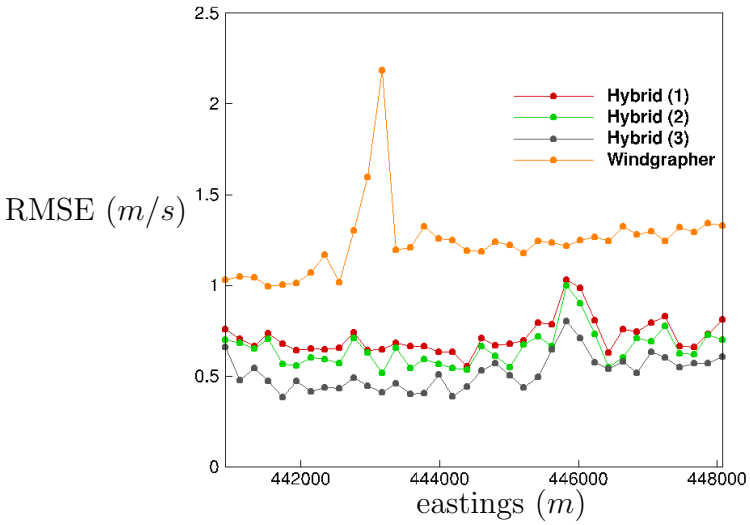

Figure 3.6: RMSE curve across Windgrapher and all three hybrid models for all 36 dual-Doppler points.

multi-fidelity GPR with a high standard industrial software. The focus is on comparing the accuracy of predictions generated by the proposed algorithm in Hybrid (3) with respect to current leading industrial standards, to test the performance of the algorithm with respect to different geographic conditions. On a separate iteration, temporal data fusion is conducted using all 3 Hybrid methods and the software from Windgrapher, which is the leading industrial software for importing, visualizing, and analysing wind resource data. Windgrapher follows the Measure-Correlate-Predict (MCP) algorithm including Linear Least Squares (LLS), the method is on correlating target and reference speed data based on the linear least squares procedure. The RMSEs for each of the 36 dual-Doppler points using all 3 hybrid methods and the industrial software are shown in Figure 3.6.

In Figure 3.6,a comparison of the results from the three Hybrid techniques

to Windgrapher by calculating the RMSE of results across all 36 dual-Doppler positions from the furthest offshore point to the most onshore one. On average, Hybrids (1), (2), and (3) are able to perform 12%, 14%, and 60%, respectively, more accurately than the industrial software. In addition, the hybrid methods showed a higher consistency in predictions that occurred offshore, where the industrial software had a relatively poor performance. Moreover, the highest RMSE for the hybrid methods is observed at 446 km eastings, where the transition from offshore to onshore takes place. Meanwhile, the industrial software not affected and the performance of the predictions is consistent in this region.

3.4.2 Spatial data fusion

The spatial data fusion aimed to project the onshore measurements to offshore locations in light of numerical simulations, to reduce the cost of direct offshore measurements. The first and last offshore and onshore points in the 36 dual-Doppler line are considered as an example. An ANN was trained using the low-fidelity WRF data at both points, and it configured the winds' relation at both points. The network is later tested on the high-fidelity lidar data of the onshore point and generated high fidelity wind speed results for the offshore point [163].

For the NARX model, eleven simulations are performed varying the number of past values (delays) for the entry variables from 1 to 10, and the number of hidden neurons from 3 to 21. The MSE of the test data set is used to assess the performance of the network, the configuration with the best performance is employed. The same number of simulations are carried for the NAR model; similarly, the past values varied from 1 to 10 and the hidden neurons from 3 to 15 and the configuration with the lowest MSE is

selected.

Table 3.2 shows the configuration and performance of NAR and NARX for the training of the network. The highest performing NAR and NARX configurations consisted of 3 delays, 15 neurons, and 4 delays, 12 neurons, respectively, which implied that the latter required less statistic training. Despite hiring a lower number of neurons, the NARX network required less computation time to outperform the NAR network by 12%.

Figure 3.7(a) shows the results from testing the NARX network, the low-fidelity WRF data and high-fidelity lidar measurements of the most offshore point. Predicted values from the network are more accurate than the low-fidelity data, where the RMSE is reduced from 1.23 m/s to 1.17 m/s. For a zoomed-up visualization of results, the data-rich high-fidelity section at hours 500-600 is shown in Figure 3.7(b).

Table 3.2: Configurations of the NARX and NAR Networks with the Best Performance.

Model	Delays and Neurons	Time steps	MSE ($\times 10^{-}2$)
NAR	3, 15	Training (6176)	1.352
		Validation (2058)	
		Test (2058)	1.559
NARX	4, 12	Training (6176)	1.241
		Validation (2058)	
		Test (2058)	1.188

3.4.3 Spatio-temporal extrapolation

In this section, temporal and spatial data fusion using 1331 10 minute lidar measurements are combined. The aim is to use the intermittent measurement at the second most onshore point to estimate the wind at the most offshore point by exploiting the numerical data.

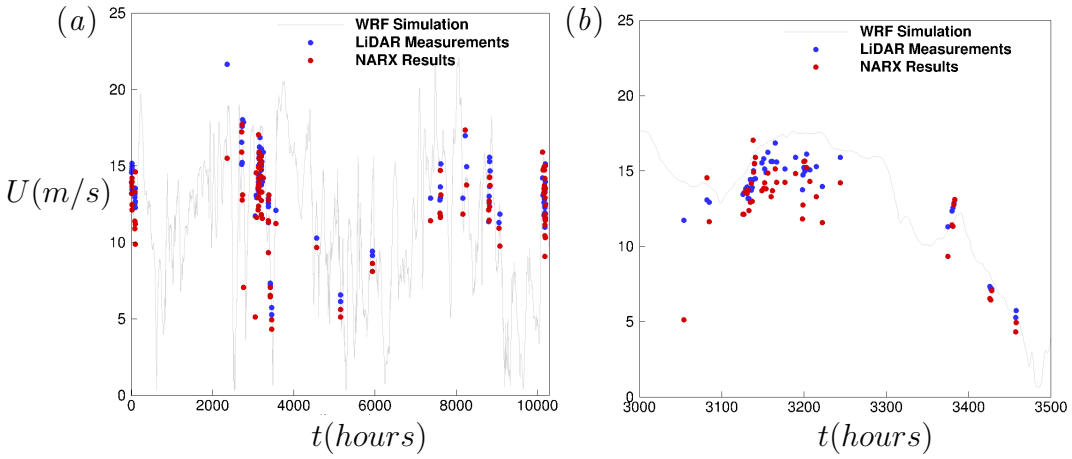

Figure 3.7: (a) Spatial data fusion results for the offshore point. (b) Close-up of spatially predicted data and high and low-fidelity data for validation.

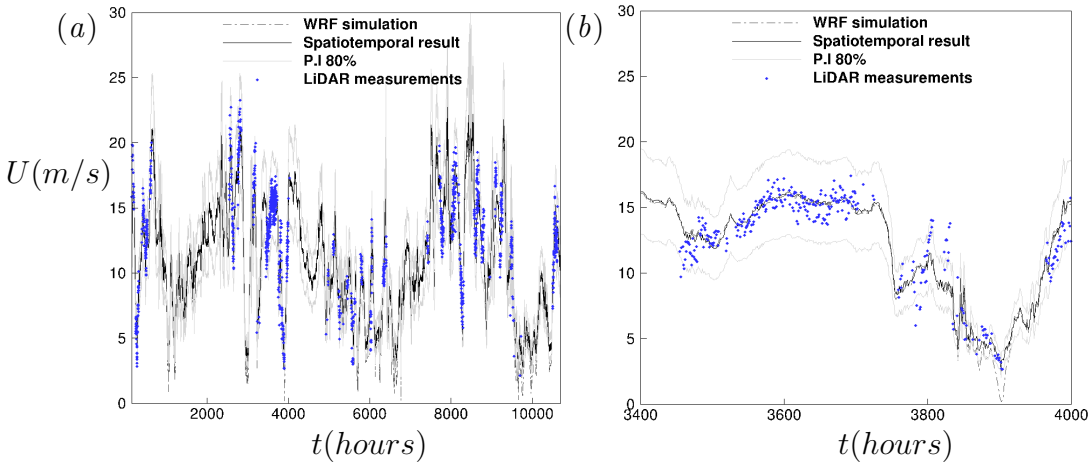

Figure 3.8: (a) Response for Spatio-temporal data fusion. (b) Close-up of the final spatiotemporal results ranging from hours: 3400-4000.

Temporal data fusion is performed following the same technique as Hybrid (3) of MF-GPR, which used the reconstructed set from the EWT algorithm for pre-processing and five predictors for the GPR algorithm: low-fidelity WRF data, first and second derivatives, North and East vector components of the wind speed set of the onshore point. Again, 30-iterations are used to optimize the hyperparameters, varying the basis and kernel functions to achieve a configuration with an RMSE of 0.84 m/s.

Following, a NARX neural network is trained using the low-fidelity WRF

data of both the second most onshore and most offshore points with 3 delays and 12 hidden neurons. The network is later tested on the time series generated from the temporal section using MF-GPR and the performance of the network is estimated at an MSE of 1.12 $\times 10^{-}2$.

Figure 3.8(a) shows the final curve of the spatial-temporal data fusion process, high-fidelity lidar data (hidden in the assessment), and 80% prediction intervals. In addition, Figure 3.8(b) is a cut-off to show the region with the richest lidar data, the results are satisfactory, as it achieved an accurate result with an RMSE 1.23 m/s impersonating the high-fidelity data of the most offshore point without discontinuity or using expensive lidars offshore. These results outperformed the WRF simulation at the offshore location where the RMSE is 1.46 m/s.

3.5 Conclusions

In this work, data fusion of numerical model results from WRF simulations (low fidelity) continuous in space and time with lidar measurements (high fidelity) sparse in space and time is performed to obtain spatial-temporal extrapolation suitable for the assessment of offshore wind. The RUNE experiment performed dual-Doppler scans which generated 114 10-minute measurements over three months, used in spatial-only and temporal-only data fusion. Then, 1331 10-minute measurements are used in the spatiotemporal experiment. Simultaneously, numerical simulations performed using the WRF model v3.6 generated an instantaneous output every 10 minutes for the same period of 3 months.

For time-domain data fusion, the model is able to represent the high-fidelity data at unobserved regions and periods by exploiting the low-fidelity data

and its functions. The addition of extra information datasets (derivatives and wind speed vector components) and pre-processing showed an improvement in the prediction performance in terms of RMSE, with a 30% average drop compared to other models that ignored them. Similarly, for the spatial fusion part of the experiment, adding extra information datasets to the NARX neural network showed improved results compared to the single input model NAR. The data from the network had a lower MSE for assessing offshore data, which could avoid sending expensive equipment offshore, hence reducing the cost of offshore wind resource assessment.

Following data fusion in both space and time, the models have been re-run at the observed optimized levels in each method. Data from the second most onshore point underwent Hybrid (3) MF-GPR for time-domain fusion, and the continuous time series output has been used along with data from the offshore point for space domain fusion. Finally, the spatial-temporal data fusion resulted in accurate offshore wind resource assessment within a 2% margin error for wind speed.

There are two major limitations in this experiment that could be addressed in future research. First, the experiment focused on data obtained from the RUNE experiment, which included only 1331 measured points from lidar equipment equivalent to 220 hours of measured data, scattered across the entire 3 months period of the experiment, which means a lot of weeks did not have high-fidelity data. Contrarily, the WRF simulation covered the entire duration with almost 10,500 points. The limited access to high-quality data is a result of low availability of lidars mainly due to the harsh weather conditions offshore that often damage the equipment. The rareness of lidar data can influence the GPR, which makes it harder to notice trends and create meaningful observations.

Second, the most onshore and offshore points of the 36 dual-Doppler are selected for the experiment, however, due to the low resolution of WRF (2 Km), the selection of corresponding WRF points is based on choosing the closest WRF point, which at utmost closeness is a few meters far off the lidar point, reducing the accuracy of predictions.

Different adaptations, tests, and improvements have been left for the future. Future work concerns further development to the multi-fidelity data fusion algorithm, by introducing additional sources of data generated by different equipment or software. An example of this is the use of a second WRF simulation with different resolutions and features. Nevertheless, concerning the results for one point-based prediction, 3-dimensional area predictions with respect to time by performing data fusion of multiple lidar measurements at different locations with full grid WRF simulation results can be performed.

The results from this chapter clearly illustrate a benefit in merging wind speed datasets generated using different sources. The temporal fusion is capable of combining the merit of continuous data generation from the WRF simulation and high accuracy from the high-fidelity lidar observations, providing a continuous accurate time-series of wind speed predictions. Additionally, the spatial propagation of data from onshore to offshore locations, successfully delivers the objective to avoid sending expensive lidar equipment to offshore locations.

In the following chapter, a hybrid forecasting model consisting of secondary decomposition pre-processing algorithm, a prediction deep neural network, and an optimization algorithm will be demonstrated. The majority of literature is focused on training models with a single source of data, usually numerical simulations due to their availability and low cost, however they

are considered as a low-fidelity source of data, which should be fused with lidar measurements to improve their accuracy. The proposed hybrid forecasting model is a novel algorithm, trained using only simulation data. Final results are compared to the lidar measurements and the overall performance is assessed with the model from this chapter to have a visualized performance comparison of the models that only train on simulations with multivariate models that consider both simulations and a higher-fidelity data source (lidar measurements).

Chapter 4

Hybrid Deep Neural Networks for Wind Speed Forecasting

Wind resource assessments are critical to the pre-construction of wind farms. For efficient exploitation of wind power, accurate and reliable wind speed forecasting are necessary. A wind speed time series is considered a signal with non-stationarity and non-linear characteristics. Therefore, pre-processing of signals using signal decomposition is proposed to improve accuracy of forecasts. However, due to the high frequency modes of the signal, single model decomposition models are no longer sufficient to obtain accurate results. Hybrid models are being developed to combine merits of signal decomposition approaches. A hybrid approach is proposed here to combine the merit of model decomposition using Complete Empirical Mode Decomposition With Adaptive Noise (CEEMDAN), and wavelet transform using Empirical Wavelet Transform (EWT). EWT is employed to further decompose the high frequency signal from CEEMDAN, thus reducing forecasting complexity. Then, an improved Bidirectional Long Short Term Memory (BiLSTM) with Grey Wolf optimizer (GWO) algorithm is

applied to forecast all of the decomposed Intrinsic Modal Functions (IMF) and modes. The numerical wind speeds (WRF) along the west coast of Denmark are used to evaluate the method. The results are then compared to lidar measurements to assess if numerical simulations are sufficient to obtain accurate forecasts when compared to a high-fidelity data source. Finally, results are also compared to a forecasting method that employs data fusion, merging lidar measurements and numerical simulations to a model to investigate the benefit of data fusion of different sources of data with different fidelities. It is shown that despite the proposed hybrid model performing efficiently in mimicking the numerical simulation, the performance is rather poor when compared to the lidar measurements, concluding that models utilizing fusion of different sources of data generate more accurate predictions.

4.1 Introduction

With rapid economic development and continuous rise in standards of living, the human demand for energy has been significantly increasing. Additionally, the use of fossil fuels such as coal, oil and natural gas have resulted in huge amounts of greenhouse gases in the atmosphere, damaging the environment and rising the global temperature, leading to global warming. Alternately, investments in renewable energy have significantly increased and have been receiving more attention lately. Wind energy has the second largest potential after solar energy and has received extensive global attention. The Global Wind Report 2021 has published that 93 GW of new wind energy was built and installed in 2020, showing a growth of 53% compared to 2019, bringing the total capacity of wind energy installed to 743 GW, a 14.3% compared to its previous year[171].

Wind energy generation has been a major component of the smart grid and has a significant role in the supply and management of electricity. However, wind energy is an intermittent source of energy due to the stochastic and low predictability nature of wind, which has a great influence on the stability and safety of the grid-integrated wind power systems [6]. To solve the resulting scheduling, management, and optimization challenges, accurate forecasting of future wind energy values is of paramount importance [8]. Forecasting of wind energy also contributes to solving other issues in the wind energy sector such as reducing operating costs and enhancing the competitiveness of wind energy. Therefore, wind forecasting is a key technology in integrating wind energy for existing different scale electricity grids. To produce wind energy forecasts, literature suggests two techniques, first, wind power is directly predicted from historic wind power data, however, this is not usually available when the forecasts are used to assess the wind resources of a new location. The second technique is to forecast wind speed first then produce wind power forecasts based on wind power curves. Scholars have paid both methodologies significant attention, however, for this work, the focus was on wind speed predictions as the main aim is to provide wind resource assessments for potential wind farm locations based on available wind speed data [7].

In the past decade, numerous approaches were proposed and implemented to forecast wind speed. With different forecasting time horizons, models can be grouped and categorised to serve different challenges. Time horizons range from very short-term predictions (few seconds to 30 minutes ahead) to long-term predictions (1 day to 1 week or more ahead). Subsequently, forecasting models are grouped under four groups based on the modelling theory followed, which include physical models, statistical models, machine learning based models, and hybrid models. In models that follow the phys-

ical approach, models such as numerical weather predictions and weather research forecasting take into consideration meteorological factors such air pressure, humidity, and temperature. The literature proposes that physical approaches had most accurate performance when forecasting wind speeds in long-term forecasting. Contrarily, traditional statistical models include Auto-regressive Moving Average (ARMA) and its variant models, which focus merely on characterizing the linear relationships in a time series. These models had generally showed superior performance in the very short-term time range [172].

On a different perspective, scientists are harnessing the advances in machine learning by employing AI-based models in wind speed forecasting [108]. Different types of neural networks such as long short-term [114], and convolutional [118] are receiving extensive attention due to their superior capability in capturing and dealing with non-linearity in datasets. Several studies showed that machine learning based models were superior in performance to statistical models [111]. Moreover, following the success of employing machine learning models in wind speed forecasting, many studies suggested combining different algorithms to characterize different aspects of wind speed fluctuation. For example, Convolutional Neural Network (CNN) were used to capture spatial features of a wind speed dataset, results were then fed into a Long Short Term Memory (LSTM) neural network to generate temporal features. Together, a combination of both neural networks proposed a hybrid forecasting model, CNN-LSTM, which captures spatio-temporal features, making full use of space-time information in the dataset [124].

Furthermore, outside the choice of predictor models, the performance of forecasting approaches was seen to be significantly enhanced by first preprocessing the data used, then setting the model configuration using in-

telligent optimization algorithms that provide optimal parameters for the predictor model [70]. For the former, wind data pre-processing considers the wind speed time series as a signal that can be decomposed into different sub-series and predicted results of all sub-series were then aggregated to produce a forecast of the original time series [71]. Signal processing methods are used as denoising algorithms that generate decomposed sub-series with removed noise more suitable for training forecasting models. From the available approaches in signal pre-processing, wavelet-based approaches can decompose the original wind speed series into different levels. Contrarily, other approaches can only produce all sub-series at one level. Wavelet-based approaches include algorithms such as Wavelet Decomposition (WD), EWT, and Empirical Mode Decomposition (EMD). In addition to employing a signal decomposition algorithm, hybrid approaches were proposed to combine merits of different algorithms [70]. A secondary decomposition approach considers more than one signal decomposition algorithm and have shown improvements for wind speed forecasting. On the optimization side, intelligent optimization algorithms are employed to provide optimal hyperparameters for the forecasting model. For machine learning based forecasting models, model configuration hyper-parameters include, but are not limited to, weights, batch size, learning rate, and number of hidden layers. Examples of optimization algorithms include grey wolf optimizer, multi-objective bat algorithm, and multi-objective multi-universe optimization [134, 138, 139].

4.1.1 Motivation and contribution of this work

In recent years, several wind speed forecasting models have been proposed and published. The literature on hybrid models for predicting wind

speed contains more than 250 articles testing different combinations of approaches, where the majority provide more accurate forecasts and enhanced performance relative to previous work. However, the majority of studies conducted experiments on a single source of data generated using simulations for its wide availability and relatively low cost. However, data generated using simulations is considered of relatively low fidelity and despite achieving high accuracy, results are rarely compared to other sources of data such as lidar observations or mast measurements. Therefore, this work investigated a hybrid approach that employed CEEMDAN and EWT algorithms for data processing, BiLSTM neural networks for the prediction model, and GWO to optimize the weights of the BiLSTM networks using WRF simulation data. Prediction results are then compared to lidar observations to accurately assess the performance of the model when compared to high-fidelity wind speed data. The main objective of the work is to demonstrate the importance and added value of data fusion between high-fidelity data generated using lidars and low-fidelity data from simulations for the assessment of wind resources.

(1) Original wind speed time series is pre-processed using Complete Empirical Mode Decomposition With Adaptive Noise (CEEMDAN) and the signal is decomposed into several Intrinsic Modal Functions (IMF).

(2) Since IMF1 is very unsystematic and is the signal with the highest frequency amongst all IMFs generated by CEEMDAN, Empirical Wavelet Transform (EWT) is employed for double decomposition, to further reduce the noise and increase the forecasting accuracy.

(3) The decomposed IMFs generated using CEEMDAN and modes by EWT excluding IMF1 are fed into Bidirectional Long Short Term Memory (BiLSTM) to predict the signals.

(4) An optimisation phase using Grey Wolf optimizer (GWO) is assigned to provide the optimum weights for each IMF and mode output from the BiLSTM models.

(5) Then results are aggregated to generate the final signal forecast.

(6) Final predicted signal is compared with results from data fusion models to demonstrate the importance of using multiple sources of data with different fidelities.

4.2 Methods for Hybrid Deep Neural Networks for Wind Speed Forecasting

In the methodology section, pre-processing of signals using a CEEMDAN model is covered in Section 4.2.1, which reduces fluctuations in time-series, by decomposing the original signal into different sub-series. Subsequently, in section 4.2.2, a special type of recurrent neural networks, LSTM is covered to then introduce BiLSTM, which provides both past and future data to train the model. A general overview of Grey Wolf Optimizer is introduced in section 4.2.3, which is an optimization algorithm that works to provide the most appropriate parameters for the forecasting model. This section is based on implementing a hybrid forecasting model for estimation of missing data from simulation data.

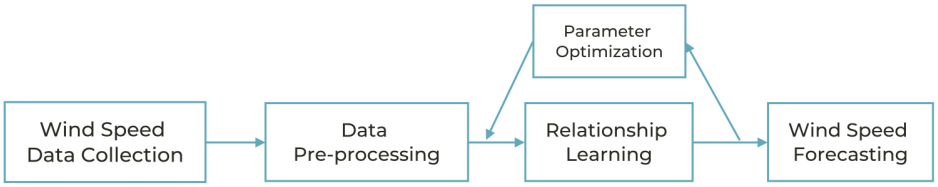

Figure 4.1: Stages of building a forecasting model, which include collection of data, pre-processing the signals, training a forecasting model with optimized hyper-parameters from optimization algorithm

4.2.1 Data Pre-processing (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise):

For this chapter, data will be pre-processed using two wavelet-based approaches, complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and empirical wavelet transform (EWT). Wavelet-based approaches are pre-processing algorithms that can decompose the original wind speed time-series into different intrinsic levels, which gives it an advantage over other pre-processing algorithms. Additionally, signal processing approaches process the entire dataset indiscriminately and hence are usually used for two purposes, data decomposition and data denoising. Combined pre-processing algorithms have shown higher efficiency for wind speed forecasting. In most cases, the data time series is decomposed into different sub series with various frequencies. Then, the higher frequency sub series is further decomposed using a different algorithm to catch in-depth trends in the dataset. The predictability of the decomposed sub series is found to be stronger than that of the original series as it contains less noise, hence will generate more accurate forecasts of each level of the original dataset [173]. Results from a study which combined Wavelet Packet Decomposition (WPD) with Fast Ensemble Empirical Mode Decomposition (FEEMD) has shown that forecasts from hybrid pre-process data is far more accurate than a single signal processing approach [174].

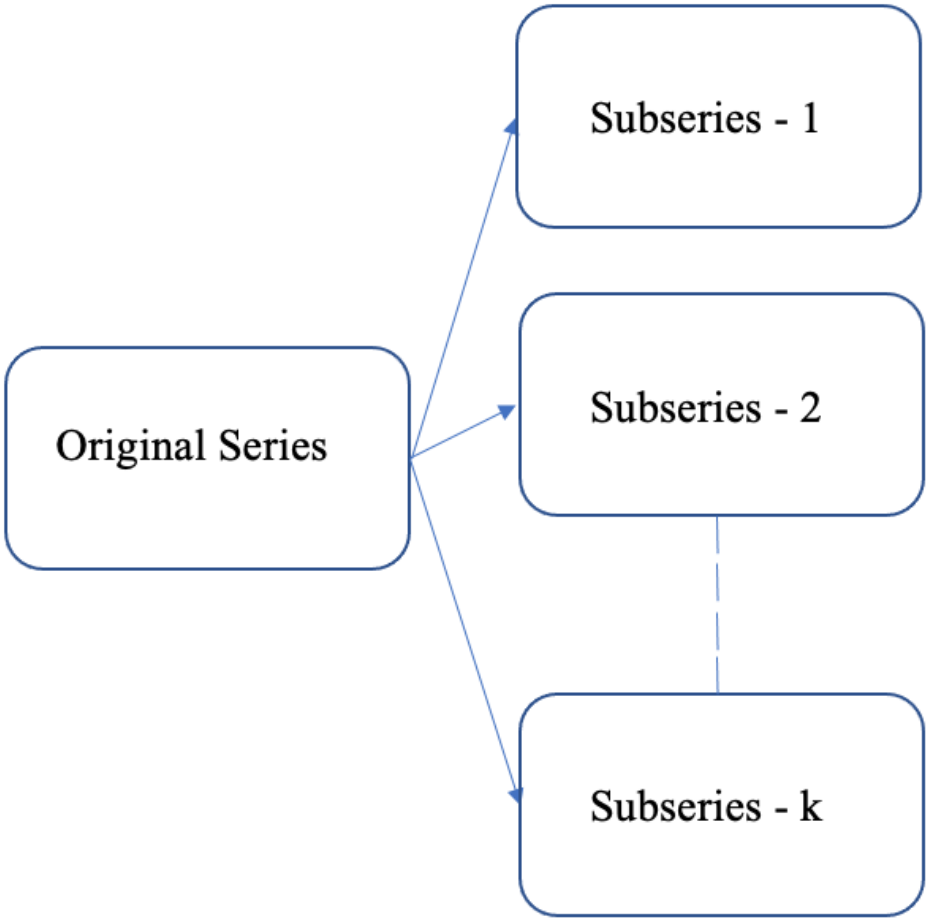

Figure 4.2: Pre-processing of original signal using CEEMDAN to decompose signal into multiple sub-series.

Therefore, further studies investigated combining different approaches and several studies have developed secondary decomposition algorithms with accurate results, some of the developed algorithms are summarized in Table 4.1. For this work, CEEMDAN will generate different frequency intrinsic mode functions and the high frequency level is later further decomposed by EWT to reduce prediction complexity.

In [175], Huang et al. developed the empirical mode decomposition technique, an adaptive data analysis method that deals with both non-linearity and non-stationarity in signals. The aim of EMD is to decompose complex signals into a finite number of IMF based on the signal's local characteris-

Table 4.1: Existing Secondary decomposition algorithms.

Article	Secondary Decomposition Method
Mi et al.	Wavelet Packet Decomposition (WPD) +EMD
Wu and Xiao	EWT + Singular Spectrum Analysis (SSA)
Moreno ET al.	Variational Mode Decomposition (VMD) + SSA
Peng et al.	CEEMDAN + VMD
Lie et al.	EEMD + WPD

tics. The IMFs comply with two conditions: (a) the entire data set contains extremes and zeros that are either equal or differ by one unit, and (b) the mean value defined by the local minima and local maxima is zero at any point of the defined envelope. The EMD process of the original signal can be defined as follows:

$$n(t) = \sum_{k=1}^{m} imf_m(t) + r_m(t) \tag{4.1}$$

Where n(t) represents the non-linearity or non-stationarity in the signal, $imf_m(t)$ is the m-th IMF of the signal, and $r_m(t)$ is the residual.

Later, to overcome the drawback of mode mixing in EMD, a noise-assisted analysis of the data, EEMD, was proposed by Wu and Huang [176]. In Ensemble Empirical Mode Decomposition (EEMD), the true IMF components are identified based on the mean of an ensemble of trials, hence the noise-added signals of EMD are decomposed by white noise incorporated into the original signal, providing a uniform reference scale, facilitating the EMD process, and helping extract the IMFs. The white noise is finally cancelled due to averaging of the ensemble. However, as each trial adds white noise to the decomposition result, the added white Gaussian noise signal

after a finite number of iterations results in a reconstruction error that may not be eliminated, and the accuracy of the forecasts will be affected. It is possible to decrease the reconstruction error by increasing the number of iterations at a high computational cost. Contrarily, a complete EEMD with adaptive noise, CEEMDAN, was developed by Torres et al. [177] providing three main advantages: (a) a noise coefficient value to control the noise level at each decomposition, (b) complete and noise-free reconstruction of signal, and (3) less trials are required compared to other approaches. The decomposition process of CEEMDAN is as follows:

Step 1: A number of noise-added series is generated:

$$n^i(t) = n(t) + p_0 w^i(t) \tag{4.2}$$

where n(t) denotes the original signal, $w^i(t)(t = 1, ..., I)$ denotes different white Gaussian noise with $N(0, 1)$ and p_0 is a noise coefficient which controls the signal-to-noise ratio.

Step 2: Decompose each of the generated noise using EMD to get the corresponding first modes $IMF_1^i(t)$. Then calculate the first mode of CEEMDAN by averaging all modes:

$$\overline{IMF_1^i(t)} = \frac{1}{I}\sum_{i=1}^{I} IMF_1^i(t) \tag{4.3}$$

Step 3: Calculate residual $r_1(t) = n(t) - \overline{IMF_1^i(t)}$ and decompose the noise-added residual $r_1(t) + p_1 E_1(w^i(t))$ to obtain the second mode:

$$\overline{IMF_2^i(t)} = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1(t) + p_1 E_1(w^i(t))) \tag{4.4}$$

where $E_1(\bullet)$ is a function to produce the first mode of EMD.

Step 4: Process is repeated to obtain the following modes until the residual component does not have at least two extreme values.

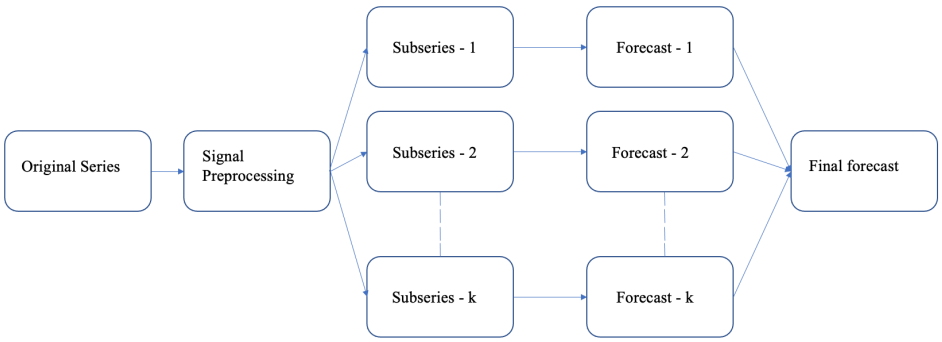

Figure 4.3: Flowchart of the proposed forecasting model.

4.2.2 Recurrent Neural Network (RNN):

Through a circular internal outline feedback connection that enables the use of past information, Recurrent Neural Network (RNN), a member of the neural network family, solves the long-term dependence of time series. Therefore, they provide a reliable and robust solution for processing time series data with variable time periods. Despite their suitability for processing time series with variable time lengths, RNNs have a major drawback, where the gradient tends to disappear for long time series and the neural network starts suffering from short-term memory [178].

Long-short Term Memory (LSTM):

In 1998, Hechreiter et al. [179] proposed the concept of LSTMs to overcome this drawback. As a special variant of RNNS, LSTMs propose a gating mechanism, which gives the traditional RNN the ability to store or

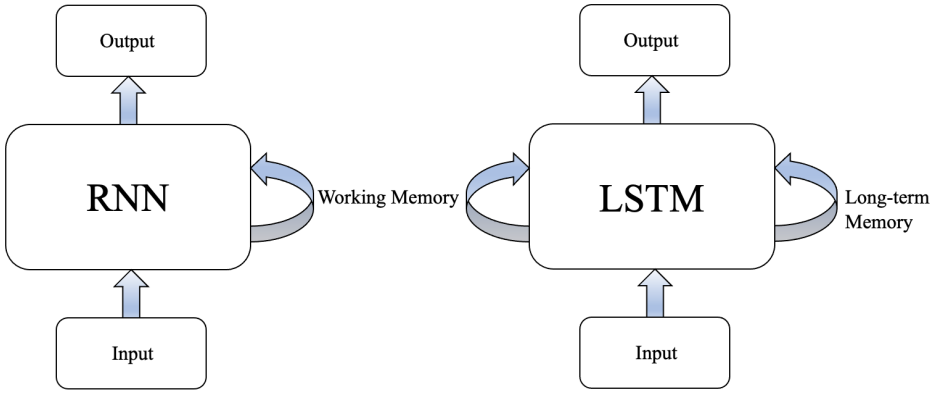

Figure 4.4: General architecture of RNN and LSTM. In RNN, the network maintains information over time in the working memory. However, LSTM networks add a long term memory for temporal changes, which allows them to solve the vanishing gradient problem in a regular RNN.

forget temporal state information. A classic LSTM is composed of a cell, an input gate, an output gate, and a forget gate. The cell stores information over arbitrary time intervals, while the memory block in the network which consists of the gates regulates the information flow into and out of the cell. Specifically, the forget gate determines the information to be retained from the past memory cell, the input gate determines the part of the information to be updated, and the output gate determines the information to exported from the memory block. Hence, LSTM holds the advantage of using additional long-term memory to remember past information, which is why they are widely used in wind speed forecasting and enhancing the accuracy of time series predictions. Being one of the most suitable approaches for time series predictions, many researchers have used LSTM to obtain deterministic or probabilistic wind speed forecasts. In [180], LSTM generated deterministic forecasts by producing prediction intervals using a beta distribution tuned for the counterpart forecasting error. In [181], an LSTM-based model was developed to generate day-ahead hourly wind speed forecasts by designing a multi-scale network that integrated information for each temporal scale. The status update process which forms

the structure of a memory block in an LSTM is shown in Figure 4.4 and a detailed implementation of the corresponding gates is as follows:

Input gate $i_t = \sigma(w_i \cdot [x_t, h_{t-1} + b_i])$

Output Gate $o_t = \sigma(w_o \cdot [x_t, h_{t-1}] + b_o)$

Forget Gate $f_t = \sigma(w_f \cdot [x_t, h_{t-1}] + b_f)$

Cell state $c_t = c_{t-1} * f_t + (tanh(w_c \cdot [x_t, h_{t-1}] + b_c)) * i_{t-1}$

Hidden state $h_t = tanh(c_{t-1}) * o_t$

Where c represents a cell state, $h(t-1)$ is the hidden state, W represents the weight matrices, b represents the biases, which are not time-dependent, additionally, the activation function for the three gates is the hyperbolic tangent function tanh(.), and the activation function of the state update is the logistic sigmoid function sigma(.), which are defined as follows:

$\sigma(x) = \frac{1}{(1+e^{-x})}$

$tanh(x) = \frac{(e^{x}-e^{-x})}{(e^{x}+e^{-x})}$

The above equations are all computed for one-time step, implying that this set of equations are recomputed for every next time step. Moreover, as weight and biases are time-independent, they remain constant in every iteration for the set of equations in each time step.

Bidirectional Long-short term memory (BiLSTM):

A major advantage for bidirectional LSTM is that they process sequential data in both forward and backward directions, where forecasting is not only dependent on past information but also future data in the time series.

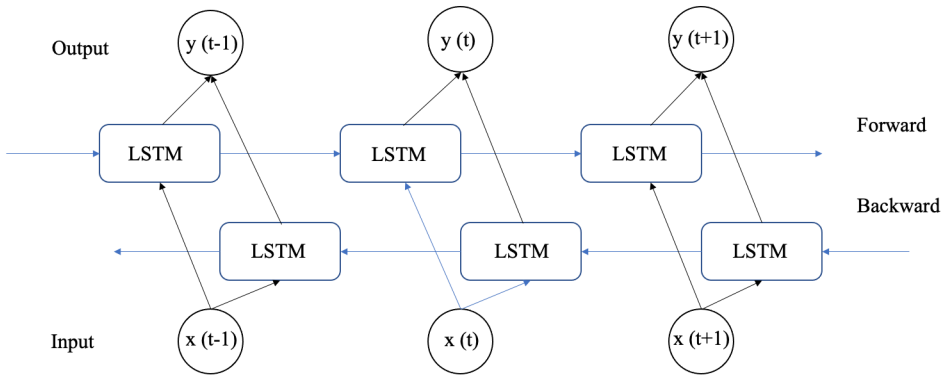

Figure 4.5: Flowchart of the architecture of BiLSTM.

With two separate hidden layers, the output is aggregated to produce a final output for predictions. The structure of a BiLSTM neural network is shown Figure 4.5. However, BiLSTM is relatively a new and advanced predictions model in the field of wind speed forecasting, hence its competitiveness has not been comprehensively demonstrated. In [182], performance of an LSTM was compared to a BiLSTM, with different epoch and unit values, the results showed that BiLSTM has an advantage to perform better due to the additional future information passed to the network for predicting. In [183], an ensemble of BiLSTM models as base predictors, generated predictions more accurate than that from other ensembles of deep neural networks. Additionally, in [184], BiLSTM was used to extract features from a time series, then extracted features are fed into another BiLSTM to obtain wind speed predictions. The results from the proposed model showed 39% improvements in terms of the coverage width compared with traditional models. Finally, in [185] different BiLSTM models are utilized to predict wind speed time series in different sub-clusters produced by k-means clustering, a machine learning approach which groups similar data points into k numbered clusters and makes predictions based on information from information from all points in the cluster.

4.2.3 Optimization algorithm (Grey Wolf Optimizer):

GWO is adopted to determine the optimal forecasting weights for combining the Complete Empirical Mode Decomposition With Adaptive Noise (CEEMDAN) decomposed signals. GWO is an optimization approach that aids deciding weights for the forecasting models. The algorithm is based on a swarm intelligence-based computation technique, where the leave-one-out strategy was developed to integrate the individual models [186].

Meta-heuristic optimization algorithms, which are swarm intelligence-based methods for simulating the hunting behaviour of grey wolves in nature, have been investigated extensively recently. On one hand, four types of wolves, alpha (α), beta (β), delta (δ), and omega (ω) are hired to imitate leadership hierarchy. The nature followed is that omega wolves follow the optimization process which is guided by the other three wolves. On the other hand, four main steps followed by grey wolves during hunting, namely, encircling prey, hunting, attacking prey, and searching for prey are key steps implemented during the process. The algorithm is categorised as follows:

Social hierarchy:

To mathematically simulate this process, the first three wolves are assigned to the first three best solution as a, b, c, respectively, while the remaining solutions - wolves - are assigned to as w. due to the nature of GWO, solutions a, b, c are employed to guide the optimization process and w will follow.

Encircling prey:

The encircling behaviour of wolves can be expressed as:

$$\overrightarrow{D} = |\overrightarrow{C} \cdot \overrightarrow{X_p}(t) - \overrightarrow{X}(t)|$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X_p}(t) - \overrightarrow{A} \cdot \overrightarrow{D}$$

Where t represents the current iteration, $\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors, $\overrightarrow{X_p}(t)$ is the position vector of the prey, and $\overrightarrow{X}(t)$ represents the position vector of a grey wolf vector and are calculated as follows:

$$\overrightarrow{A} = 2\overrightarrow{e} \cdot \overrightarrow{r_1} - \overrightarrow{e}$$

$$\overrightarrow{C} = 2\overrightarrow{r_2}$$

Where $\overrightarrow{e}$ components are linearly decreased from 2 to 0 over the course of iterations and r_1 and r_2 are random vectors in [0,1].

Hunting:

Grey wolves have the capability to identify the position and location of the prey and then encircle it. To mathematically simulate this process of hunting the prey, the previous statement that wolves a, b, and c are the optimum solutions and will have a better understanding of the positions and locations of prey is followed. Subsequently, the fittest three solutions obtained are assigned and the remaining search agents, including w wolves, update their positions accordingly. This is mathematically represented as follows:

$$\overrightarrow{D_a} = |\overrightarrow{C_1} \cdot \overrightarrow{X_a} - \overrightarrow{X}|, \overrightarrow{D_b} = |\overrightarrow{C_2} \cdot \overrightarrow{X_b} - \overrightarrow{X}|, \overrightarrow{D_c} = |\overrightarrow{C_3} \cdot \overrightarrow{X_c} - \overrightarrow{X}|$$

$$\overrightarrow{X_1} = \overrightarrow{X_a} - \overrightarrow{A_1}(\overrightarrow{D_a}), \overrightarrow{X_2} = \overrightarrow{X_b} - \overrightarrow{A_2}(\overrightarrow{D_b}), \overrightarrow{X_3} = \overrightarrow{X_c} - \overrightarrow{A_3}(\overrightarrow{D_c})$$

$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1}+\overrightarrow{X_2}+\overrightarrow{X_3}}{3}$

Where $\overrightarrow{A_1}$, $\overrightarrow{A_1}$, $\overrightarrow{A_1}$ are random vectors, and $\overrightarrow{X_a}$, $\overrightarrow{X_b}$ and $\overrightarrow{X_c}$ represent the positions of and wolves, respectively.

Attacking prey:

Following, the wolves attack when the movement of the prey stops. To approach the prey in the mathematical simulation, the value of $\overrightarrow{e}$ is reduced and the fluctuation range of $\overrightarrow{A}$ is also decreased with $\overrightarrow{A}$. That is decreasing $\overrightarrow{e}$ from 2 to 0, the random value of $\overrightarrow{A}$ is changed in the interval [-e, e] during the iterations. Furthermore, the next position of the search agent could be anywhere between the current position and the position of the prey, given the value of is in [-1, 1].

4.2.4 Hybrid model: CEEMDAN + EWT + BiLSTM + GWO

The proposed hybrid CEEMDAN-EWT-BiLSTM-GWO model, which is based on the hybrid decomposition algorithm and the BiLSTM-GWO model, is demonstrated in Figure 4.6, the main steps of the model are based on the hybrid decomposition approach and bidirectional LSTM optimized using GWO. In the first step, the CEEMDAN is used to decompose the original wind speed time series into a set of IMF each with a different frequency. The first IMF generated, IMF1, is the most fluctuating and disorderly signal amongst all outputs and hence is not suitable for forecasting and would reduce the prediction accuracy. Therefore, another signal decomposition iteration aimed at reducing the fluctuations and noise in that signal is essential to reduce the forecast difficulty. The EWT can decompose

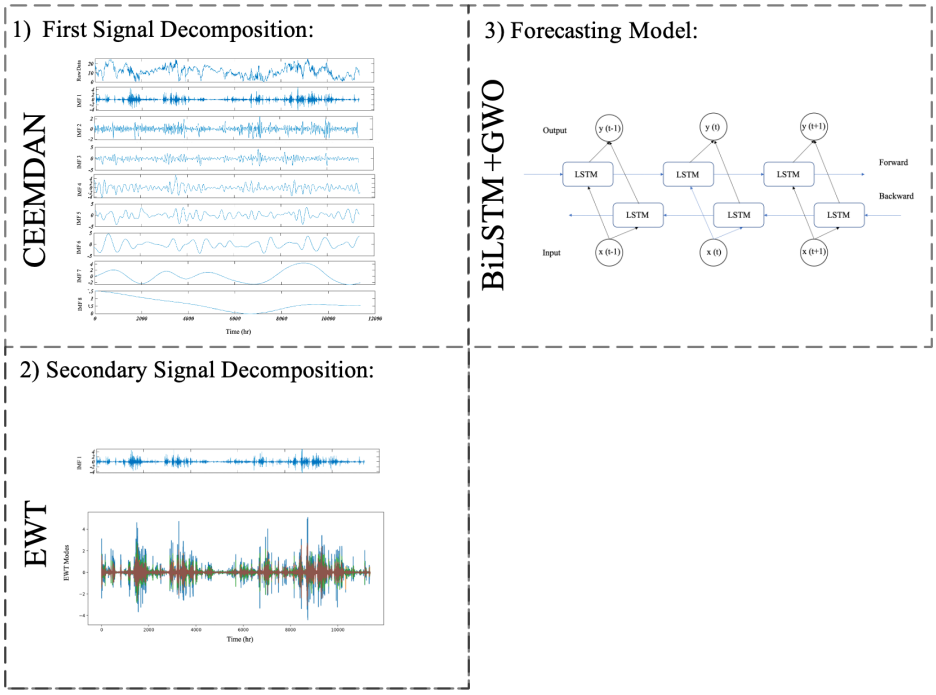

Figure 4.6: Architecture of the proposed hybrid model.

high frequency signals such as IMF1 into relatively more steady components. Hence, EWT is employed to further decompose the wind speed signal before feeding signals into the neural networks. In the following step, predictions on the decomposed signals are generated, including the modes generated by EWT and IMFs generated by CEEMDAN, excluding IMF1, using BiLSTMs optimised by GWO. In the last step, all forecasting results from modes and IMFs are aggregated to obtain the final forecasting time series.

4.3 Experiments and Analysis

To assess the effectiveness of the proposed combined forecasting model, experiments using wind speed datasets collected from the Technical University of Denmark (DTU) are used as illustrative examples.

4.3.1 Data set

To test the proposed model, a case associated with the RUNE project is considered, which is a near-shore experiment conducted at the west coast of Denmark. Here, a numerical experiment is used, which is part of a number of numerical simulations performed using the WRF model v3.6. This particular experiment was setup with 4 nested domains, the outermost covering northwestern Europe and a 2-km horizontal resolution innermost domain covering the west coast of Denmark [168]. Spectral nudging to the ERA5 reanalysis was used in the upper model levels of the outermost domain. The simulation has 8 vertical levels within the first 100 m and instantaneous output was produced every 10 min. The experiment also used the Mellor-Yamada-Janjic planetary boundary layer scheme, a sea surface temperature product from the Danish Meteorological Institute, and the CORINE land cover description [169]. For this experiment, the same 4 WRF simulation points across 2 km horizontal line, 50 m above sea level are used, which are used in chapter 3, to have a fair comparison with the model that used data fusion and lidar observations.

4.3.2 Evaluation metrics

Many evaluation metrics are researched and applied to evaluate the effectiveness of different forecasting models. In many studies, no general standard for performance metrics is followed. The most common error metrics include mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). However, after analysing the results from multiple studies, it is noticed that some models outperform other models when the evaluation metric is changed. For example, model (a) can show better performance than model (b) in an iteration, while performing much

worse in a different iteration. This could be explained by referring to the approach of a model, as every iteration tries new parameters and functions to generate predictions. However, this should not significantly affect the metric used. Alternately, after investigating the equations for each metric and comparing the results from different studies, the abnormality is found to be associated with RMSE and MSE. This is because as models are more advanced, errors have reduced significantly to less than 1, i.e 0.9, 0.6 and 0.4. In the equations for RMSE and MSE, the error is squared before it is aggregated to find the mean error, which means that the error is massively decreased for iterations where performance is high (error less than 1), which is a property of both RMSE and MSE, as they penalise both high and low errors more than other metrics. Hence, an iteration which generates forecasts with an error less than 1, when squared, the error will be smaller compared to using other metrics. Contrarily, when the iteration generates forecasts with errors greater than 1, when squared, the error will increase, thus, final mean error is significantly affected.

Table 4.2: Table of different metrics and their equations to calculate the forecasting error.

Metric	Definition	Equation
MAE	Mean absolute error of D forecasting results	$\frac{1}{D}\sum_{i=1}^{D}\lvert x_i - y_i\rvert$
MSE	Mean squared error of D forecasting results	$\frac{1}{D}\sum_{i=1}^{D}(x_i - y_i)^2$
RMSE	Root mse of D forecasting results	$\sqrt{\frac{1}{D}\sum_{i=1}^{D}(x_i - y_i)^2}$

MAE, MSE, and RMSE are used to evaluate the average magnitude between the predicted result and the original data, to avoid different signs forecasting errors cancelling each other out. For these metrics, the model with lowest values is considered of the highest performance. The equations and definitions for each of the three metrics are given in Table 4.2, where x and y represent the actual and predicted values, respectively, and D is the sample size.

4.4 Results and Discussion

This work proposed a hybrid forecasting model to predict data from a WRF simulation dataset for 4 points. First, Intrinsic Modal Functions (IMF)s from pre-processing algorithm Complete Empirical Mode Decomposition With Adaptive Noise (CEEMDAN) are presented. Then, presented modes of IMF 1 after second decomposition using Empirical Wavelet Transform (EWT) to reduce the high frequency in the signal. Following, predicted signals are aggregated and compared to the original dataset. Second, predicted signal is compared to lidar observations and predicted signal from the model in Chapter 3 to investigate and assess the benefits of using data fusion of lidar observations with WRF simulations.

4.4.1 Pre-processing using CEEMDAN and EWT

To build a forecasting model with high performance, it is critical to fully analyze and consider the features of the original time series. In this work, the CEEMDAN approach is first used to decompose the original wind speed time series to reduce the non stationary and non linear characteristics. As presented in Figure 4.7, the raw data in addition to the IMFs extracted from the original wind speed data is shown from highest to lowest frequency. Results show that each decomposition contains its own characteristics, reflecting the different oscillatory nature in the time series provided. For this experiment, 8 total IMF components are generated, named IMF 1 to IMF 8, with IMF 1 representing the decomposition with the highest frequency, additionally, IMF 1 has the most detailed information of the original time series. Contrarily, IMF 8 represents the decomposition with the lowest frequency, which presents the variational trend of the wind speed series.

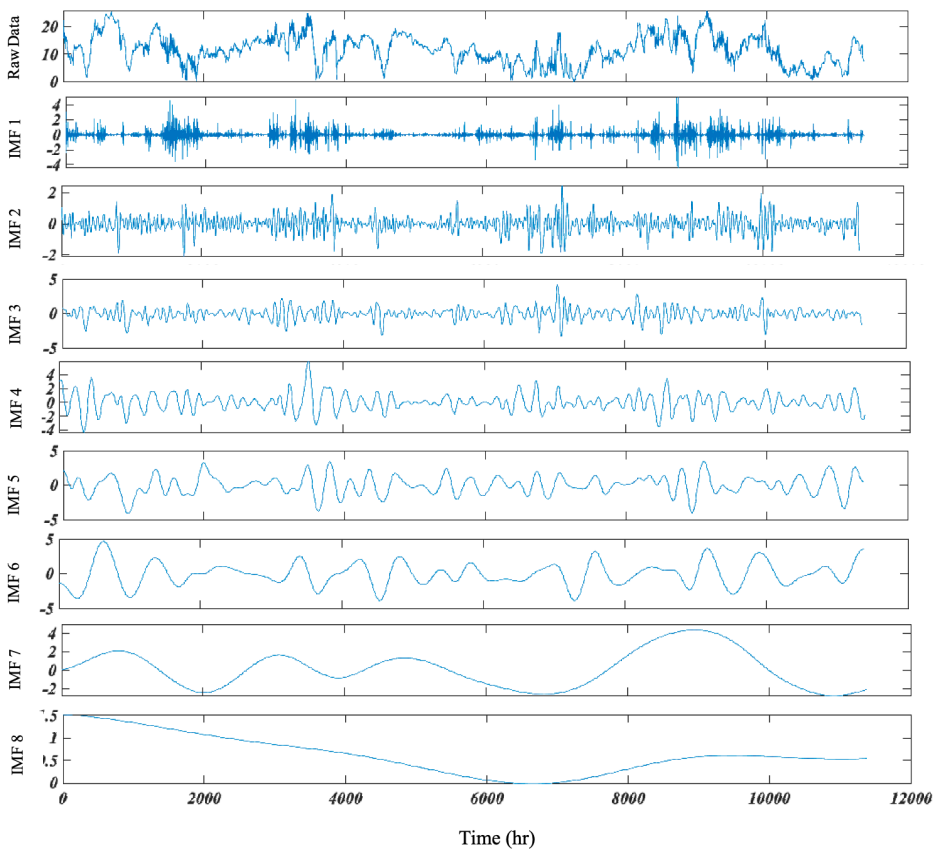

Figure 4.7: Decomposed Intrinsic Modal Functions (IMF) sub-series from original signal using CEEMDAN.

Forecasting results of the IMFs show good performance, except for IMF 1 due to its high frequency oscillatory nature. Therefore, IMF 1 is concluded to require additional pre-processing due to its poor predictions performance in addition to its importance.

Thus, to improve the status of IMF 1, EWT is employed as a secondary signal decomposition approach to further decompose that signal. The result is 6 decomposed modes of IMF 1 (shown in Figure 4.8) and all IMFs and modes, except IMF 1 are then fed into the BiLSTM model for predictions.

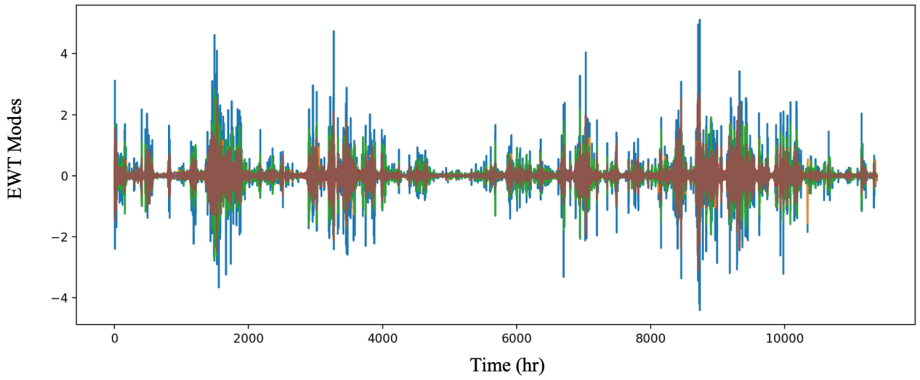

Figure 4.8: Secondary signal decomposition result of IMF 1 using EWT.

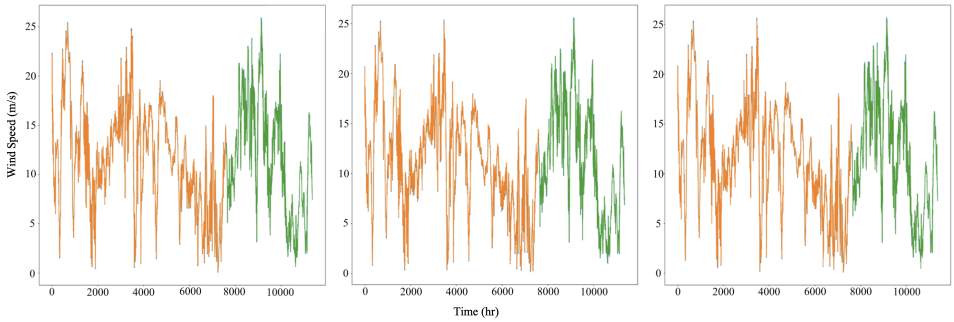

Figure 4.9: Prediction result for all three data points, where the orange line represents the training points and green represents the test points.

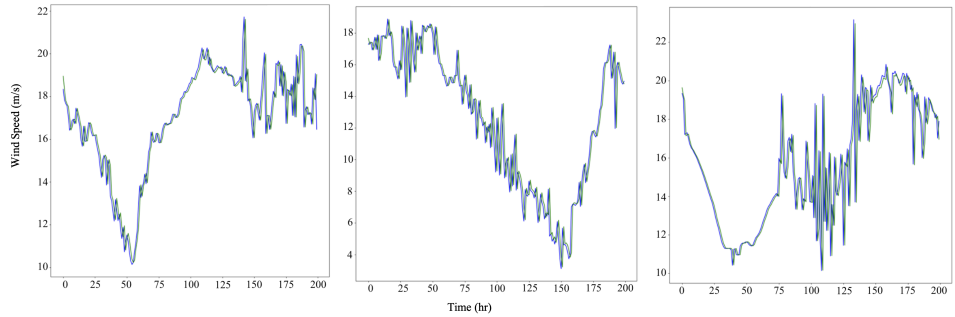

Figure 4.10: A close up look of the test points for all three data points.

4.4.2 Predictions using BiLSTM optimized using GWO

Following the pre-processing of the three furthest offshore points from the WRF simulation using CEEMDAN and EWT, data is fed into Bidirectional Long Short Term Memory (BiLSTM) neural networks optimised using Grey Wolf optimizer (GWO) for signal forecasting. The performance of the model is optimised, through the hyperparameters, by adjusting the weights,

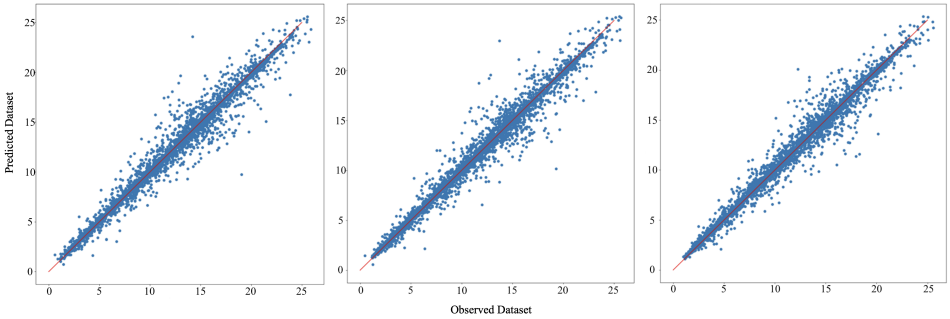

Figure 4.11: Plots of observed data against predicted data from the proposed model for all three data points.

bias, and number of layers for the neural networks. First, the performance of the model is explored by hiding 40% of the data, the remaining 60% are used in training the neural networks. Figure 4.9 and Figure 4.10 show the prediction results of training and testing the three signals and a close up on the generated results. The performance of predictions is evaluated using 3 evaluation metrics, MAE, MSE, and RMSE. The values of the evaluation metrics for the proposed model can be found in Table 4.3 and are 0.4, 0.56, and 0.75 m/s for the training section of the data and 0.48, 0.79, and 0.88 m/s, respectively. Figure 4.11 shows the observed vs predicted curve for three data points.

4.4.3 Proposed model compared to lidar observations and data fusion model

The experiment that is performed earlier revealed that building hybrid models that employ more than one pre-processing algorithms, combined with advanced and optimized deep neural networks, is sufficient at forecasting signals of wind speed time series. However, in wind resource assessment, the objective is not only to predict wind speed data accurately, but to predict data as close as possible to the actual data that is later measured. For the majority of experiments, models are trained and tested on simulation

Table 4.3: Configurations and accuracy of the Gaussian Process Regression models.

WRF Point	Metric	Value
Point 1	MAE (Training)	0.4
	MAE (Test)	0.48
	MSE (Training)	0.56
	MSE (Test)	0.79
	RMSE (Training)	0.75
	RMSE (Test)	0.88
Point 2	MAE (Training)	0.41
	MAE (Test)	0.49
	MSE (Training)	0.55
	MSE (Test)	0.76
	RMSE (Training)	0.74
	RMSE (Test)	0.87
Point 3	MAE (Training)	0.38
	MAE (Test)	0.51
	MSE (Training)	0.59
	MSE (Test)	0.83
	RMSE (Training)	0.77
	RMSE (Test)	0.91

data, since they are relatively easier and cheaper to obtain. Contrarily, simulation data is considered of relatively low-fidelity compared to other wind speed measurement techniques such as lidars, yet delivers continuous streaming data across large spaces for long time periods, which is not possible with lidars. A possible solution to the scarcity of lidar data and low-fidelity of simulations is proposed in Chapter 3, where lidar measurements are merged with WRF simulations in a hybrid solution for offshore wind resource assessment using a data fusion approach. For this chapter of the thesis, the significant importance of data fusion and having other sources of wind speed data that are of high-fidelity are demonstrated, by comparing results from a model trained with only simulation data, and a model trained with both lidar measurements and simulations (Chapter 3) of the lidar measurements as they are considered closer to the real values.

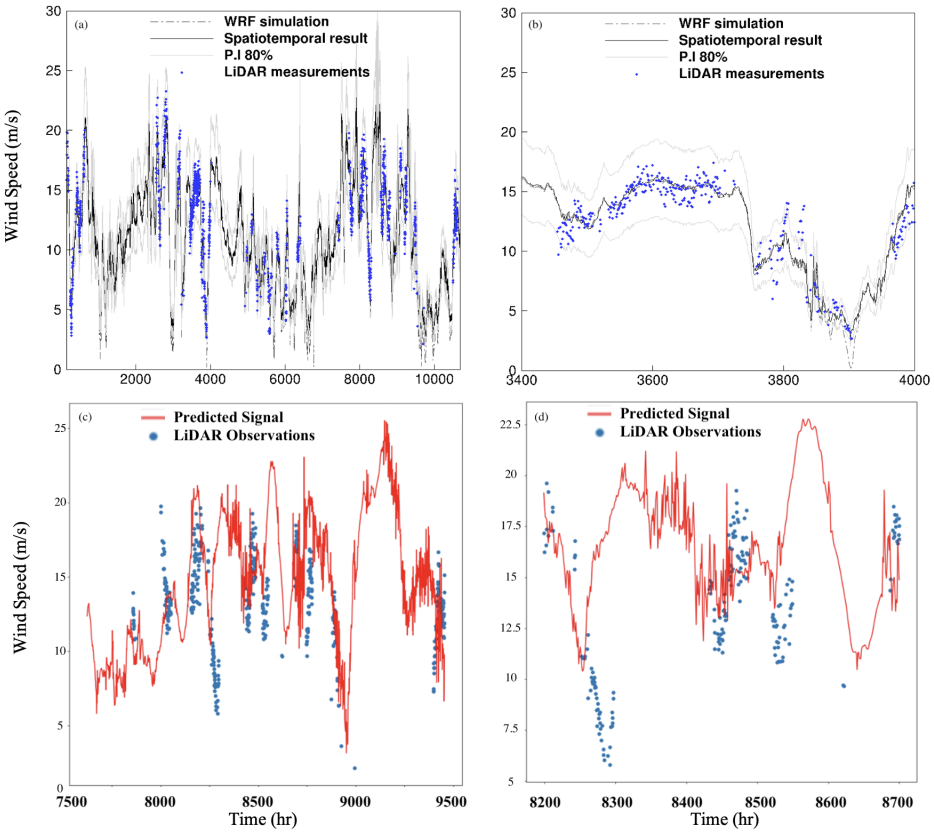

Figure 4.12: Panels (a) and (b) showing predicted signal and close up of the test data using proposed model (CEEMDAN+EWT)+ BiLSTM + GWO compared to lidar measurements. Panels (c) and (d) showing predicted signal and close up of the test data using proposed data fusion model from chapter 3 against lidar measurements.

Figure 4.12 panels (a) and (b) show results from the CEEMDAN + EWT + BiLSTM + GWO model proposed in this chapter compared to lidar measurements. Results showed that despite the model performing effectively in predicting the WRF simulation time series, performance is poor compared to the high-fidelity lidar measurements, where the average RMSE is 1.36 m/s. Alternately, the model trained with lidar measurements showed significant out-performance, where the RMSE is 0.63 m/s, with 53% improvement in performance. In panels (c) and (d) a presentation of the results from the model proposed in chapter 3, which uses lidar measurements as additional input information. Results show that the model following data fusion and using lidar measurements as secondary information outperforms

models that do not consider lidar data.

4.5 Conclusion

This work performed wind speed data forecasting using a novel hybrid approach. The proposed model decomposed the wind speed time series using two pre-processing approaches to reduce high frequency and decrease non-stationarity and non-linearity in the signal. The obtained IMFs and modes from CEEMDAN and EWT, respectively, are fed into BiLSTM neural networks optimized by a GWO algorithm to predict the final signal of the wind speed time series. Results from the model show high performance and accuracy in mimicking the simulated time series (WRF), however, the simulation data are not a good representation of the actual wind speed observed, as they are considered of poor quality compared to high-fidelity data (lidar).

In the field of wind resource assessment, and specifically when predicting wind speed data, the task is to get better forecasts relative to the high-fidelity data, as the main objective of the process to estimate the power generated from the wind turbine or farm. Minor differences in the wind speed would have significant impact on the estimation of wind power, since the formula for calculating the output power is the cube of the wind speed. Thus, a complete dependence on simulations may risk calculating power values that are far off real values. Contrarily, data fusion of lidar measurements and simulations could provide a solution that combines the merits of both techniques for the assessment of wind resources.

For the following chapter, a tensor (three dimensional matrix) is employed to ingest wind speed data in two space and one time dimensions to develop

a three dimensional space-space-time model that can capture both spatial and temporal features of wind speed and provide predictions for a full 3D domain. The model is a multivariate algorithm that will have both WRF simulations and lidar measurements of the wind speed. Additionally, it will aim to reduce the number of missing entries in the matrix by finding lower rank product matrices of the original matrix, then Gaussian process regression is employed to impute the missing data. Finally, the product of the complete lower rank matrices is calculated to obtain a full original matrix with no missing entries. The proposed forecasting model will allow for data fusion of different sources of wind speed data to generate predictions in multiple dimensions, taking into consideration several neighbouring and distant lidar and WRF data points.

Chapter 5

3D Probabilistic Matrix Factorization

In the pre-construction of wind farms, wind resource assessment is of paramount importance. Measurements by lidars are considered a source of high-fidelity data. However, they are expensive and provide an incomplete scarce time-series in both space and time domains, in particular, high unavailability for measurements offshore. Contrarily, Weather Research and Forecasting (WRF) models, with numerical simulations, generate continuous temporal and spatial data with a relatively low fidelity. A hybrid approach to combine the merit of measurements and numerical simulations for the assessment of offshore wind is proposed. Unlike other models, which consider either spatial or temporal extrapolations for point or interval predictions, this model investigates assessments of large spatial areas at three different heights with multiple spatial points for the duration of three months, providing direct plane and site level assessments. Firstly, the measurements and numerical datasets are fed onto a sparse matrix, where the columns represent the spatial lidar and WRF points, and the rows represent the

time steps. Entries of the matrix reflect the wind speed at a given time and specified location and technique, empty entries reflect unobserved data. Then, a non-linear probabilistic matrix factorization using Gaussian process model is used, to train and test for matrix completion, which fills the missing data with predictions. The model is optimised with Stochastic Gradient Descent, to apply Gaussian process without approximate methods. Additionally, SGD performs well on large sparse matrices, as it scales linearly with the number of observations. To evaluate the method, numerical and measured wind speed data along the west coast of Denmark are used. The proposed data fusion technique, using gappy measurements, resulted in accurate offshore wind resource assessment with matrix completion results of higher accuracy than industrial and academic models, with 58% and 40% improvements, respectively. The experiment in this chapter is able to reduce the computational cost of forecasting a 3 dimensional space-space-time wind speed location by reducing the number of iterations required to predict multiple wind speed time-series and the total number of wind speed points to predict. Additionally, in the lower rank matrices, the forecasting model is able to correlate both the spatial and temporal latent features to a single domain, using both features simultaneously to predict the wind speed.

5.1 Introduction and Literature Review

Over the past decade, the energy industry has seen great changes due to a worldwide demand for sustainable energy. Clean energy is recognised as the pathway for a sustainable future, which lead to a dramatic expansion and an increase in renewable energy capacity, and a rise in global investments. In 2019, the global wind power market added 60 GW to its arsenal, the

second largest wind power annual increase, reaching a total of 743 GW for both onshore and offshore sites [187]. In 2021, Denmark wind energy accounted for an estimated 57% of electricity generation, with high shares also in both Ireland and the UK, 32% and 24.8%, respectively. As of 2020, the UK has a total set onshore record of 10.2 TWh and offshore of 9.2 TWh [188]. In addition, Europe intends to increase the demand for wind energy and its capacity by 35% within the next decade [189].

Evaluating the wind speed condition of a potential location is a critical early step before the construction of any wind farm. As minimal changes in speed can drastically have large deviations in the power output [190], and as the wind varies both geographically and temporally over a wide range of scales, an accurate wind resource assessment is essential and is considered of a paramount significance for a successful wind energy project [191]. Moreover, the assessment provides aid to the selection of wind turbines, their layouts, and for planning a wind project, which wind power developers use to estimate the future energy production of a wind farm to meet their demand [192].

Instruments that measure wind can yield accurate observations of the wind speed but are expensive, and the data is generally sparse in both the space and time domains. This equipment includes e.g. lidars, which measure the line-of-sight (LOS) velocity by computing the Doppler shift of the signal of an infrared laser based on the movement of aerosols. However the lidar output is usually intermittent with high unavailability at offshore locations. Also, satellite synthetic aperture radars (SARs) are employed, which measure wind at 10 m above sea level (ASL) with low temporal resolution and apply only to offshore measurements. There are also buoy systems, which are equipped with an array of sensors that measure a number of parameters with regards to the state of the atmosphere and the water, but they are ex-

pensive and require regular maintenance. Finally, there are meteorological masts, which are the traditional way to measure the wind climatology.

Contrarily, numerical weather prediction (NWP) models offer output that cover large geographical areas and long-time horizons simultaneously and continuously, but the data is significantly of a lower fidelity. The models include e.g. Weather Research Forecast (WRF), Global Forecast System (GFS), and European Centre for a Medium-range Weather Forecasts (ECMWF) [193].

Measurement instruments and numerical simulations complement each other, which suggests hybrid data fusion techniques to combine their merits. It is desirable to extend the information from coastal vertical lidars (wind profilers) for the reconstruction of offshore time series, as they are easier to maintain [194]. Information can be numerically extended from coastal measurements to offshore time series at lower cost and higher accuracy, compared to having complete data dependent on lidars at all the positions. This technique has been widely used in predictions of future developments based on various inputs [195]. Further, wind resource assessment is commonly requested to cover large areas, over a long-time-interval (e.g. a few months or years), therefore, recommending spatial-temporal fusion of physically and numerically measured wind [196].

Missing value estimation is a significant problem in many research areas, including recommender systems [197], geostatistics [198], and image restoration [199]. In most cases, the cost of acquiring high fidelity data or repeating an experiment, due to low availability, is high, therefore, filling out missing data is the method of preference [200]. Most missing value estimation approaches include but are not limited to e.g. clustering algorithms and probabilistic matrix factorization (PMF). Clustering based approaches

became so popular after performing very well in the recommender system, which analyse patterns of user interest in products to provide personalised recommendations by employing a smaller set of highly similar users instead of the entire database [201-202]. However, matrix factorization techniques proved to be superior to clustering-based methods because the former allows incorporation of additional information [203]. In its basic form, PMF factorizes a matrix to find two lower rank matrices such that their dot product is the original matrix, which minimizes the sum squared difference of matrix values. The model focuses on data-dense locations in the matrix and hence provides lower rank matrices with much lower number of missing cells. After factorizing the partially observed matrix, each row as well as each column are assigned a latent vector, and the estimation of the missing cell becomes the inner product of the latent vectors for the corresponding row and the corresponding column. PMF characterizes both time steps and spatial points by vectors of factors inferred from point time series patterns. High correspondence between space and time factors leads to estimation. These methods are becoming increasingly popular by combining good scalability with predictive accuracy. Different types of input data are placed in a matrix with one dimension representing space and the other dimension representing time [204].

In this work, a non-linear probabilistic matrix factorization model is employed to map both space and time to a joint latent factor space of dimensionality F, such that space-time interactions are modelled as inner products in that space. For large and sparse matrices, stochastic gradient descent technique is used to optimize the Gaussian process, which successfully handles large-scale and sparse machine learning problems, and the parameters can all be learned using maximum likelihood via the technique [202]. For a dataset with N spatial points and M time steps, the

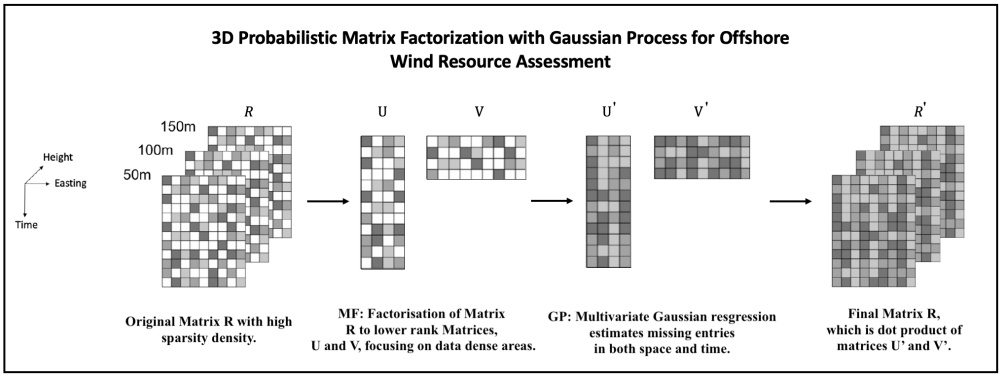

Figure 5.1: Flow chart for 3D spatio-temporal probabilistic matrix factorization for wind resource assessment.

matrix is considered as $R \in N \times M$. The objective is to obtain a lower rank factorized form of R, $R = U^T V$, where $U \in R^{D \times N}$ and $V \in R^{D \times M}$. The algorithm focuses on locations in matrix R with dense data to reduce the number of missing cells in matrices U and V. Predictions can then be performed on missing entries by estimating (U, V) from the training data and computing the resulting approximation to R. This is graphically demonstrated in Figure 5.1. The results from this model provide direct predictions of wind speed across multiple spatial points located at three different planes, representing three different heights, each representing a different height, and at any given time. Previous forecasting models focused mainly on predicting the wind speed either spatially, by including multiple spatial points for a single time-step, which is also known as point predictions, besides, interval predictions; where the target is to forecast the wind speed for a specific spatial point for a specified time interval, dealing with a single wind speed time-series. The multi-fidelity Gaussian model [209] deals with multivariate data to merge several time-series and make predictions for the time series of a single spatial point, and then data is fed to a neural network to predict the time-series for a different spatial point.

The novelty of this work relies on the utilisation and deployment of a probabilistic matrix factorization model with Gaussian process algorithm

for the accurate assessment of offshore wind resources with reduced cost and higher accuracy, testing multiple points both spatially and temporally. It combines the generally continuous but low-fidelity numerical data and high-fidelity but limited physical measurements. Efforts are also devoted to pre-processing the time series and taking into account additional information not considered in existing methods to lift the accuracy of the fusion. This algorithm enables the projection of limited nearshore measurements to offshore locations in light of numerical simulations and limited lidar measurements with significantly higher accuracy than the industry standard approach. The application uses tests for the prediction of wind speed in 2 space dimensions across time using multiple lidar, WRF, and processed data inputs, to deliver a model capable of performing spatial and temporal predictions in a single step for multiple spatial points at different heights and locations, feeding 64 lidar points and 20 WRF spatial points along 12,960 time steps. The algorithm considers all neighboring data in the space and time domains for the prediction of every missing entry.

This chapter is organized as follows. In Section 2, related and previous work is discussed, where algorithms of missing value estimation are employed. In Section 3, a demonstration of the methodology for the temporal extrapolation using multi-fidelity GPR, probabilistic matrix factorization, stochastic gradient descent optimizer, and proposed application model, non-linear matrix factorization with Gaussian process algorithms is discussed. Section 4 describes the work case location and the collection of the high and low fidelity data along with their temporal and spatial distributions. In Section 5, main results of both experiments conducted are shown and the performance of our model is compared against the industrial and academic standards, before drawing conclusions in Section 6.

5.2 Related Work

In this section, a discussion of the previous methods used to estimate the parameters of missing data is demonstrated. In addition, machine learning models and some of their applications in other sectors for data imputation are stated.

Based on available data, Expectation Maximization (EM) and Maximum Likelihood (ML) are two very common methods of estimating parameters of missing data. ML approaches the missing value estimation by finding the underlying probability distribution of the available data. Due to the sparseness of the dataset, it is imperative that both methods follow an iterative way in estimating the missing values. The first step of the approach is to estimate the parameters of interest from the available data and the probable value of the missing data. Following, parameters are recalculated using the available data along with estimates from the first round, and new parameters are applied to re-estimate the missing values, and so forth. This process is repeated until the estimated data has a high correlation with that of the previous cycle [201]. The aforementioned methods are more appropriate for uni-modal probability distributions of sample, which is not the case in data collected from a wind site, and especially for wind speed. In the case of stochastic intermittent data, a mixture of Gaussian methods is a more suitable probability distribution method.

In machine learning, factorization-based methods are a well-established and powerful technique for analysing data for matrix completion. A significant amount of research was conducted to achieve careful data preprocessing, hyperparameters tuning, handling very large, sparse, and imbalanced datasets. A probabilistic framework for matrix factorization, PMF, was presented in [206], which was integrated to a fully Bayesian model

later [196]. The model scales linearly with the number of observations in the original matrix and performed well on the Netflix dataset, where the rows represented users, the columns represented the items and the data is review-based. The model included adaptive prior on the model parameters and showed how the model capacity can be controlled automatically. Further, MF was generalised to a full Bayesian model in [207] and [208], which incorporated multiple sources of side information and combined multiple priori estimates for the missing data using real-world drug-target interaction datasets. Additionally, Agathokleous and Tsapatsoulis [201] inspected the Voting Advice Application (VAA) data from the Cypriot presidential elections to estimate missing data for the party and candidate recommender system using several collaborative filtering methods. This work applies the idea of missing data completion in a matrix, in the wind industry sector.

5.3 Methods for Probabilistic Matrix Factorization

In the methodology section, pre-processing of signals using empirical wavelet transform (EWT), is covered in Section 5.3.1, which reduces fluctuations in time-series, by removing unnecessary noise. Subsequently, in section 5.3.2, non-probabilistic matrix factorisation with Gaussian process is covered. A general overview of probabilistic matrix factorization is introduced, which focuses on imputation of missing values, optimization with stochastic gradient descent, and combining with the Gaussian approach to have a non-linear approach are discussed. This section is based on implementing a Gaussian probability for estimation of missing data in a numerical method that allows imputation of data in matrices and tensors. The targeted algo-

rithm allows for wind speed estimations in 3 dimensions, two of which are in the space domain relative to a third time dimension.

5.3.1 Pre-processing: Empirical Wavelet Transform

Empirical Wavelet Transform (EWT) is an algorithm used to achieve good forecasting results for non-stationary wind speed time series. The five-model decomposition algorithm can extract meaningful information from a given series by designing an appropriate wavelet filter bank. Previous work is followed, in which a dataset derived from model simulations is pre-processed by generating the adaptive wavelet and then decomposing the signal into a finite number of modes. The process starts by identifying and extracting the different intrinsic modes of the wind time-series, by relying on robust preprocessing for peak detection. Then, spectrum segmentation is performed based on the detected maxima, hence constructing a corresponding wavelet filter bank. In this work, the EWT algorithm is used to preprocess 15 grid points, 5 at three different heights, from a WRF-based numerical simulation for the comparison model EWT + MF-GPR. The process consists of five main steps: extending the signal, Fourier transforms, extracting boundaries, building a filter bank, and extracting the sub band. The five level decomposition attained by the preprocessing algorithm, EWT, is able to describe the signal in a meaningful way with much less fluctuations, by extracting five uncorrelated filter modes from the wind speed signal and a residual from the extraction. The reconstructed signal will be used as additional input for the MF-GPR.

5.3.2 Non-linear Probabilistic Matrix Factorization with Gaussian Process

The aforementioned Gaussian process is an excellent method for temporal data fusion and 1D time-series predictions. However, in geostatistics and specifically in wind resource assessments, it is necessary to predict multiple points at different locations, and hence spatial extrapolation is required along temporal extrapolation. In previous studies [210], neural networks such as NAR and NARX were trained to connect the time-series of two spatial points to train data with observed time-steps using WRF simulations and predict offshore time steps at unobserved ones using lidar measurements of the onshore point. However, this work experimented with 16 spatial points distributed across 4 km range along a three month time-series at three different heights.

Probabilistic Matrix Factorization PMF

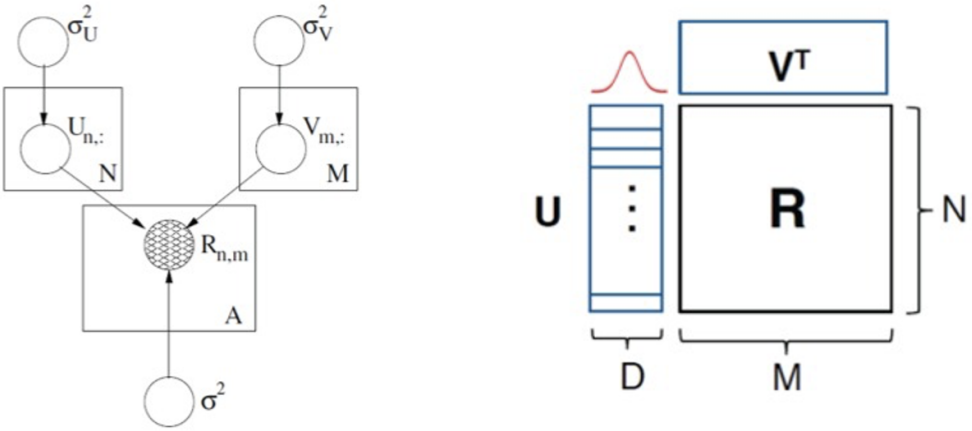

Figure 5.2: Architecture of probabilistic matrix factorization with uncertainties (Left) and byproduct matrices (Right).

In this work, non-linear matrix factorization with Gaussian process is employed to predict and assess wind speed for 16 different spatial points,

concurrently, where the learning is done following Stochastic Gradient Descent. An $N \times M$ real valued matrix with several missing cells is considered, where N is the number of time-steps and M is the number of lidar points. The goal of the matrix completion is to predict and fill these missing cells based on the available data. Two latent matrices are employed: $U \in R^{D \times N}$ and $V \in R^{D \times M}$, to capture the row and column features of the original matrix R, respectively. PMF favours a probabilistic perspective to solve the problem from the matrix factorization aspect. Let $R_{i,j}$ represent the wind speed of time-step i (time) for lidar point j (space), with column vectors U_i and V_j representing the time-specific and space-specific latent feature vectors respectively, the generative algorithmic development process is as follows [205]:

1. For each row i in R, $[i]_1^N$, generate $U_{i,:}$ $\mathcal{N}(0, \sigma_U^2 I)$, where I denotes the identity matrix.
2. For each column j in R,$[j]_1^M$, generate $V_{j,:}$ $\mathcal{N}(0, \sigma_V^2 I)$.
3. For the non-missing cells (i, j), generate $R_{i,j}$ $\mathcal{N}(U_{i,:} V_{j,:}^T, \sigma_2)$.

A conditional distribution is defined over the observed wind speeds as:

$$p(R|U, V, \sigma^2) = \prod_i^N \prod_j^M [\mathcal{N}(R_{ij}|U_i^T V_j, \sigma^2)]^{I_{ij}}, \tag{5.1}$$

where $\mathcal{N}(x|\mu, \sigma^2)$ represents the probability density function of the Gaussian distribution with mean μ, variance σ^2, and I_{ij} is the indicator function that can either be equal to 1 when the matrix cell has a wind speed value or 0 otherwise. In PMF, matrix R is modelled as a low rank matrix with noise corruption, where the matrix factorization, $U^\top V$, is the mean of the distribution and the noise is Gaussian with variance σ^2.Thus, zero mean spherical Gaussian priors are placed on time and space feature vectors [205]:

$$p(U|\sigma_U^2) \;=\; \prod_{i=1}^{N} \mathcal{N}(U_i|0,\sigma_U^2 I),\;\; p(V|\sigma_V^2) \;=\; \prod_{j=1}^{N} \mathcal{N}(V_j|0,\sigma_V^2 I). \quad (5.2)$$

The log-posterior distribution over the latent matrices U and V is given by [205]:

$$\begin{array}{l} \log\; p(U,V|R,\; \sigma^2,\sigma_U^2,\sigma_V^2) \\ =\; -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(R_{ij}-U_iV_j^T)^2 \\ -\frac{1}{2\sigma_U^2}\sum_{i=1}^{N} U_i^T U_i - \frac{1}{2\sigma_V^2}\sum_{j=1}^{M} V_j^T V_j \\ -\frac{1}{2}(A\log\sigma^2 + ND\log\;\sigma_U^2 + MD\log\;\sigma_V^2) + C \end{array} , \quad (5.3)$$

where constant C does not depend on the latent parameters. Ideally, the marginal likelihood of the model would be calculated, which is how this model learns, but in practice this is not tractable. Instead, maximum a posteriori, MAP, inference maximizes the logarithmic likelihood with respect to U and V, which is equivalent to minimizing the sum of squared error function with quadratic regularization terms [205]:

$$\begin{array}{l} E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}(R_{ij}-U_i^T V_j)^2 \\ +\frac{\lambda_U}{2}\sum_{i=1}^{N}\|U_i\|_{FRO}^2 + \frac{\lambda_V}{2}\sum_{j=1}^{M}\|V_i\|_{FRO}^2 \end{array} \quad (5.4)$$

where $\lambda_U = \alpha_U/\alpha$, $\lambda_V = \alpha_V/\alpha$, and $\|.\|_{FRO}^2$ denotes the Frobenius norm. Performing gradient descent in U and V will give a local minimum of the objective function.

For each column of the latent matrices, $U_{:,d}$ and $V_{:,d}$, the prior distribution is a zero-mean Gaussian process, which is a generalization of the multivariate Gaussian distribution. A Gaussian process $GP(m(t), k(t,s)$ is determined by a mean function $m(t)$ and a covariance function $k(t,t')$, contrarily, a multivariate Gaussian is determined by a mean vector and a covariance matrix, $GP(m(t), k((t,s),(t',s')))$, where the algorithm considers the high-

fidelity data, $f_h(t)$, in the multivariate set as a function of two variables (t, s), and s is the low-fidelity dataset, $f_1(t)$, $f_h(t) = g(t, f_l(t))$ [205].

A major drawback is the need for manual complexity control for this training procedure to allow the model to generalize well, specifically on very large sparse datasets. Controlling the regularization parameters mentioned above (λ_U and λ_V), is one way of controlling the complexity of the model, where the model can be trained by a set of parameter values one setting of the parameters at a time, and choose the best performing model. However, this is highly computational and hence very expensive, as it requires multiple training of models instead of just one. Alternately, priors could be introduced for the hyperparameters and maximizing the log-posterior of the model over both hyperparameters and parameters, which will lead to automatic control of the complexity based on the training data. Despite this approach proving beneficial in practice, it is theoretically not well grounded [205].

A demonstration of the architecture of a PMF is shown in Figure 7.1, where full matrix R is factorised into matrices U and V with uncertainties and dimensions $D \times N$ and $M \times D$, respectively.

With little changes to the notations, PMF is probabilistically equivalent to Bayesian PCA. The Bayesian treatment provides fully automatic complexity control as model parameters and hyperparameters are integrated. Considering a matrix of latent variables, $X \equiv U^\top \in R^{N \times D}$, and a mapping matrix that goes from the latent space to the space of observed data, $W \equiv V^\top \in R^{M \times D}$. Following the new notation, the probabilistic model can be written in the form [206]:

$$p(R|W, X, \sigma^2) = \prod_{I=1}^{N} \mathcal{N}(r_{i,:}|W_{X_i,:}, \sigma^2 I), \tag{5.5}$$

where $X_i, :$ is the ith column of U, and $r_{i,:}$ represents the column vector from the ith row of R containing wind speeds of the ith time-step for a point in space. The previous equation is a multi-output linear regression from a D dimensional feature matrix V to matrix targets R. Placing a prior over X gives the following, which can be marginalised later to give the marginal likelihood used in missing values imputation,

$$p(X) = \prod_{i=1}^{N} \prod_{j=1}^{D} \mathcal{N}(x_{i,j}|0, \alpha_x^{-1}), \tag{5.6}$$

Missing values imputation

The method discussed is a Gaussian matrix factorization model with a particular covariance structure, which means marginalizing is straightforward in finding the missing values. A Gaussian distribution is considered over the following parameters: a vector y with mean μ and covariance σ, in the form y $N(\mu, \sigma)$. An observed subset of y is represented by y_i, where i is an index for the observed values. When marginalizing the missing values, getting the Gaussian form y_i $N(\mu_i, \sigma_{i,i})$ where μ_i, and $\sigma_{i,i}$, represent the mean vector with the rows for the sum of σ columns associated with the unobserved elements of the removed y. Hence, for a sparse data matrix, the likelihood is given by [206]:

$$p(Y|W, \sigma^2, \alpha_x) \;=\; \prod_{i=1}^{N} N(y_{i,j_i}|0, \alpha_x^{-1} W_{ji,:} W_{ji,:}^T + \sigma^2 I) \tag{5.7}$$

Optimizing with respect to the parameters leads to α_x being part of W, which leaves the likelihood function associated with PPCA, and hence be-

comes intractable when marginalizing W. Instead the prior is taken over W,

$$p(W) = \prod_{I=1}^{M} \prod_{j=1}^{D} \mathcal{N}(W_{i,j}|0, \alpha_W^{-1}), \tag{5.8}$$

and the marginal likelihood is then in the form

$$p(Y|X, \sigma^2, \alpha_w) \; = \; \prod_{j=1}^{D} N(y_{i_j,j}|0, \alpha_w^{-1} X_{i_{j,:}} X_{i_{j,:}}^T + \sigma^2 I). \tag{5.9}$$

which is the marginal likelihood of a Bayesian linear regression model with multiple outputs. These equivalences imply that with marginalisation of either W or X, will eventually lead to optimizing the resulting marginal likelihood for the remaining matrix and model hyperparameters.

The selection of likelihood is based entirely on the one with fewer parameters. In this case there are less columns since they represent the space domain, rather than the rows that represent the time steps. As using EM for a large matrix would be highly computationally expensive, it makes sense to consider SGD, which converges much faster [206].

Stochastic Gradient Descent

Instead of maximizing the likelihood through an EM approach presents the wind speed for each lidar point one at a time, computing the gradient of the logarithmic likelihood for the lidar point. The gradients of that point are then used to update the parameters X, σ^2, and α_w; then a new lidar point is presented. If the objective for the previous approach is to maximize the likelihood, the negative logarithmic likelihood is minimised. For the j^{th} lidar point this is given by [206]:

$$E_j(X) \;=\; \frac{N_j}{2}\;\log\;|C_j| + \frac{1}{2}(y_{i_j,j}^T C_j^{-1} y_{i_j,j}) + \text{const.}, \tag{5.10}$$

where $C_j \;=\; \alpha_w^{-1} X_{i_{j,:}} X_{i_{j,:}}^T + \sigma^2 I$ and N_j is the number of time steps with measured wind speeds. Then compute the gradient with respect to X as

$$\frac{dE_j(X)}{dX_{i_{j,:}}} = -GX_{i_{j,:}}. \tag{5.11}$$

with $G = (C_j^{-1} y_{i_j,j}\, y_{i_i,j}^T,_j\, C_j^{-1} - C_j^{-1})$. Other gradients with respect to the other parameters can also be found.

The model Non-Linear PMF via GP-LVMs

A probabilistic matrix factorization with parameters marginalized is considered a Bayesian multi-output regression model, where optimization occurs with respect to the inputs to the regression [16], which is equivalent to probabilistic PCA. In addition, it also belongs to a category of models called Gaussian process latent variable models (GP-LVM). The Gaussian process with covariance function $C \;=\; \alpha_w^{-1} X X^T + \sigma^2 I$ is a linear model that can be transferred nonlinear by replacing the inner product matrix, XX^T by a Mercer kernel. Consequently, maximizing over the logarithmic likelihood can no longer be attained through an eigenvalue problem; however, SGD in the manner described above would be straightforward.

The regression model followed can be written in the form of a product of univariate Gaussian distributions,

$$p(R|U, V, \sigma^2) = \prod_{i=1}^{N} N(y_{i,:}|W_{x_{i,:}}, \sigma^2 I), \tag{5.12}$$

where the inner product $f_j(x_{i,:}) \;=\; w_{j,:}^T x_{i,:}$ represents the mean of each

Gaussian. A Gaussian process with latent variable models can be recovered by recognising the placed prior distribution directly over the function through a Gaussian process. A commonly used covariance function that gives a prior over nonlinear functions is known as the Radial Basis Function (RBF), covariance,

$$k(x_{l,:}, x_{i,:}) = \alpha_m \exp\left(-\frac{\gamma m}{2}||x_{l,:} - x_{i,:}||^2\right), \tag{5.13}$$

which can be substituted directly in the likelihood to give a probabilistic model,

$$p(R|U, \sigma^2, \theta) = \prod_{j=1}^{D} N(y_{i_j,j}|0, K + \sigma^2 I), \tag{5.14}$$

where parameters of the covariance function are presented in the θ hyper-parameter.

Predictions

Following learning based on the SGD provides an estimate of the latent matrices U and V, where for a missing cell, $R_{n.m}$, the maximum likelihood becomes the inner product of the corresponding latent vectors [206].

5.4 Case Description

In this work, probabilistic matrix factorization with Gaussian process is used for wind speed predictions at multiple points for continuous time. In particular, given a time-velocity matrix with missing entries, where the goal is to predict the missing wind speeds at the unobserved time-steps; by observing both high-fidelity sparse lidar measurements and continuous simulation output from the WRF model (see Figure 5.4 for time-series plot

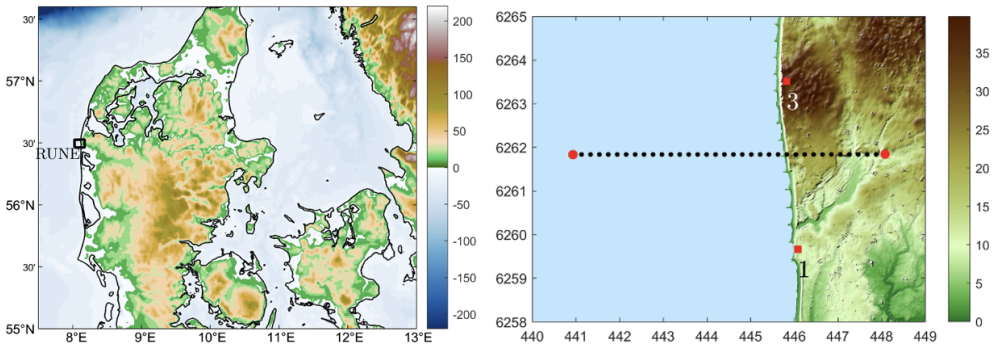

Figure 5.3: (Left) RUNE experimental area (in the rectangle) in western Denmark. (Right) RUNE coastal experimental area on a digital model of the surface (UTM32 WGS84, Zone 32V). The positions of the lidars (1 and 3) are shown in squares and of the dual-Doppler scans (36) in black markers. The scans are taken at three heights, 50, 100 and 150 m ASL. Colorbars indicated the height in meters ASL [25-26]

of lidar measurements and WRF simulations, and Figures 5.5, and 5.6 for the distribution of the observations across the experiment period).

In addition, Figure 5.7 demonstrates the correlation between all lidar and WRF time series, the correlations range from 0.8 to 0.99, indicating high correlation between all the sets, which reflects that even far points can be used to influence the prediction of any point in the matrix, and the strong relation between all the data points. Strong correlation between the independent features and dependent variable is a strong indication that accurate estimations could be yielded. Two numerical experiments are run on the data acquired from the dual-Doppler scans of the RUNE experiment (see Figure 5.3). The scans were acquired between the period from December 2015 until March 2016. The dual-Doppler scans (36 in total per height measured) were performed with two scanning lidars (positions 1 and 3), which were configured to match their scanning patterns at three heights 50, 100, and 150 m ASL. One 'virtual line', i.e. a line perpendicular to the coast took about 45s.

The dual-Doppler points used are at the far most offshore 4 km range. In

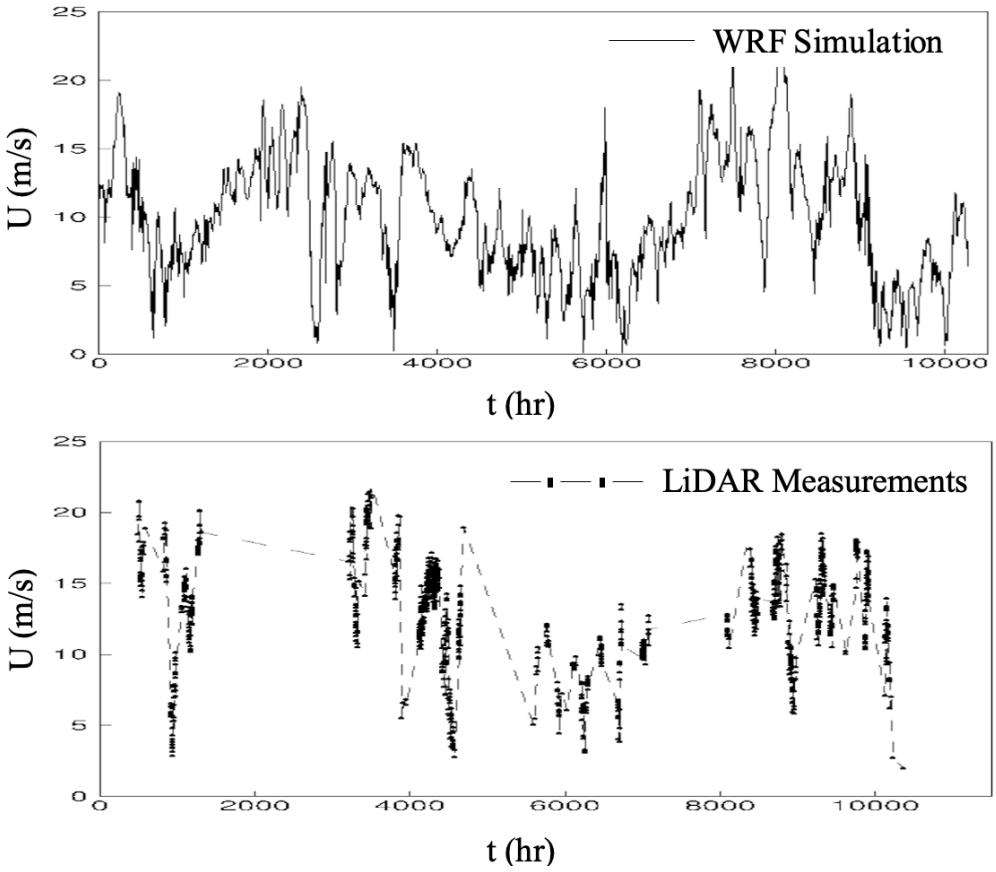

Figure 5.4: (a) Low-fidelity data from numerical simulation (WRF) at the most offshore point. (b) High-fidelity data from the dual-Doppler lidar setup at the most offshore point.

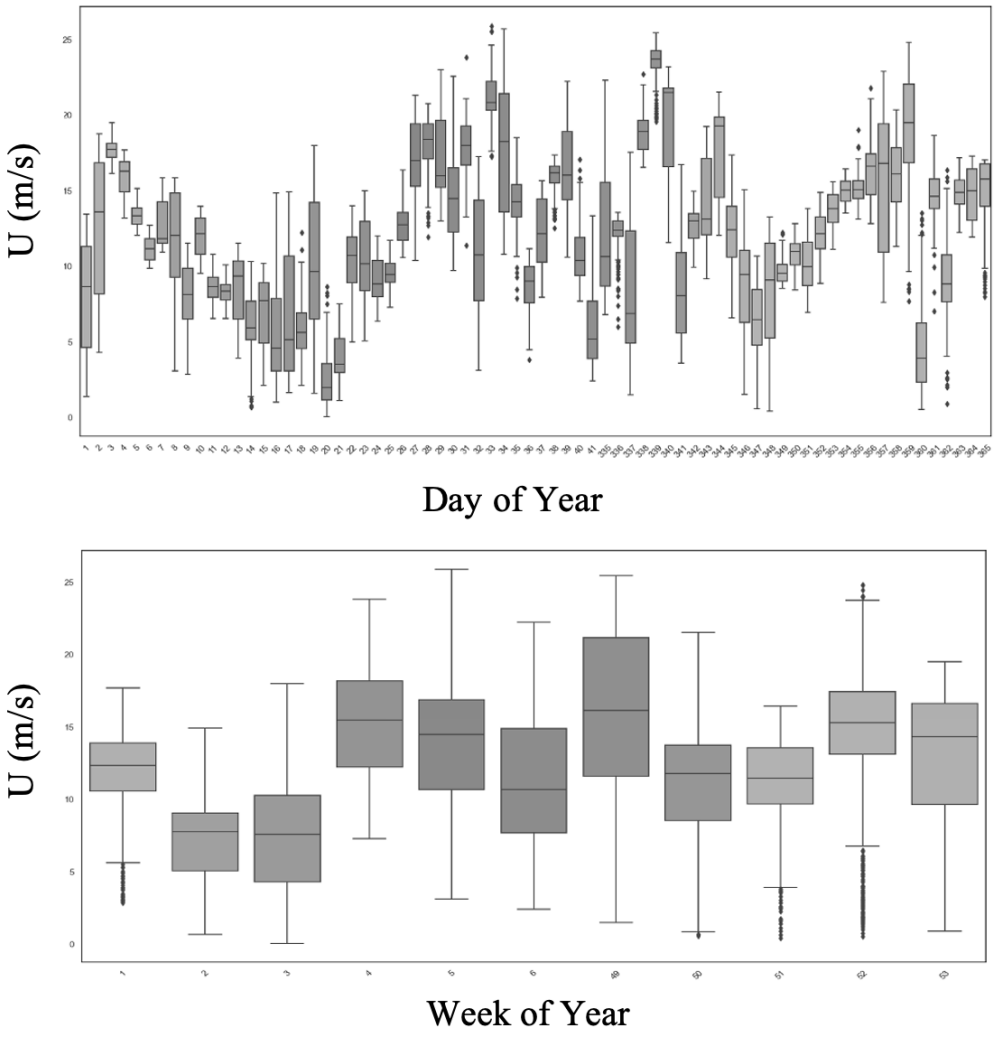

Figure 5.5: (a) Daily box plot for wind speed time-series.(b) Weekly box plot for wind speed time-series.

the first experiment, the effect of increasing the observations in the matrix and varying the lidar points by using three different matrices with different number of lidar points is evaluated. As for each height, there are about four to five dual-Doppler positions per km, which is referred to as lidar points. The 3 matrices available are: Matrix A has the first 2 lidar points (from the west) per km, which leads to a total of 8 new lidar points; Matrix B has the same lidar points as matrix A with an additional lidar point (the following point moving east), hence, 3 lidar points per km leading to a total of 12 new lidar points; and Matrix C has 4 lidar points per km leading to a total of 16 lidar points. The first experiment takes place at the 50 meters height level. The statistics of the datasets for experiment 1 are given in Table 1, which also shows the division of data for all three matrices A, B and C with percentage of sparsity in every matrix. Figure 5.6 (a), shows the number of observations for every week of the experiment and (b) shows the histogram for the number of observations. For all experiments,the data is partitioned, 20% of dataset for the validation, 20% for testing and 60% for training.

For experiment 2, matrix C (16 lidar points) alone is used, but at the three different heights, (50, 100, and 150 m). For experiment 2, results of the experiment are compared to those using EWT and GPR, an academic prediction model for wind forecast predictions, and Windgrapher, a leading industrial software. However, these techniques are one dimensional interval prediction techniques, hence can only be applied to one spatial point at a time and require 16 iterations for comparison with one iteration from the probabilistic matrix factorization with Gaussian process model.

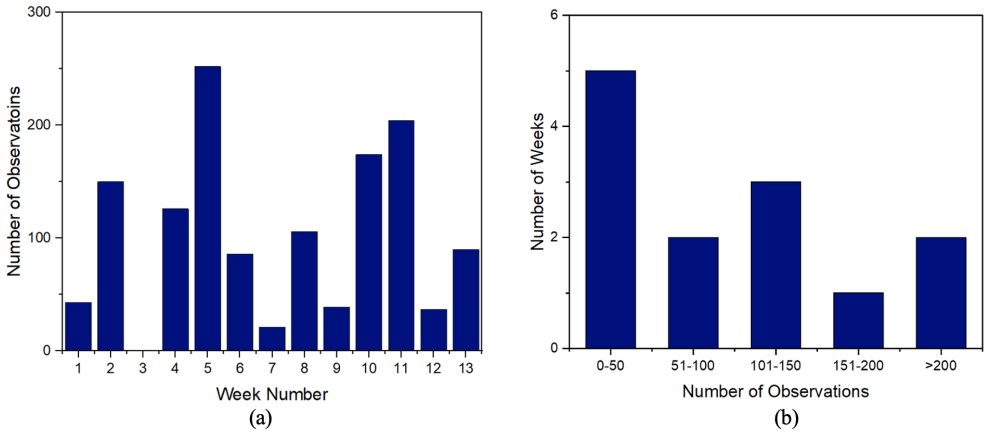

Figure 5.6: Histograms of (a) wind speed measurement frequencies; and, (b) the division of observations by weeks.

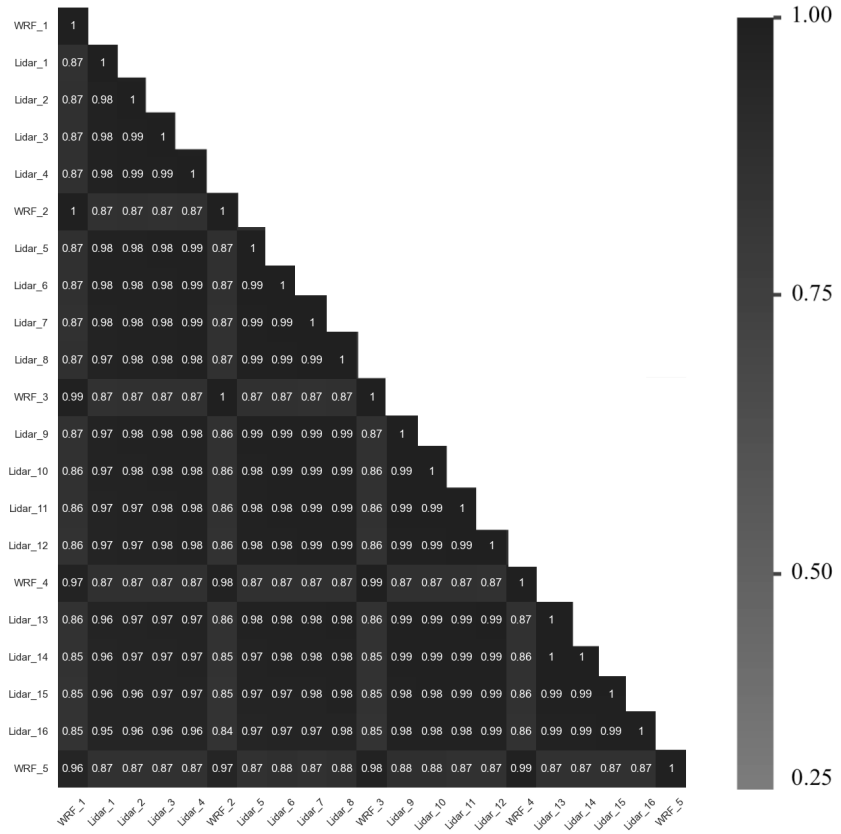

Figure 5.7: Correlation between all lidar and WRF points.

Table 5.1: Statistics of the datasets used in Experiment 1.

	A	B	C
lidar Points	8	12	16
Measurements observed (lidar)	10624	15936	21248
Simulation data (WRF)	56880	56880	56880
Total No. of Cells	739440	966960	1194480
Missing Data (Sparsity)	91%	93%	94%

Table 5.2: Statistics of the datasets used in Experiment 2.

	50 m	100 m	150 m
lidar Points	16	16	16
Time-steps	11376	11376	11376
Measurements observed	21248	21248	21248
Measurements per Point	1328	1328	1328
Measurements Density (Sparsity)	11.8%	11.8%	11.8%

5.5 Results and Discussion

Results from both experiment 1 and 2 are discussed and shown in this section. First, the results from experiment 1, where 3 matrices had different number of lidars to study the effect of adding lidar observations and empty cell to the matrix. Then for the second experiment, the results compare the RMSE for 16 lidar points at 3 different heights using the discussed non-linear Gaussian probabilistic matrix factorisation using two k factors (2 and 5), which are compared with an industrial predictions software, an academic method, and finally the WRF simulation as benchmark.

5.5.1 Experiment 1: One height and different number of lidar points

Three different datasets for the furthest west offshore distance of 4 km are used to train three iterations of probabilistic matrix factorization. The performance of the three models is optimized, through the hyperparameter,

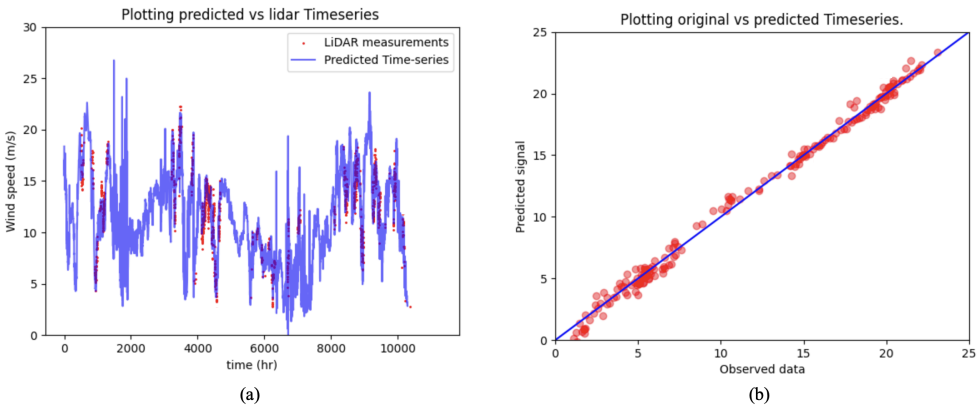

Figure 5.8: (a) Predicted Time-series and lidar observations.(b) Prediction points against counter observed measurements.

by applying different learning rates and K-factors. Firstly, exploring the matrix with 5 WRF points and 2 lidar points per km, with a total of 8 lidar points, the time steps represent 3 months worth of data. The dimensions of the matrix are 13 columns and 11,376 rows, represented as model (A). Then, for model (B), the number of lidar points is increased to 3 per km, and finally model (C) had 4 lidar points per km. The slight increase in the number of lidar points helped increase the sparsity of the models from 91% to 94%. This experiment is employed to test how the algorithm reacts to different sparsities (density of missing data) and how increasing the number of lidar data affects the predictions.

Table 5.1 shows the statistics of the datasets used in Experiment 1. All models had a constant number of WRF data (5 points), while the lidar data and hence the total number of cells in the matrix varied, resulting in three different sparsity percentages for models A, B, and C around 91%, 93% and 94%, respectively.

Figure 5.8 panel (a) shows the predicted time series of the first lidar point and the lidar observations, additionally, panel (b) shows the observed lidar points against their counterparts from the predicted time series for the 20% test dataset. Figure 5.9 top panels (a), (b), and (c) shows the time-series

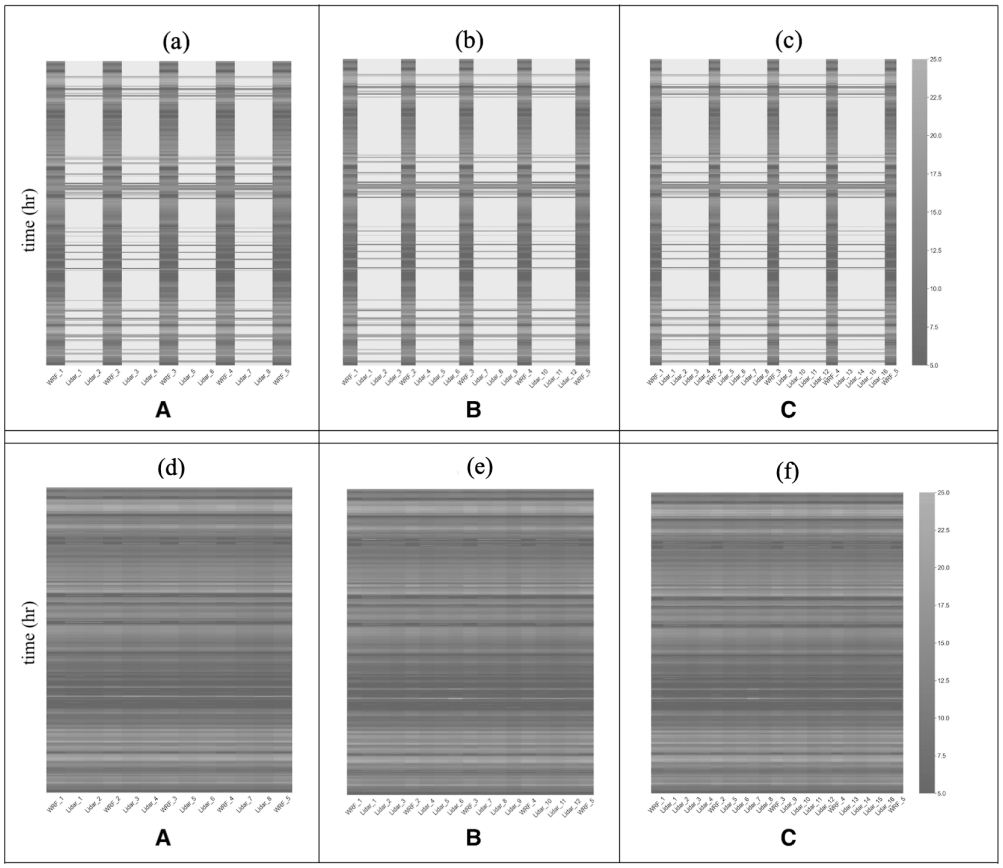

Figure 5.9: Original matrices of all 3 setups with 2, 3, and 4 lidar points, A, B, and C, respectively (Top). Final processed matrices for all setups, A, B, and C, respectively respectively (bottom).

of the three original matrices with 8, 12, and 16 lidar points, respectively. Panels (d), (e), and (f) show the 3 result matrices after performing PMF. The RMSE for each point in the matrix with respect to the 20% hidden test data within the time-series is measured and compared to the WRF output and lidar data. The results from all matrices outperform that of the WRF simulation. The experiment showed that increasing the number of lidar points, hence the sparsity of the matrix and amount of high accurate data, does not affect the accuracy of the prediction as the RMSE of the lidar points is unchanged amongst all points. This is due to the multicollinearity in the dataset, as all lidar and WRF points are highly correlated, it can be concluded that using any of the neighbouring points would be beneficial, despite its distance from the targeted point, and once points are to an extent linearly correlated, missing data is not a problem.

5.5.2 Experiment 2: 3 different heights for 16 lidar points

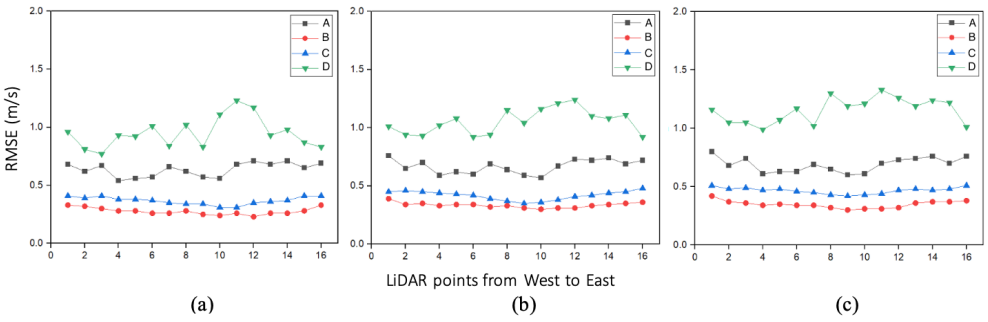

Figure 5.10: Comparison between RMSE results from Windgrapher (A), EWT + MGPR (B), our tested method NPMF with Gaussian process (k = 5 and k = 2), (c) and (D), respectively, across all 16 grid points for heights 50m (a), 100m (b), and 150m (c).

The second numerical experiment aimed to test how the models perform at different heights compared to other academic and industrial algorithms. Three models are trained with 5 WRF points, the number of lidar points

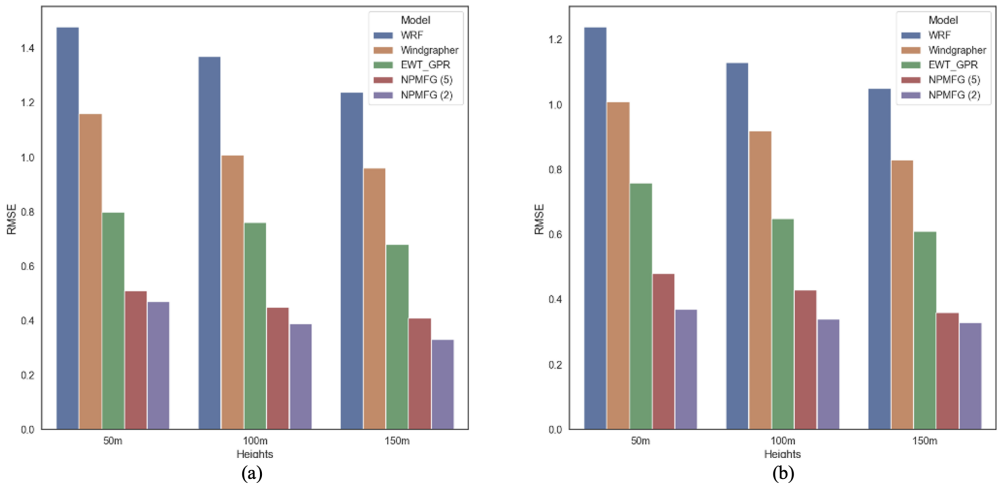

Figure 5.11: (a) Comparison between RMSE results from WRF data (Dark blue), Windgrapher (Orange), EWT + MGPR (Green), our current method NPMF with Gaussian process (k = 5 (Red) and k = 2 (Purple)) for the far east lidar point for each height. (b) Comparison between RMSE results from WRF data (Dark blue), Windgrapher (Light blue), EWT + MGPR (Green), NPMF with Gaussian process (k = 5 and k = 2) for the far west lidar point for each height.

is constant at 4 points per km leading to a total of 16 lidar points in each matrix. The three different heights of the experiment are 50, 100, and 150 m. The sparsity of all three models is unchanged at 11.8%, since this experiment focused on testing how the algorithm works at different heights, and on different sets of similar data compared to other prediction algorithms.

Table 5.2 shows the statistics used in Experiment 2. For this iteration, the total number of lidar points is fixed to 16, and had three different heights, 50, 100, and 150m ASL. All matrices have the same number of missing cells and lidar/WRF ratio, with a measurement density of 11.8%, 11,376 time-steps equivalent to 79 days, and 238,896 total number of cells per matrix.

Subsequently, the performance of the Matrix factorization algorithm is evaluated against an industrial software, Windgrapher, a leading software for

importing, visualising and analysing wind resource data. Windgrapher follows the Measure-Correlate-Predict (MCP) algorithms including Linear Least Squares; the method is on correlating target and reference speed data, based in the linear least squares procedure. Another algorithm in the comparison is the Empirical wavelet Transform (EWT) and Multi-Fidelity Gaussian Process Regression (MF-GPR). The EWT is used to pre-process the WRF time-series reducing the spikes and high frequency fluctuations, which results in a more accurate Gaussian process. Finally, for the PMF testing, different k numbers, 2 and 5 are used.

The RMSEs for each of the 16 lidar points are measured at one iteration from the Matrix using PMF with k= 2 and k= 5, then are compared to their counterparts using EWT + MF-GPR, and Windgrapher, but after 16 different iterations (as they can predict for a single point at a time only). Figure 5.10 shows the RMSE for each of the 16 lidar points using all algorithms tested, where panels (a), (b), and (c) represent heights 50, 100, and 150 m, respectively. On average, the PMF algorithm using k = 2 and k = 5 is able to outperform the other algorithms at all measured points for all three heights. The RMSE of PMF algorithm is reduced by at least 65% compared to the industrial software and 40% compared to the academic algorithm for 50 m ASL. At 100 and 150 m ASL, the RMSEs are very similar, however, increasing the height reduced the RMSE and increased the percentage drop in RMSE.

Figure 5.11 panels (a) and (b) show the RMSE of the most and least offshore points using all the algorithms followed for all three heights using the 3 algorithms discussed above and the 2 PMF models. The figure demonstrates the order of algorithms showing the least to most accurate, with PMF leading for both k numbers. Similarly, the results indicate that by increasing the height of the measurements more accurate predictions are

achieved, which is due to higher accuracy in the WRF data.

5.6 Conclusion

In this work, data fusion of lidar measurements sparse in space and time (high-fidelity) with output from the WRF model has been performed, which is continuous in space and time (low-fidelity), to obtain spatio-temporal predictions at unobserved space and time points, suitable for offshore wind resource assessment. During the RUNE experiment dual-Doppler scans are performed, which resulted in 36 lidar points across 10 km, both offshore and onshore, with 1331 measurements. Subsequently, the numerical simulations performed using the WRF model generated an instantaneous constant output every 10 min for the same period, resulting in a total of 11,376 data points. In this work, the performance of the algorithm has been tested on the offshore data at 16 points at 3 different heights, 50, 100, and 150 meters ASL.

In the first experiment, the number of lidar points per km is varied, to test the accuracy of the model with less valuable lidar data of sparsity percentages, which varied between 91%, 93%, and 94%. As aforementioned, the high-fidelity lidar data is presented at unobserved regions and periods by exploiting the available data and the low-fidelity WRF simulations. The addition of more lidar points caused an increase in the sparsity of the matrix, despite giving more valuable lidar data. The RMSE of predictions is not affected as it ranged from 0.45 m/s and 0.52 m/s across all three matrices; in this experiment, only the data at 50 m ASL is used.

Contrarily, the second experiment aimed to test the accuracy of prediction of the PMF model when the K-factor is varied at 2 and 5, and compare

the results to an academic model with pre-processing (EWT+ MF-GPR) and an industrial leading software (Windgrapher). First, 16 lidar points and 5 WRF points with 11.8% sparsity are used. The results showed that using a lower K-factor is more accurate and results in improved predictions; this is observed significantly at lower heights. Additionally, the results for all models showed that by using data at higher heights results in more accurate predictions, as the WRF outputs are more accurate at higher levels at this particular site. Hence, the results for 50 and 150 m had the highest and lowest RMSEs, respectively. The difference between both heights is around 0.18 m/s, equivalent to 15% drop in RMSE. Subsequently, both PMF models with 2 and 5 K-factors are able to outperform both the industrial and the academic models by at least 65% and 40%, respectively. Additionally, results from

There are three major limitations in this work, which could be addressed in future work. First, only data obtained from the RUNE experiment is addressed, which is a very sparse dataset, including only 1331 measurements for each lidar point equivalent to 220 h of measured data, reflecting several weeks with no data. Second, due to the low resolution of the WRF model output (2 km), the generated WRF points are all to the nearest 100 m, which is a considerable distance, as points this far will have significant variances in the wind speed. Hence, interpolation is necessary to obtain WRF data at the corresponding lidar points. Third, the starting time for both data sets generated by WRF and lidar is not constant, where the WRF data is available a few days earlier. This problem is called a 'cold' start and is a common issue in matrix completion problems.

Future work may also concern further development to the matrix input data. Additional datasets such as derivatives and WRF data of a higher resolution could be a great source of information. This would reduce the

sparsity of the matrix and improve the Gaussian process, hence the accuracy of predictions. Furthermore, preprocessing of the WRF data with EWT before processing in the matrix should be tested.

Finally, in this chapter, the experiments have been able to deliver the objectives of developing a 3 dimensional space-space-time hybrid model that combines wind speed data generated using different methods, namely lidar measurements and WRF simulations, to predict wind speed data at a reduced computational cost. The model is able to provide an accurate and continuous tensor of wind speed data than WRF simulations and lidars, respectively.

Chapter 6

Conclusions

This section summarises the work carried out during this thesis project and lists the findings from the experiments covered and how they supported solving and achieving the aim and objectives defined. Then, it covers the challenges encountered in building the wind speed prediction models. Finally, it provides a discussion on the possible future trends to enhance wind speed forecasting for wind resource assessment.

6.1 Summarising the models and reflection on objectives

The utilization of wind speed forecasting is a key solution to overcome the energy crisis and global pollution. It is very important to have an accurate estimation of the wind energy generated from wind turbines and integrated into the power grid. However, as the nature of wind is stochastic and intermittent, the stability of the supply is significantly affected, and accurate forecasting models are required. This thesis investigated various algorithms

and combinations of approaches that hired Deep Neural Network (DNN) to enhance the accuracy of forecasting models, and to use surrounding onshore measurements and offshore simulations to estimate a continuous stream of accurate offshore data without sending expensive equipment offshore. Additionally, the concept of using multi-variate data from different sources with different fidelities and propagation of forecasting results from one location to another is investigated. Multiple hybrid models are developed to perform data fusion of wind speed data and assess wind resources, which promises to reduce the cost of offshore wind resource assessment by using onshore data and limited lidar observations on onshore locations.

First, a model comprised of an Empirical Wavelet Transform (EWT) for pre-processing and a Nonlinear Autoregression With External Input (NARX) forecasting model is used to assess the merits of data fusion by combining limited lidar observations (high-fidelity) with continuously available Weather Research Forecast (WRF) simulations (low-fidelity) data. The model is able to successfully represent the high-fidelity data at unobserved regions and periods by exploiting the low-fidelity data and its functions taking additional side information such as wind speed magnitudes, pre-processed signals of the wind dataset, and derivatives of the low-fidelity dataset. The resulting time series mimics the lidar observations better than the WRF simulation with only onshore lidar observations, reducing the number of lidars required, fulfilling the objective of assessing the significance of data fusion of combining different sources of wind data and reducing of costs for offshore wind resource assessment.

Second, the effect of using data fusion and combining simulations with observations is investigated to further assess the significance of data fusion, and how using a high-fidelity data source improves forecasting performance. An advanced model is proposed, which used a second signal decomposition

approach to decompose the wind speed signal using two pre-processing approaches, namely Complete Empirical Mode Decomposition With Adaptive Noise (CEEMDAN) and EWT, and optimized the forecasting neural network using a Grey Wolf optimizer (GWO) to generate predictions based on WRF simulations only. Results showed that despite the -MF-GPR with NARX- model from chapter 3 lacking advanced pre-processing approaches and intelligent optimization algorithms, it is still able to outperform the proposed simulation hybrid model, which signified the value of obtaining lidar observations and using them as additional input data in the forecasting models.

Third, a novel model that considers three dimensional data based on matrix factorization is developed to predict wind speed data in two space dimensions with respect to time, to have predictions in the full 3D space-time domain, capturing both spatial and temporal features of wind speed at a wind farm location. The model uses EWT for pre-processing the signals and both lidar observations and WRF simulations are fed into the matrix to obtain spatio-temporal predictions at unobserved space and time domains simultaneously. Results from this model significantly outperformed the WRF simulations, results from the previous model hybrid models, and a leading industrial software, successfully delivering a forecasting model that does not require offshore lidars and can generate predictions in multiple space-time dimensions, hence reducing the cost of offshore wind resource assessment.

The three experiments carried out are able to establish a basis for understanding the benefits of data merge using advanced hybrid models that employ pre-processing algorithms, forecasting deep neural networks, and optimization algorithms. Additionally, during the training of each experiment, the hyper-tuning of the parameters showed major improvements

with respect to the accuracy of predictions, reflecting the importance of combining optimization algorithms to forecasting models.

In conclusion, it is important to address the computational cost aspect of the proposed models. By leveraging the developed models, the need for offshore deployment of lidars is minimized, resulting in significant cost reduction. Furthermore, the factorization of the matrix contributes to a decrease in missing data cells, further reducing computational requirements. Although a detailed cost analysis to quantify the monetary difference before and after implementing these models was not conducted, it is essential to acknowledge that the proposed models offer a promising solution to reducing costs associated with offshore wind speed predictions. By clearly articulating the impact on computational costs and emphasizing the avoidance of offshore equipment deployment, the thesis demonstrates the potential for significant economic benefits in wind energy forecasting.

Through this clarification, it becomes evident that the developed models not only enhance prediction accuracy but also provide tangible advantages in terms of cost efficiency, supporting the feasibility and practicality of their implementation in real-world scenarios.

6.2 Challenges and recommendations in wind speed forecasting

There are three main challenges that are encountered during the conducted work of this thesis.

(1) Understanding the complex uncertainties in a wind speed time series.

Due to the intermittent and stochastic nature of wind, the wind speed time series is considered a signal with a highly complex nature and uncertainties. The high non-stationarity and non-linearity in the wind speed signal reduces the performance of any prediction model used. Hence, it is very difficult for regression models to learn the relations and trends between input features and predicted output. Therefore, it is critical to employ a pre-processing stage to reduce the complexity of signals before they are fed into a prediction model. Many pre-processing approaches are employed to reduce the uncertainties, where the effectiveness of predictions are proven to improve.

Several pre-processing approaches were investigated and proposed in the literature, however, the degrees to which the forecasting accuracy is enhanced remains different for each approach. It has not been proven that a single particular approach can outperform others, which explains the variety in wind speed data, where each dataset contains its complexity and uncertainty. The variance is related to each dataset, and thus, different particular pre-processing models may be more suitable than others for each dataset. Therefore, it is recommended to employ hybrid pre-processing approaches such as CEEMDAN and EWT to benefit from the merit of model decomposition and wavelet transform in processing the dataset, which is proven beneficial in the third methodology chapter.

(2) Efficiently finding and extracting useful features in the dataset.

Another important factor in enhancing the performance of predictions is to accurately select input features of the forecasting model. A major challenge in producing accurate forecasts is the limited availability of data, it is essential to have datasets rich in data to avoid unsatisfactory wind speed forecasting results. An approach to counter the low availability of data

is to mine for more effective features in the dataset, which supports the model with explanations to the underlying fluctuations and improvements to the accuracy of training the model, hence increasing the accuracy and predictability of the regression model. Classic feature selection approaches are easy to implement at a low computational cost and provide a selection criterion for selecting important features from the available candidates. Contrarily, feature extraction algorithms often can not deal with highly complex features when the dataset contains data with highly non-linear characteristics. However, this effective drawback is less critical when the forecasting model employs a deep neural network, but on the other side, the computational cost is seen to increase. Another approach in dealing with low availability of data is to introduce additional information to the dataset such as first and second derivatives, and vector components of the wind speed time series, where more relations between the inputs and outputs can be observed and investigated. This technique is proven beneficial in the first methodology chapter.

(3) Automatically characterizing the complex relations between input features and output wind speed data.

In many recent works, DNNs were employed due to their powerful modelling of complex nonlinear relations, however, their performance is highly affected by the model configuration determined after several trials and errors based on future wind speed data. Therefore, it is highly computational and time consuming to assign DNNs with optimal configuration without knowing the future wind speed data. Therefore, Bidirectional Long Short Term Memory (BiLSTM) neural networks, which provide the forecasting model with future data as input, provide a quicker method for models to reach optimal configurations. Additionally, using optimization algorithms such as GWO is a good approach to design proper DNN-based models that

can best describe very complex relations in a dataset.

6.3 Future work in wind speed forecasting

To overcome the aforementioned described challenges, the following four approaches are proposed in the work, however, future improvements can be made to further improve the wind speed forecasting accuracy.

(1) Developing novel hybrid pre-processing approaches.

The uncertainties in wind speed datasets have not yet been fully analyzed, hence, more hybrid pre-processing models should be proposed and implemented to overcome this challenge. This thesis proposed using EWT in the first section of the methodology, which provided improvements to the forecasting model. Additionally, within the third methodology section, a hybrid pre-processing model that consider two different signal processing approaches are proposed, where CEEMDAN is employed to decompose the signal into different Intrinsic Modal Functions (IMF)s and the decomposed signal with highest frequency is further decomposed using EWT to different modes. There is an urgent need to continue developing novel models of pre-processing signals, to sufficiently analyze the uncertainties in wind data. New models consisting of different kinds of approaches will be suitable for dealing with the complex uncertainties.

(2) Increasing the number of input features.

In practice, fluctuations in wind speed data are caused by many factors. Hence, using topographical and meteorological factors, such as those obtained from WRF simulations, will be useful in precisely describing the fluctuations in wind speed. Therefore, using models that combine datasets

from WRF simulations and other sources of data will allow additional factors and features to be implemented in the algorithms. Additionally, finding sets that can provide additional side information for the provided datasets, such as derivatives and vector components, will improve the accuracy of the model.

(3) Developing novel hybrid forecasting models.

With the case with pre-processing of signals, different forecasting algorithms provide different approaches to tackling relations in a dataset. Therefore, new hybrid forecasting approaches that combine different deep neural networks are required to further enhance the accuracy of predictions. For example, traditional statistical forecasting models, such as Auto-regressive Integrated Moving Average (ARIMA), can be combined with new optimized and advanced neural networks to better characterize the fluctuations of a wind speed time series. This methodology will lead to an optimally combined model that targets forecasting nonlinear signals with different fitting abilities acquired from various forecasting models, to better describe very complex wind speed fluctuations. Additionally, optimizer algorithms make it easier to choose the optimal weights for combining different forecasting approaches, therefore optimizer algorithms are very critical in the forecasting process. Intelligent approaches offer optimal configurations of parameters to combine forecasting models, and hence more attention should be given to investigating the effect of various optimization approaches and how their design can be applied to combine various forecasting models.

(4) Testing the developed models on different datasets and variables.

This thesis focused on developing wind speed prediction models using a specific dataset and considering wind speed as the sole variable, it is important to acknowledge that further research and testing are required to

assess the robustness and generalisability of the results. One limitation of our study lies in the single geographical location used for testing, which may not adequately represent the diverse range of wind patterns and characteristics present in different regions. Additionally, by solely considering wind speed as the predictor variable, other influential factors such as temperature, humidity, and topography were not taken into account. Therefore, future work should aim to expand the scope of the study by including multiple geographic locations and incorporating a broader range of relevant variables. This will enable a more comprehensive evaluation of the model's performance and ensure its applicability across various settings. By addressing these limitations, we can enhance the robustness and reliability of wind speed prediction models for practical applications in renewable energy systems and other related fields.

Chapter 7

Appendix

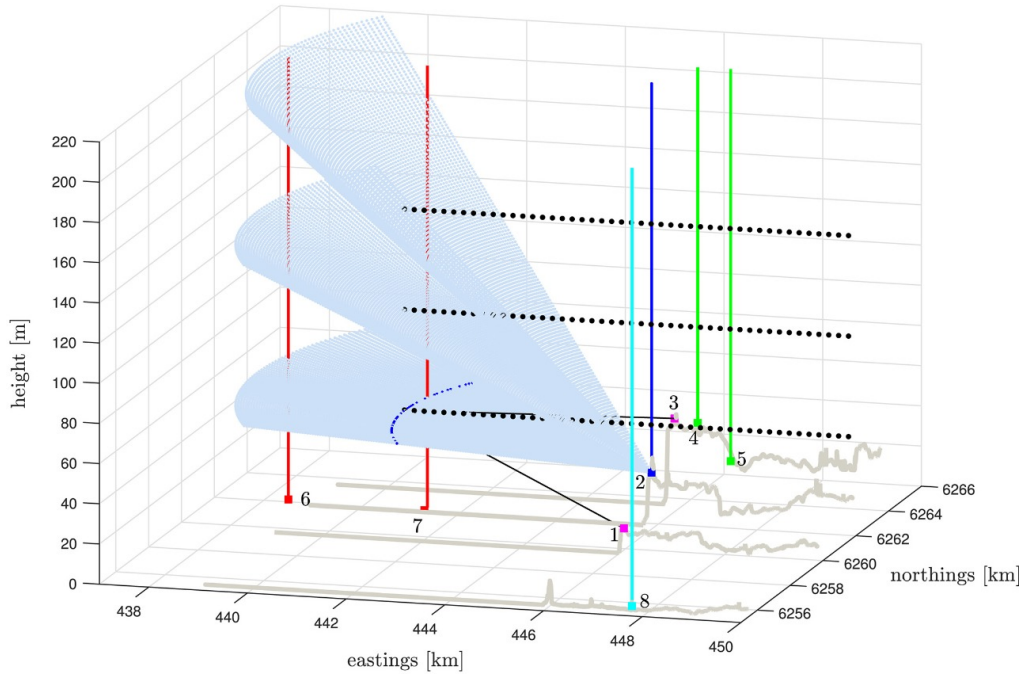

Figure 7.1: Map of LiDAR positions and technology used to generate the datasets.

7.1 Generation of LiDAR Data

7.1.1 Setup of LiDARs

Dual Setup for Koshava (pos. 1) and Sterenn (pos. 3) were configured to match their scans along three horizontal virtual lines at 50, 100 and 150 m ASL from 5 km west to 4 km east (inland) of position 2 (Figure 4). They acquired 45 LOSs per virtual line in 45 s, corresponding to horizontal distances of 200 m between points. Like the PPI scenario in phases 2 and 3, the total scanning time amounted to 145 s, including the 10 s that the scanner heads need to get back to their starting positions. LOS velocities were retrieved separately for each point and lidar and, from these, the horizontal wind speed components were reconstructed over a given period. To ensure the spatial proximity of two opposite range gates for the dual-setup reconstruction, 89 and 91 range gates were acquired per LOS and per system for phases 2 and 3, respectively.

In order to include all acceptable range-gate combinations, a collocating algorithm filtered out data that did not fulfil a certain distance threshold. During phase 2, the range-gate positions were not well collocated in the x-direction (i.e., west–east), and therefore, the threshold in the x-direction was set to 51 m. For the y (north–south) and z (vertical) directions and during phase 3, the threshold was 10 m. The reconstruction algorithm was applied to the 10, 30 and 60-min averaged LOS velocities of both systems, after filtering out data that did not fulfil the CNR threshold (26.5 dB). Reconstructed points situated near the coast using the dual setup cannot be well established. This is due to errors introduced by the very small beam crossing angles.

7.1.2 Setup of WRF

WRF (Weather Research and Forecasting) simulations of wind speed are generated using numerical weather prediction models. These models utilize mathematical equations and computational algorithms to simulate the behavior and evolution of atmospheric variables, including wind speed. Here's a general overview of the process:

Model Configuration: The WRF model requires initial conditions, boundary conditions, and various configuration settings. Initial conditions include atmospheric variables such as temperature, humidity, pressure, and wind speed at a specific starting time. Boundary conditions provide information at the model's outer boundaries to account for the influence of surrounding areas. Configuration settings specify model parameters, grid resolution, physics options, and other model-specific choices.

Grid Setup: The model domain is divided into a grid system consisting of multiple grid points. Each grid point represents a specific location in the atmosphere, and the resolution determines the spacing between grid points. The WRF model supports various grid types, such as Cartesian, latitude-longitude, and rotated pole grids.

Numerical Integration: The model's equations, including the fundamental equations of fluid dynamics and thermodynamics, are solved numerically to simulate atmospheric processes. These equations describe the conservation of mass, momentum, energy, and other relevant variables. The numerical integration process advances the atmospheric state in time, considering the interactions between neighboring grid points and accounting for physical processes like advection, diffusion, and parameterized physics.

Initialization and Time Stepping: The model uses the initial conditions to

set the starting state of the atmosphere. It then progresses through time in discrete steps, known as time steps. During each time step, the model calculates the changes in atmospheric variables based on the computed tendencies and the selected time step size. The model iteratively updates the state of the atmosphere by repeatedly applying the numerical integration process.

Parameterizations: To account for processes that occur at scales smaller than the model's grid resolution, parameterizations are used. These parameterizations represent sub-grid-scale phenomena such as turbulence, convection, radiation, and surface interactions. Parameterization schemes include the Mellor-Yamada-Janjic turbulence scheme, the Kain-Fritsch convection scheme, the Rapid Radiative Transfer Model (RRTM) for radiation, and others. These schemes provide estimates of wind speed changes due to these small-scale processes.

Output and Visualization: Once the simulation is complete, the model generates output files containing the simulated atmospheric variables, including wind speed. These output files can be analyzed and visualized using post-processing tools to understand the simulated wind patterns, variations, and their relationship to other meteorological variables.

Following are the equations used in the generation of the WRF simulation:

1. Conservation of Mass (Continuity equation):

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

This equation represents the conservation of mass, where ρ is the air density, t is time, $\mathbf{u}$ is the three-dimensional wind vector (u, v, w), and $\nabla\cdot$ denotes the divergence operator.

2. Conservation of Momentum (Navier-Stokes equations):

$$\frac{\partial(\rho \mathbf{u})}{\partial t} + \nabla \cdot (\rho \mathbf{u}\mathbf{u}) = -\nabla p + \nabla \cdot \tau + \rho \mathbf{g} + \mathbf{F}$$
$$\frac{\partial(\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v}\mathbf{u}) = -\nabla p + \nabla \cdot \tau + \rho \mathbf{g} + \mathbf{F}$$
$$\frac{\partial(\rho \mathbf{w})}{\partial t} + \nabla \cdot (\rho \mathbf{w}\mathbf{u}) = -\nabla p + \nabla \cdot \tau + \rho \mathbf{g} + \mathbf{F}$$

These equations represent the conservation of momentum in the x, y, and z directions, respectively. p is the pressure, τ is the stress tensor, $\mathbf{g}$ is the acceleration due to gravity, and $\mathbf{F}$ represents external forces (such as friction and Coriolis force).

3. Thermodynamic Energy Equation:

$$\frac{\partial(\rho \theta)}{\partial t} + \nabla \cdot (\rho \theta \mathbf{u}) = -\nabla \cdot (\rho \mathbf{v} \theta') + \nabla \cdot (K \nabla \theta) + Q$$

This equation describes the conservation of potential temperature (θ), which is a measure of the temperature normalized by pressure. θ' represents perturbations from the mean potential temperature, K is the thermal diffusivity, and Q represents heat sources/sinks.

4. Radiative Transfer Equation:

$$\frac{\partial I}{\partial t} + \mathbf{u} \cdot \nabla I = \varepsilon B - \kappa I$$

This equation represents the radiative transfer equation, where I is the radiation intensity, $\mathbf{u}$ is the wind vector, ε is the emission coefficient, B is the blackbody radiation, and κ is the absorption coefficient. This equation governs the transfer of energy due to radiation in the atmosphere.

5. Turbulence Parameterization Equations: Various sub-grid-scale parameterizations are used to represent turbulent processes. Examples include the

Reynolds-Averaged Navier-Stokes (RANS) equations, the TKE (Turbulent Kinetic Energy) equation, and the vertical mixing schemes

Bibliography

[1] Chen, H., Zhu, Q., Peng, C., Wu, N., Wang, Y., Fang, X., Gao, Y., Zhu, D., Yang, G., Tian, J., Kang, X., Piao, S., Ouyang, H., Xiang, W., Luo, Z., Jiang, H., Song, X., Zhang, Y., Yu, G., Zhao, X., Gong, P., Yao, T. and Wu, J. (2013), The impacts of climate change and human activities on biogeochemical cycles on the Qinghai-Tibetan Plateau. Glob Change Biol, 19: 2940-2955. https://doi.org/10.1111/gcb.12277.

[2] R.M. Errera, S. Yvon-Lewis, J.D. Kessler, L. Campbell, Responses of the dinoflagellate Karenia brevis to climate change: CO2 and sea surface temperatures, Harmful Algae, Volume 37, 2014, Pages 110-116, ISSN 1568-9883, https://doi.org/10.1016/j.hal.2014.05.012.

[3] T Tokioka, Climate changes predicted by climate models for the increase of greenhouse gases, Progress in Nuclear Energy, Volume 29, Supplement, 1995, Pages 151-158, ISSN 0149-1970, https://doi.org/10.1016/0149-1970(95)00038-L.

[4] Gail Rajgor, Building wind farms: Part five: the precarious construction phase needs careful preparation, Renewable Energy Focus, Volume 12, Issue 6, 2011, Pages 28-32, ISSN 1755-0084, https://doi.org/10.1016/S1755-0084(11)70150-8.

[5] C. Hiroux, M. Saguan, Large-scale wind power in European electricity markets: Time for revisiting support schemes and market designs?, Energy

Policy, Volume 38, Issue 7, 2010, Pages 3135-3145, ISSN 0301-4215, https://doi.org/10.1016/j.enpol.2009.07.030.

[6] Jian Chen, Yu Zhang, Zhongyun Xu, Chun Li, Flow characteristics analysis and power comparison for two novel types of vertically staggered wind farms, Energy, Volume 263, Part E, 2023, 126141, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2022.126141.

[7] Lars Ødegaard Bentsen, Narada Dilp Warakagoda, Roy Stenbro, Paal Engelstad, Spatio-temporal wind speed forecasting using graph networks and novel Transformer architectures, Applied Energy, Volume 333, 2023, 120565, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2022.120565.

[8] Yixiang Ma, Lean Yu, Guoxing Zhang, Short-term wind power forecasting with an intermittency-trait-driven methodology, Renewable Energy, Volume 198, 2022, Pages 872-883, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2022.08.079.

[9] Yangyang Li, Tao Zhang, Xintao Deng, Biao Liu, Jugang Ma, Fuyuan Yang, Minggao Ouyang, Active pressure and flow rate control of alkaline water electrolyzer based on wind power prediction and 100% energy utilization in off-grid wind-hydrogen coupling system, Applied Energy, Volume 328, 2022, 120172, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2022.120172.

[10] Sadjad Galvani, Behnam Mohammadi-Ivatloo, Morteza Nazari-Heris, Saeed Rezaeian-Marjani, Optimal allocation of static synchronous series compensator (SSSC) in wind-integrated power system considering predictability, Electric Power Systems Research, Volume 191, 2021, 106871, ISSN 0378-7796, https://doi.org/10.1016/j.epsr.2020.106871.

[11] Ahmad K. ALAhmad, Voltage regulation and power loss mitigation by optimal allocation of energy storage systems in distribution systems considering wind power uncertainty, Journal of Energy Storage, Volume 59, 2023, 106467, ISSN 2352-152X, https://doi.org/10.1016/j.est.2022.106467.

[12] Guodao Zhang, Yisu Ge, Zi Ye, Mohammed Al-Bahrani, Multi-objective planning of energy hub on economic aspects and resources with heat and power sources, energizable, electric vehicle and hydrogen storage system due to uncertainties and demand response, Journal of Energy Storage, Volume 57, 2023, 106160, ISSN 2352-152X, https://doi.org/10.1016/j.est.2022.106160.

[13] Weiwei Dong, Guohua Zhao, Serhat Yüksel, Hasan Dinçer, Gözde Gülseven Ubay, A novel hybrid decision making approach for the strategic selection of wind energy projects, Renewable Energy, Volume 185, 2022, Pages 321-337, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2021.12.077.

[14] Paria Mansourshoar, Ahmad Sadeghi Yazdankhah, Mohsen Vatanpour, Behnam Mohammadi-Ivatloo, Impact of implementing a price-based demand response program on the system reliability in security-constrained unit commitment problem coupled with wind farms in the presence of contingencies, Energy, Volume 255, 2022, 124333, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2022.124333.

[15] Maria Grazia De Giorgi, Antonio Ficarella, Marco Tarantino, Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods, Energy, Volume 36, Issue 7, 2011, Pages 3968-3978, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2011.05.006.

[16] Joseph C.Y. Lee, Caroline Draxl, Larry K. Berg, Evaluating wind speed and power forecasts for wind energy applications using an open-source and systematic validation framework, Renewable Energy, Volume 200, 2022, Pages 457-475, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2022.09.111.

[17] M. Lydia, S. Suresh Kumar, A. Immanuel Selvakumar, G. Edwin Prem Kumar, A comprehensive review on wind turbine power curve modelling techniques, Renewable and Sustainable Energy Reviews, Volume 30, 2014, Pages 452-460, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2013.10.030.

[18] Harsh S. Dhiman, Dipankar Deb, Valentina Emilia Balas, Chapter 2 – Wind energy fundamentals, Editor(s): Harsh S. Dhiman, Dipankar Deb, Valentina Emilia Balas, In Wind Energy Engineering, Supervised Machine Learning in Wind Forecasting and Ramp Event Prediction, Academic Press, London, UK 2020, Pages 9-21, ISBN 9780128213537, https://doi.org/10.1016/B978-0-12-821353-7.00013-2.

[19] J.T. DAVIES, CHAPTER 1 – VELOCITIES AND STRESSES IN TURBULENT FLOWS, Editor(s): J.T. DAVIES, Turbulence Phenomena, Academic Press, London, UK 1972, Pages 1-78, ISBN 9780122060700, https://doi.org/10.1016/B978-0-12-206070-0.50006-7.

[20] Brandi, Aldo. (2016). "Evening transition between anabatic and katabatic wind flow regimes in complex terrain". 10.13140/RG.2.2.16134.98888.

[21] C. Pérez Albornoz, M.A. Escalante Soberanis, V. Ramírez Rivera, M. Rivero, Review of atmospheric stability estimations for wind power applications, Renewable and Sustainable Energy Reviews, Volume 163, 2022, 112505, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2022.1125

05.

[22] Alexander Kalmikov, Chapter 2 – Wind Power Fundamentals, Editor(s): Trevor M. Letcher, Wind Energy Engineering, Academic Press, London, UK 2017, Pages 17-24, ISBN 9780128094518, https://doi.org/10.1016/B978-0-12-809451-8.00002-3.

[23] Sohoni, Vaishali, S. C. Gupta and Rahul Nema. "A Critical Review on Wind Turbine Power Curve Modelling Techniques and Their Applications in Wind Based Energy Systems." Journal of Energy 2016 (2016): 1-18.

[24] Terence C. Mills, Chapter 1 – Time Series and Their Features, Editor(s): Terence C. Mills, Applied Time Series Analysis, Academic Press, London, UK 2019, Pages 1-12, ISBN 9780128131176, https://doi.org/10.1016/B978-0-12-813117-6.00001-6.

[25] Han Wang, Ning Zhang, Ershun Du, Jie Yan, Shuang Han, Yongqian Liu, A comprehensive review for wind, solar, and electrical load forecasting methods, Global Energy Interconnection, Volume 5, Issue 1, 2022, Pages 9-30, ISSN 2096-5117, https://doi.org/10.1016/j.gloei.2022.04.002.

[26] Shuai Hu, Yue Xiang, Hongcai Zhang, Shanyi Xie, Jianhua Li, Chenghong Gu, Wei Sun, Junyong Liu, Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction, Applied Energy, Volume 293, 2021, 116951, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2021.116951.

[27] Anne Gharaibeh, Abdulrazzaq Shaamala, Rasha Obeidat, Salman Al-Kofahi, Improving land-use change modelling by integrating ANN with Cellular Automata-Markov Chain model, Heliyon, Volume 6, Issue 9, 2020, e05092, ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2020.e05092.

[28] Yu Jiang, Zhe Song, Andrew Kusiak, Very short-term wind speed forecasting with Bayesian structural break model, Renewable Energy, Volume 50, 2013, Pages 637-647, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2012.07.041.

[29] Tinghui Ouyang, Andrew Kusiak, Yusen He, Modeling wind-turbine power curve: A data partitioning and mining approach, Renewable Energy, Volume 102, Part A, 2017, Pages 1-8, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2016.10.032.

[30] Jianzhou Wang, Jianming Hu, A robust combination approach for short-term wind speed forecasting and analysis – Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model, Energy, Volume 93, Part 1, 2015, Pages 41-56, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2015.08.045.

[31] Zhengtang Liang, Jun Liang, Li Zhang, Chengfu Wang, Zhihao Yun, Xu Zhang, Analysis of multi-scale chaotic characteristics of wind power based on Hilbert–Huang transform and Hurst analysis, Applied Energy, Volume 159, 2015, Pages 51-61, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2015.08.111.

[32] J.P.S. Catalão, H.M.I. Pousinho, V.M.F. Mendes, Short-term wind power forecasting in Portugal by neural networks and wavelet transform, Renewable Energy, Volume 36, Issue 4, 2011, Pages 1245-1251, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2010.09.016.

[33] Ma, Qingwen, Sihan Liu, Xinyu Fan, Chen Chai, Yangyang Wang, and Ke Yang. 2020. “A Time Series Prediction Model of Foundation Pit

Deformation Based on Empirical Wavelet Transform and NARX Network" Mathematics 8, no. 9: 1535. https://doi.org/10.3390/math8091535.

[34] Li Zhang et al 2019 IOP Conf. Ser.: Earth Environ. Sci. 252 032052, Xi'an, China. DOI 10.1088/1755-1315/252/3/032052.

[35] Hossain M, Mekhilef S, Afifi F, Halabi LM, Olatomiwa L, Seyedmahmoudian M, et al. (2018) Application of the hybrid ANFIS models for long term wind power density prediction with extrapolation capability. PloS ONE 13(4): e0193772. https://doi.org/10.1371/journal.pone.0193772.

[36] Dhiman, Harsh S., and Dipankar Deb. "A review of wind speed and wind power forecasting techniques." arXiv preprint arXiv:2009.02279 (2020), https://doi.org/10.48550/arXiv.2009.02279.

[37] Luis Torgo and Orlando Ohashi. 2011. 2D-interval predictions for time series. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11). Association for Computing Machinery, New York, NY, USA, 787–794. https://doi.org/10.1145/2020408.2020546.

[38] J.L. Torres, A. García, M. De Blas, A. De Francisco, Forecast of hourly average wind speed with ARMA models in Navarre (Spain), Solar Energy, Volume 79, Issue 1, 2005, Pages 65-77, ISSN 0038-092X, https://doi.org/10.1016/j.solener.2004.09.013.

[39] Pinson, P. (2012), Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions. Journal of the Royal Statistical Society: Series C (Applied Statistics), 61: 555-576. https://doi.org/10.1111/j.1467-9876.2011.01026.x.

[40] Tanveer Ahmad, Hongcai Zhang, Biao Yan, A review on Renewable energy and electricity requirement forecasting models for smart grid and buildings, Sustainable Cities and Society, Volume 55, 2020, 102052, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2020.102052.

[41] Tinghui Ouyang, Xiaoming Zha, Liang Qin, Yusen He, Zhenhao Tang, Prediction of wind power ramp events based on residual correction, Renewable Energy, Volume 136, 2019, Pages 781-792, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2019.01.049.

[42] Sommerfeld, M., Dörenkämper, M., De Schutter, J., and Crawford, C.: Offshore and onshore ground-generation airborne wind energy power curve characterization, Wind Energ. Sci. Discuss. [preprint], https://doi.org/10.5194/wes-2020-120, in review, 2020.

[43] Wang Y, Hu Q, Srinivasan D, Wang Z. Wind power curve modelling and wind power forecasting with inconsistent data. IEEE Trans Sustain Energy 2018;10(1):16–25.

[44] Riahy G, Abedi M. Short term wind speed forecasting for wind turbine applications using linear prediction method. Renewable Energy 2008;33(1):35–41.

[45] Zhang J, Yan J, Infield D, Liu Y, Lien F-s. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and gaussian mixture model. Applied Energy 2019;241:229–44.

[46] Prósper MA, Otero-Casal C, Fernández FC, Miguez-Macho G. Wind power forecasting for a real onshore wind farm on complex terrain using WRF high resolution simulations. Renewable Energy 2019;135:674–86.

[47] Yan J, Zhang H, Liu Y, Han S, Li L, Lu Z. Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping.

IEEE Trans Power Syst 2018;33(3):3276–84.

[48] Yu Y, Han X, Yang M, Yang J. Probabilistic prediction of regional wind power based on spatiotemporal quantile regression. IEEE Trans Ind Appl 2020;56:6117–27.

[49] Santamaría-Bonfil G, Reyes-Ballesteros A, Gershenson C. Wind speed forecasting for wind farms: A method based on support vector regression. Renewable Energy 2016;85:790–809.

[50] Liu H, Chen C, Lv X, Wu X, Liu M. Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods. Energy conversion Manage 2019;195:328–45.

[51] Wang H, Wang G, Li G, Peng J, Liu Y. Deep belief network based deterministic and probabilistic wind speed forecasting approach. Applied Energy 2016;182:80–93.

[52] Zhao X, Jiang N, Liu J, Yu D, Chang J. 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 conversion Manage 2020;203:112239.

[53] Yuan X, Chen C, Jiang M, Yuan Y. Prediction interval of wind power using parameter optimized Beta distribution based LSTM model. Applied Soft Computing 2019;82:105550.

[54] Hu J, Heng J, Wen J, Zhao W. Deterministic and probabilistic wind speed forecasting with de-noising-reconstruction strategy and quantile regression based algorithm. Renewable Energy 2020;162:1208–26.

[55] Banik A, Behera C, Sarathkumar TV, Goswami AK. Uncertain wind power forecasting using LSTM-based prediction interval. IET Renewable

Power Gener 2020;14(14):2657–67.

[56] Zhang Y, Wang J, Wang X. Review on probabilistic forecasting of wind power generation. Renewable Sustain Energy Rev 2014;32:255–70.

[57] Khosravi A, Nahavandi S. Combined non-parametric prediction intervals for wind power generation. IEEE Trans Sustain Energy 2013;4(4):849–56.

[58] Khosravi A, Nahavandi S, Creighton D, Atiya AF. Comprehensive review of neural network-based prediction intervals and new advances. IEEE Trans Neural Network 2011;22(9):1341–56.

[59] Zhang Z, Qin H, Liu Y, Yao L, Yu X, Lu J, Jiang Z, Feng Z. Wind speed forecasting based on quantile regression minimal gated memory network and kernel density estimation. Energy conversion Manage 2019;196:1395–409.

[60] Khosravi A, Nahavandi S, Creighton D, Atiya AF. Lower upper bound estimation method for construction of neural network-based prediction intervals. IEEE Trans Neural Netw 2010;22(3):337–46.

[61] Wang Y, Wang H, Srinivasan D, Hu Q. Robust functional regression for wind speed forecasting based on sparse Bayesian learning. Renewable Energy 2019;132:43–60.

[62] Li Y, Wang Y, Chen Z, Zou R. BayesIan robust multi-extreme learning machine. Knowl-Based Syst 2020;210:106468.

[63] Guo Z, Zhao W, Lu H, Wang J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renewable Energy 2012;37(1):241–9.

[64] Peng Z, Peng S, Fu L, Lu B, Tang J, Wang K, Li W. A novel deep learning ensemble model with data denoising for short-term wind speed forecasting. Energy conversion Manage 2020;207:112524.

[65] Liu Z, Jiang P, Zhang L, Niu X. A combined forecasting model for time series: Application to short-term wind speed forecasting. Applied Energy 2020;259:114137.

[66] Osório G, Matias J, Catalão J. Short-term wind power forecasting using adaptive neuro-fuzzy inference system combined with evolutionary particle swarm optimization, wavelet transform and mutual information. Renewable Energy 2015;75:301–7.

[67] Liu H, Tian H-q, Li Y-f. Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, wavelet packet-MLP and wavelet packet-ANFIS for wind speed predictions. Energy conversion Manage 2015;89:1–11.

[68] Liu H, Mi X, Li Y. 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 conversion Manage 2018;155:188–200.

[69] Zhang D, Peng X, Pan K, Liu Y. A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme learning machine. Energy conversion Manage 2019;180:338–57.

[70] Wu Z, Xiao L. A secondary decomposition based hybrid structure with meteorological analysis for deterministic and probabilistic wind speed forecasting. Applied Soft Computing 2019;85:105799.

[71] Liu H, Mi X-w, Li Y-f. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and elman neural network. Energy conversion Manage 2018;156:498–514.

[72] Han L, Jing H, Zhang R, Gao Z. Wind power forecast based on im-

proved long short term memory network. Energy 2019;189:116300.

[73] Moreno SR, da Silva RG, Mariani VC, dos Santos Coelho L. Multi-step wind speed forecasting based on hybrid multi-stage decomposition model and long short-term memory neural network. Energy conversion Manage 2020;213:112869.

[74] Sun N, Zhou J, Chen L, Jia B, Tayyab M, Peng T. An adaptive dynamic short-term wind speed forecasting model using secondary decomposition and an improved regularized extreme learning machine. Energy 2018;165:939–57.

[75] Peng T, Zhang C, Zhou J, Nazir MS. Negative correlation learning-based RELM ensemble model integrated with OVMD for multi-step ahead wind speed forecasting. Renewable Energy 2020;156:804–19.

[76] Mi X, Liu H, Li Y. Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine. Energy conversion Manage 2019;180:196–205.

[77] Santhosh M, Venkaiah C, Kumar DV. Ensemble empirical mode decomposition based adaptive wavelet neural network method for wind speed prediction. Energy conversion Manage 2018;168:482–93.

[78] Qu Z, Mao W, Zhang K, Zhang W, Li Z. Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network. Renewable Energy 2019;133:919–29.

[79] Zhang W, Qu Z, Zhang K, Mao W, Ma Y, Fan X. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. Energy conversion Manage 2017;136:439–51.

[80] Li C, Zhu Z, Yang H, Li R. An innovative hybrid system for wind

speed forecasting based on fuzzy preprocessing scheme and multi-objective optimization. Energy 2019;174:1219–37.

[81] Dong Q, Sun Y, Li P. A novel forecasting model based on a hybrid processing strategy and an optimized local linear fuzzy neural network to make wind power forecasting: A case study of wind farms in China. Renewable Energy 2017;102:241–57.

[82] Moreno SR, dos Santos Coelho L. Wind speed forecasting approach based on singular spectrum analysis and adaptive neuro fuzzy inference system. Renew Energy 2018;126:736–54.

[83] Niu T, Wang J, Zhang K, Du P. Multi-step-ahead wind speed forecasting based on optimal feature selection and a modified bat algorithm with the cognition strategy. Renewable Energy 2018;118:213–29.

[84] Yan X, Liu Y, Xu Y, Jia M. Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition. Energy conversion Manage 2020;225:113456.

[85] Liu H, Tian H-q, Liang X-f, Li Y-f. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. Applied Energy 2015;157:183–94.

[86] Xiang L, Li J, Hu A, Zhang Y. Deterministic and probabilistic multi-step forecasting for short-term wind speed based on secondary decomposition and a deep learning method. Energy conversion Manage 2020;220:113098.

[87] Shreya Dutta, Yanling Li, Aditya Venkataraman, Luis M. Costa, Tianxiang Jiang, Robert Plana, Philippe Tordjman, Fook Hoong Choo, Chek Fok Foo, Hans B. Puttgen, Load and Renewable Energy Forecasting for a Microgrid using Persistence Technique, Energy Procedia, Volume 143, 2017,

Pages 617-622, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2017.12.736.

[88] Charney, J.G., Fjortoft, R. and Von Neumann, J. (1950) Numerical Integration of the Barotric Vorticity Equation. Tellus, 2, 237-254. http://dx.doi.org/10.1111/j.2153-3490.1950.tb00336.x.

[89] Chapter 2 – The first and second laws, Editor(s): Paul D. Williams, Maarten H. P. Ambaum, In Developments in Weather and Climate Science, Thermal Physics of the Atmosphere (Second Edition), Elsevier, 2021, Pages 17-38, ISBN 9780128244982, https://doi.org/10.1016/B978-0-12-824498-2.00009-4.

[90] Ghali Yakoub, Sathyajith Mathew, Joao Leal, Intelligent estimation of wind farm performance with direct and indirect 'point' forecasting approaches integrating several NWP models, Energy, Volume 263, Part D, 2023, 125893, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2022.125893.

[91] Jung J, Broadwater RP. Current status and future advances for wind speed and power forecasting. Renewable Sustainable Energy Rev 2014;31:762–77.

[92] Sultan Al-Yahyai, Yassine Charabi, Adel Gastli, Review of the use of Numerical Weather Prediction (NWP) Models for wind energy assessment, Renewable and Sustainable Energy Reviews, Volume 14, Issue 9, 2010, Pages 3192-3198, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2010.07.001.

[93] S. Alessandrini, S. Sperati, P. Pinson, A comparison between the ECMWF and COSMO Ensemble Prediction Systems Applied to short-term wind power forecasting on real data, Applied Energy, Volume 107, 2013, Pages 271-280, ISSN 0306-2619, https://doi.org/10.1016/j.ap

energy.2013.02.041.

[94] B. Navascués, J. Calvo, G. Morales, C. Santos, A. Callado, A. Cansado, J. Cuxart, M. Díez, P. del Río, P. Escribà, O. García-Colombo, J.A. García-Moya, C. Geijo, E. Gutiérrez, M. Hortal, I. Martínez, B. Orfila, J.A. Parodi, E. Rodríguez, J. Sánchez-Arriola, I. Santos-Atienza, J. Simarro, Long-term verification of HIRLAM and ECMWF forecasts over Southern Europe: History and perspectives of Numerical Weather Prediction at AEMET, Atmospheric Research, Volumes 125–126, 2013, Pages 20-33, ISSN 0169-8095, https://doi.org/10.1016/j.atmosres.2013.01.010.

[95] M Lindskog, H Järvinen, D.B Michelson, Assimilation of radar radial winds in the HIRLAM 3D-var, Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere, Volume 25, Issues 10–12, 2000, Pages 1243-1249, ISSN 1464-1909, https://doi.org/10.1016/S1464-1909(00)00187-8.

[96] J. Martín-Vaquero, B.A. Wade, On efficient numerical methods for an initial-boundary value problem with nonlocal boundary conditions, Applied Mathematical Modelling, Volume 36, Issue 8, 2012, Pages 3411-3418, ISSN 0307-904X, https://doi.org/10.1016/j.apm.2011.10.021.

[97] Matija Perne, Marija Zlata Božnar, Boštjan Grašič, Primož Mlakar, Juš Kocijan, Improving wind vector predictions for modelling of atmospheric dispersion during Seveso-type accidents, Atmospheric Pollution Research, Volume 12, Issue 2, 2021, Pages 76-83, ISSN 1309-1042, https://doi.org/10.1016/j.apr.2020.10.010.

[98] S. Buhan and I. Çadırcı, "Multistage Wind-Electric Power Forecast by Using a Combination of Advanced Statistical Methods," in IEEE Transactions on Industrial Informatics, vol. 11, no. 5, pp. 1231-1242, Oct. 2015,

doi: 10.1109/TII.2015.2431642.

[99] Troen, I, L. Landberg: Short-Term Prediction of Local Wind Conditions. Proceedings of the European Community Wind Energy Conference, Madrid (ES), September10-14, 1990, pp. 76-78, ISBN 0-9510271-8-2

[100] Nielsen, T. S., Madsen, H., Tøfting, J. (1998). WPPT, A Tool for On-Line Wind Power Prediction.

[101] Xiaochen Wang, Peng Guo, Xiaobin Huang, A Review of Wind Power Forecasting Models, Energy Procedia, Volume 12, 2011, Pages 770-778, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2011.10.103.

[102] Dazhi Yang, Jan Kleissl, Summarizing ensemble NWP forecasts for grid operators: Consistency, elicitability, and economic value, International Journal of Forecasting, 2022, , ISSN 0169-2070, https://doi.org/10.1016/j.ijforecast.2022.08.002.

[103] Muhammad Uzair Yousuf, Ibrahim Al-Bahadly, Ebubekir Avci, Statistical wind speed forecasting models for small sample datasets: Problems, Improvements, and prospects, Energy conversion and Management, Volume 261, 2022, 115658, ISSN 0196-8904, https://doi.org/10.1016/j.enconman.2022.115658.

[104] Kim, Yeojin, and Jin Hur. 2020. "An Ensemble Forecasting Model of Wind Power Outputs Based on Improved Statistical Approaches" Energies 13, no. 5: 1071. https://doi.org/10.3390/en13051071.

[105] Wang, Jingmin Zhou, Qingwei Zhang, Xueting. (2018). Wind power forecasting based on time series ARMA model. IOP Conference Series: Earth and Environmental Science. 199. 022015. 10.1088/1755-1315/199/2/022015.

[106] S. Tian, Y. Fu, P. Ling, S. Wei, S. Liu and K. Li, "Wind Power Forecasting Based on ARIMA-LGARCH Model," 2018 International Conference on Power System Technology (POWERCON), Guangzhou, China, 2018, pp. 1285-1289, doi: 10.1109/POWERCON.2018.8601740.

[107] Rajesh G. Kavasseri, Krithika Seetharaman, Day-ahead wind speed forecasting using f-ARIMA models, Renewable Energy, Volume 34, Issue 5, 2009, Pages 1388-1393, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2008.09.006.

[108] Tanveer Ahmad, Huanxin Chen, A review on machine learning forecasting growth trends and their real-time applications in different energy systems, Sustainable Cities and Society, Volume 54, 2020, 102010, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2019.102010.

[109] Aghajani, Afshin Kazemzadeh, Rasool Ebrahimi, Afshin. (2016). A novel hybrid approach for predicting wind farm power production based on wavelet transform, hybrid neural networks and imperialist competitive algorithm. Energy conversion and Management. 121.232-240.10.1016/j.enconman.2016.05.024.

[110] Wei Sun, Bin Tan, Qiqi Wang, Multi-step wind speed forecasting based on secondary decomposition algorithm and optimized back propagation neural network, Applied Soft Computing, Volume 113, Part A, 2021, 107894, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2021.107894.

[111] Zounemat-Kermani, M. Hourly predictive Levenberg–Marquardt ANN and multi linear regression models for predicting of dew point temperature. Meteorol Atmos Phys 117, 181–192 (2012). https://doi.org/10.1007/s00703-012-0192-x.

[112] Wang J, Wang S, Yang W. A novel non-linear combination system for short-term wind speed forecast. Renewable Energy 2019;143:1172–92.

[113] 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.

[114] Liu H, Mi X, Li Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, lstm network and elm. Energy conversion Manage 2018;159:54–64.

[115] Song J, Wang J, Lu H. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. Applied Energy 2018;215:643–58.

[116] He X, Nie Y, Guo H, Wang J. Research on a novel combination system on the basis of deep learning and swarm intelligence optimization algorithm for wind speed forecasting. IEEE Access 2020;8:51482–99.

[117] Yu Y, Han X, Yang M, Yang J. Probabilistic prediction of regional wind power based on spatiotemporal quantile regression. IEEE Trans Ind Applied 2020;56:6117–27.

[118] Neshat M, Nezhad MM, Abbasnejad E, Mirjalili S, Tjernberg LB, Garcia DA, Alexander B, Wagner M. A deep learning-based evolutionary model for shortterm wind speed forecasting: A case study of the lillgrund offshore wind farm. Energy conversion Manage 2021;236:114002.

[119] Niu Z, Yu Z, Tang W, Wu Q, Reformat M. Wind power forecasting using attention-based gated recurrent unit network. Energy 2020;196:117081.

[120] Ding M, Zhou H, Xie H, Wu M, Nakanishi Y, Yokoyama R. A gated recurrent unit neural networks based wind speed error correction model for

short-term wind power forecasting. Neurocomputing 2019;365:54–61.

[121] Wan J, Liu J, Ren G, Guo Y, Yu D, Hu Q. Day-ahead prediction of wind speed with deep feature learning. Int J Pattern Recognit Artif Intell 2016;30(05):1650011.

[122] H. Mezaache and H. Bouzgou, "Auto-Encoder with Neural Networks for Wind Speed Forecasting," 2018 International Conference on Communications and Electrical Engineering (ICCEE), El Oued, Algeria, 2018, pp. 1-5, doi: 10.1109/CCEE.2018.8634551.

[123] Zhu Q, Chen J, Shi D, Zhu L, Bai X, Duan X, Liu Y. Learning temporal and spatial correlations jointly: A unified framework for wind speed prediction. IEEE Trans Sustain Energy 2019;11(1):509–23.

[124] Chen Y, Zhang S, Zhang W, Peng J, Cai Y. Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting. Energy conversion Manage 2019;185:783–99.

[125] Ghaderi A, Sanandaji BM, Ghaderi F. Deep forecast: Deep learning-based spatio-temporal forecasting. 2017, arXiv preprint arXiv:1707.08110.

[126] S. Liang, L. Nguyen and F. Jin, "A Multi-variable Stacked Long-Short Term Memory Network for Wind Speed Forecasting," 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 2018, pp. 4561-4564, doi: 10.1109/BigData.2018.8622332.

[127] Chen J, Zeng G-Q, Zhou W, Du W, Lu K-D. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. Energy conversion Manage 2018;165:681–95.

[128] Shi J, Guo J, Zheng S. Evaluation of hybrid forecasting approaches for

wind speed and power generation time series. Renewable Sustain Energy Rev 2012;16(5):3471–80.

[129] Da Liu, Dongxiao Niu, Hui Wang, Leilei Fan, Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm, Renewable Energy, Volume 62, 2014, Pages 592-597, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2013.08.011.

[130] Hye Un Cho, Yujin Nam, Eun Ji Choi, Young Jae Choi, Hongkyo Kim, Sangmu Bae, Jin Woo Moon, Comparative analysis of the optimized ANN, SVM, and tree ensemble models using Bayesian optimization for predicting GSHP COP, Journal of Building Engineering, Volume 44, 2021, 103411, ISSN 2352-7102, https://doi.org/10.1016/j.jobe.2021.103411.

[131] Zhongyue Su, Jianzhou Wang, Haiyan Lu, Ge Zhao, A new hybrid model optimized by an intelligent optimization algorithm for wind speed forecasting, Energy conversion and Management, Volume 85, 2014, Pages 443-452, ISSN 0196-8904, https://doi.org/10.1016/j.enconman.2014.05.058.

[132] Pei Du, Jianzhou Wang, Zhenhai Guo, Wendong Yang, Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting, Energy conversion and Management, Volume 150, 2017, Pages 90-107, ISSN 0196-8904, https://doi.org/10.1016/j.enconman.2017.07.065.

[133] Xiao L, Shao W, Jin F, Wu Z. A self-adaptive kernel extreme learning machine for short-term wind speed forecasting. Applied Soft Computing 2021;99:106917.

[134] Li L-L, Chang Y-B, Tseng M-L, Liu J-Q, Lim MK. Wind power prediction using a novel model on wavelet decomposition-support vector

machines-improved atomic search algorithm. J Cleaner Prod 2020;270:121817.

[135] Chen Y, Dong Z, Wang Y, Su J, Han Z, Zhou D, Zhang K, Zhao Y, Bao Y. Short-term wind speed predicting framework based on eemd-ga-lstm method under large scaled wind history. Energy conversion Manage 2021;227:113559.

[136] Hu Y-L, Chen L. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm. Energy conversion Manage 2018;173:123–42.

[137] Wang J, Yang W, Du P, Niu T. A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm. Energy conversion Manage 2018;163:134–50. [138] Luo L, Li H, Wang J, Hu J. Design of a combined wind speed forecasting system based on decomposition-ensemble and multi-objective optimization approach. Applied Math Model 2021;89:49–72.

[139] Wu C, Wang J, Chen X, Du P, Yang W. A novel hybrid system based on multi-objective optimization for wind speed forecasting. Renewable Energy 2020;146:149–65.

[140] Hu Y-L, Chen L. A nonlinear hybrid wind speed forecasting model using LSTM network, hysteretic ELM and differential evolution algorithm. Energy conversion Manage 2018;173:123–42.

[141] Wen X. Modeling and performance evaluation of wind turbine based on ant colony optimization-extreme learning machine. Applied Soft Comput 2020;94:106476.

[142] Zhang D, Peng X, Pan K, Liu Y. A novel wind speed forecasting based on hybrid decomposition and online sequential outlier robust extreme

learning machine. Energy conversion Manage 2019;180:338–57.

[143] Mahmoud T, Dong Z, Ma J. An advanced approach for optimal wind power generation prediction intervals by using self-adaptive evolutionary extreme learning machine. Renewable Energy 2018;126:254–69.

[144] Fu W, Wang K, Li C, Tan J. Multi-step short-term wind speed forecasting approach based on multi-scale dominant ingredient chaotic analysis, improved hybrid GWO-SCA optimization and ELM. Energy conversion Manage 2019;187:356–77.

[145] Xiao L, Shao W, Jin F, Wu Z. A self-adaptive kernel extreme learning machine for short-term wind speed forecasting. Applied Soft Computing 2021;99:106917.

[146] 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. Applied Soft Computing 2021;100:106996.

[147] Wang J, Heng J, Xiao L, Wang C. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. Energy 2017;125:591–613.

[148] Liu H, Chen C. Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction. Applied Energy 2019;254:113686.

[149] M. Negnevitsky, P. Johnson and S. Santoso, "Short term wind power forecasting using hybrid intelligent systems," 2007 IEEE Power Engineering Society General Meeting, Tampa, FL, USA, 2007, pp. 1-4, doi:10.1109/PES.2007.385453.

[150] Hoolohan Victoria, Tomlin Alison S, Cockerill Timothy. Improved

near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data. Renewable Energy 2018;126.

[151] Murthy KSR, Rahi OP. A comprehensive review of wind resource assessment. Renewable Sustain Energy Rev 2017;72.

[152] Sempreviva AM, Barthelmie RJ, Pryor SC. Review of methodologies for offshore wind resource assessment in European seas. Surv Geophys 2008;29.

[153] Singh Shikha, Bhatti TS, Kothari DP. A review of wind-resource-assessment technology. J Energy Eng 2006;132.

[154] Zhang Chi, Wei Haikun, Zhao Xin, Liu Tianhong, Zhang Kanjian. A Gaussian process regression based hybrid approach for short-term wind speed prediction. Energy conversion Manage 2016;126.

[155] S. S. Soman, H. Zareipour, O. Malik and P. Mandal, "A review of wind power and wind speed forecasting methods with different time horizons," North American Power Symposium 2010, Arlington, TX, USA, 2010, pp. 1-8, doi:10.1109/NAPS.2010.5619586.

[156] Hu Jianming, Wang Jianzhou. Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. Energy 2015;93.

[157] Kandasamy Kirthevasan, Dasarathy Gautam, Oliva Junier, Schneider Jeff, Póczos Barnabás. Multi-fidelity Gaussian process bandit optimisation. J Artificial Intelligence Res 2019;66.

[158] Tastu Julija, Pinson Pierre, Kotwa Ewelina, Madsen Henrik, Nielsen Henrik Aa. Spatio-temporal analysis and modeling of short-term wind

power forecast errors. Wind Energy 2011;14.

[159] Cadenas Erasmo, Rivera Wilfrido, Campos-Amezcua Rafael, Cadenas Roberto. Wind speed forecasting using the NARX model, case: La mata, oaxaca, México. Neural Comput Applied 2016;27.

[160] Liu Jingwen, Gu Yanlei, Kamijo Shunsuke. Customer pose estimation using orientational spatio-temporal network from surveillance camera. Multimedia Syst 2018;24.

[161] Dalmau, R., Perez-Batlle, M., Prats, X. Estimation and prediction of weather variables from surveillance data using spatio-temporal Kriging. A: Digital Avionics Systems Conference. "DASC 2017: 36th Digital Avionics Systems Conference: St. Petersburg, Florida, USA: September 17-21, 2017: proceedings papers". St. Petersburg, Florida: 2017, p. 1-8.

[162] Higdon Dave, Picard Dominique, Parussini L, Venturi D, Perdikaris P, Karniadakis GE. Multi-fidelity Gaussian process regression for computer experiments Thèse dirigée par Josselin Garnier Thèse rapportée par. J Comput Phys 2017;336.

[163] Alamaniotis Miltiadis, Karagiannis Georgios. Integration of Gaussian processes and particle swarm optimization for very-short term wind speed forecasting in smart power. Int J Monit Surveillance Technol Res 2018;5.

[164] Kennedy MC, O'Hagan A. Predicting the output from a complex computer code when fast approximations are available. Biometrika 2000;87.

[165] Perdikaris P, Raissi M, Damianou A, Lawrence ND, Karniadakis GE. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. Proc R Soc A 2017;473.

[166] Rasmussen Carl Edward, Williams Christopher KI. Gaussian pro-

cesses for machine learning. 2018 https://doi.org/10.7551/mitpress/3206.001.0001.

[167] Li Yunhua, Ling Lina, Chen Jiantao. Combined grey prediction fuzzy control law with application to road tunnel ventilation system. J Applied Res Technol 2015;13.

[168] Floors Rogier, Peña Alfredo, Lea Guillaume, Vasiljević Nikola, Simon Elliot, Courtney Michael. The RUNE experiment-a database of remote-sensing observations of near-shore winds. Remote Sens 2016;8.

[169] Floors R, Hahmann AN, Peña A. Evaluating mesoscale simulations of the coastal flow using lidar measurements. J Geophys Res: Atmos 2018;123.

[170] Høyer Jacob L, Karagali Ioanna. Sea surface temperature climate data record for the north sea and baltic sea. J Clim 2016;29.

[171] Global wind report 2021 - Global Wind Energy Council - GWEC (no date). Available at: https://gwec.net/global-wind-report-2021/.

[172] Yun Wang, Runmin Zou, Fang Liu, Lingjun Zhang, Qianyi Liu, A review of wind speed and wind power forecasting with deep neural networks, Applied Energy, Volume 304, 2021, 117766, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2021.117766.

[173] Zongxi Qu, Wenqian Mao, Kequan Zhang, Wenyu Zhang, Zhipeng Li, Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network, Renewable Energy, Volume 133, 2019, Pages 919-929, ISSN 0960-1481, https://doi.org/10.1016/j.renene.2018.10.043.

[174] Hui Liu, Hong-qi Tian, Xi-feng Liang, Yan-fei Li, Wind speed forecast-

ing approach using secondary decomposition algorithm and Elman neural networks, Applied Energy, Volume 157, 2015, Pages 183-194, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2015.08.014.

[175] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., Liu, H. H. (1998). The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. Proceedings: Mathematical, Physical and Engineering Sciences, 454(1971), 903–995. http://www.jstor.org/stable/53161.

[176] Na Sun, Jianzhong Zhou, Lu Chen, Benjun Jia, Muhammad Tayyab, Tian Peng, An adaptive dynamic short-term wind speed forecasting model using secondary decomposition and an improved regularized extreme learning machine, Energy, Volume 165, Part B, 2018, Pages 939-957, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2018.09.180.

[177] Torres M.E., Colominas M.A., Schlotthauer G. A complete ensemble empirical mode decomposition with adaptive noise; Proceedings of the 36th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Prague, Czech Republic. 22–27 May 2011; pp. 4144–4147.

[178] Jikai Duan, Hongchao Zuo, Yulong Bai, Jizheng Duan, Mingheng Chang, Bolong Chen, Short-term wind speed forecasting using recurrent neural networks with error correction, Energy, Volume 217, 2021, 119397, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2020.119397.

[179] Sepp Hochreiter, Jürgen Schmidhuber; Long Short-Term Memory. Neural Comput 1997; 9 (8): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

[180] Xiaohui Yuan, Chen Chen, Min Jiang, Yanbin Yuan, Prediction interval of wind power using parameter optimized Beta distribution based

LSTM model, Applied Soft Computing, Volume 82, 2019, 105550, ISSN 1568-4946, https://doi.org/10.1016/j.asoc.2019.105550.

[181] Banik, A., Behera, C., Sarathkumar, T.V. and Goswami, A.K. (2020), Uncertain wind power forecasting using LSTM-based prediction interval. IET Renew. Power Gener., 14: 2657-2667. https://doi.org/10.1049/iet-rpg.2019.1238.

[182] Huang, L., Li, L., Wei, X. et al. Short-term prediction of wind power based on BiLSTM–CNN–WGAN-GP. Soft Comput 26, 10607–10621 (2022). https://doi.org/10.1007/s00500-021-06725-x.

[183] Singla, P., Duhan, M. Saroha, S. An ensemble method to forecast 24-h ahead solar irradiance using wavelet decomposition and BiLSTM deep learning network. Earth Sci Inform 15, 291–306 (2022). https://doi.org/10.1007/s12145-021-00723-1.

[184] S. Siami-Namini, N. Tavakoli and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 3285-3292, doi: 10.1109/BigData47090.2019.9005997.

[185] Tao Liang, Qing Zhao, Qingzhao Lv, Hexu Sun, A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers, Energy, Volume 230, 2021, 120904, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2021.120904.

[186] Seyedali Mirjalili, Seyed Mohammad Mirjalili, Andrew Lewis, Grey Wolf Optimizer, Advances in Engineering Software, Volume 69, 2014, Pages 46-61, ISSN 0965-9978, https://doi.org/10.1016/j.advengsoft.2013.12.007.

[187] Murdock, Hannah E., Gibb, Duncan, Andre, Thomas, Sawin, Janet L., Brown, Adam, Andre, Thomas, Appavou, Fabiani, Brown, Adam, Ellis, Geraint, Epp, Baerbel, Gibb, Duncan, Guerra, Flavia, Joubert, Fanny, Kamara, Ron, Kondev, Bozhil, Levin, Rachele, Murdock, Hannah E., Sawin, Janet L., Seyboth, Kristin, Skeen, Jonathan, Sverrisson, Freyr, Wright, Glen, Corcoran, Fiona, Yaqoob, Hend, Assoum, Dalia, Gicquel, Stephanie, Hamirwasia, Vibhushree, Ranalder, Lea, Satzinger, Katharina, Williamson, Laura E., Findlay, Katherine, Swenson, Anna, Urbani, Florencia, Froning, Sabine, Reise, Niels, Mastny, Lisa (2020). Renewables 2020 - Global status report (INIS-FR–20-1110). France.

[188] Fs-unep-centre.org [online] Available at: https://www.fs-unep-centre.org/wp-content/uploads/2020/06/GTR_2020.pdf, 2022.

[189] Chen Wang, Shenghui Zhang, Peng Liao, Tonglin Fu, Wind speed forecasting based on hybrid model with model selection and wind energy conversion, Renewable Energy 196 (2022) 763–781. ISSN 0960-1481.

[190] H. Victoria, S. Tomlin Alison, Cockerill Timothy, Improved near surface wind speed predictions using Gaussian process regression combined with numerical weather predictions and observed meteorological data, Renew. Energy 126 (2018).

[191] S.M. Weekes, A.S. Tomlin, S.B. Vosper, A.K. Skea, M.L. Gallani, J.J. Standen, Long-term wind resource assessment for small and medium-scale turbines using operational forecast data and measure–correlate–predict, Renew. Energy 81 (2015) 760–769. ISSN 0960-1481.

[192] A.M. Sempreviva, R.J. Barthelmie, S.C. Pryor, Review of methodologies for offshore wind resource assessment in European seas, Surv. Geophys. 29 (2008).

[193] S. Shikha, T.S. Bhatti, D.P. Kothari, A review of wind-resource-assessment technology, J. Energy Eng. 132 (2006).

[194] C. Zhang, W. Haikun, Z. Xin, L. Tianhong, Z. Kanjian, A Gaussian process regression based hybrid approach for short-term wind speed prediction, Energy conversion. Manag. 126 (2016).

[195] D.R. Drew, J.F. Barlow, T.T. Cockerill, M.M. Vahdati, The importance of accurate wind resource assessment for evaluating the economic viability of small wind turbines, Renew. Energy 77 (2015) 493–500, https://doi.org/10.1016/j. renene.2014.12.032. ISSN 09601481.

[196] Tastu, J., Pinson, P., Kotwa, E., Madsen, H. and Nielsen, H.A. (2011), Spatio-temporal analysis and modeling of short-term wind power forecast errors. Wind Energ., 14: 43-60. https://doi.org/10.1002/we.401.

[197] Benjamin M. Marlin, Richard S. Zemel, Sam T. Roweis, and Malcolm Slaney. 2011. Recommender systems: missing data and statistical model estimation. In Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Three (IJCAI'11). AAAI Press, 2686–2691.

[198] B. Munoz, V.M. Lesser, R.A. Smith, B.M. Munoz Virginia Lesser Ruben A Smith, Applying multiple imputation with geostatistical models to account for item nonresponse in environmental data, J. Mod. Appl. Stat. Methods 9 (2010).

[199] Q. Zhang, Q. Yuan, C. Zeng, X. Li, Y. Wei, Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network, IEEE Trans. Geosci. Rem. Sens. 56 (8) (2018) 4274–4288.

[200] Wang, X., Li, A., Jiang, Z. et al. Missing value estimation for DNA microarray gene expression data by Support Vector Regression imputation and orthogonal coding scheme. BMC Bioinformatics 7, 32 (2006). https://doi.org/10.1186/1471-2105-7-32.

[201] M. Agathokleous, N. Tsapatsoulis, Voting Advice Applications: missing value estimation using matrix factorization and collaborative filtering, IFIP Adv. Inf. Commun. Technol. 412 (2013) 20–29, https://doi.org/10.1007/978-3-642-41142-73. Springer New York LLC.

[202] Ungar, L. H. Foster, D. P. (1998). Clustering Methods for Collaborative Filtering. Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence (AAAI'98) (p./pp. 112–125), July, Madison, Wisconsin, USA: AAAI Press.

[203] Yehuda K, Robert B, Chris V. "Matrix factorization techniques for recommender systems," in Computer, vol. 42, no. 8, pp. 30-37, doi: 10.1109/MC.2009.263.2009.

[204] T. Zhou, H. Shan, A. Banerjee, G. Sapiro, Kernelized probabilistic matrix factorization: exploiting graphs and side information, in: SIAM International Confer- Ence on Data Mining, 2012, pp. 403–414.

[205] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic Matrix Factorization. In Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS'07). Curran Associates Inc., Red Hook, NY, USA, 1257–1264.

[206] B.B. Hu, A.B. Hu, EXTENDED BAYESIAN MATRIX FACTORIZATION WITH NON-random MISSING DATA bayesian matrix factorization with NonRandom missing data using informative Gaussian process priors and soft evidences bence Bolg'ar. Proceedings of the eighth inter-

national conference on probabilistic graphical models, in: Proceedings of Machine Learning Research vol. 52, 2016, pp. 25–36.

[207] H. Shan and A. Banerjee, "Generalized Probabilistic Matrix Factorizations for Collaborative Filtering," 2010 IEEE International Conference on Data Mining, Sydney, NSW, Australia, 2010, pp. 1025-1030, doi:10.1109/ICDM.2010.116.

[208] J. Gilles, Empirical wavelet transform, IEEE Trans. Signal Process. 61 (16) (Aug.15, 2013) 3999–4010, https://doi.org/10.1109/TSP.2013.2265222.

[209] R. Sarkar, S. Julai, S. Hossain, W.T. Chong, M. Rahman, A comparative study of activation functions of NAR and NARX neural network for long-term wind speed forecasting in Malaysia, Math. Probl Eng. 2019 (2019).

[210] B. Elshafei, A. Pe˜na, D. Xu, J. Ren, J. Badger, F.M. Pimenta, et al., A hybrid solution for offshore wind resource assessment from limited onshore measurements, Appl. Energy 298 (2021), https://doi.org/10.1016/j.apenergy.2021.117245.